Perishable Capacity

Airline network planning and revenue management from first principles



Prepared for Enzo Simier

2026

Perishable Capacity: Airline Network Planning and Revenue Management from First Principles

Personal study edition prepared for Enzo Simier, Network Planning Analyst, Montréal. October 2026.

Written from the Network Planning Reading Library assembled on 27 October 2026: the MIT OpenCourseWare courses 16.75J **Airline Management** (Peter P. Belobaba, Spring 2006) and 1.206J **Airline Schedule Planning** (Cynthia Barnhart, Spring 2003); chapters of Kenneth Train's **Discrete Choice Methods with Simulation**; the network planning preparation handbook; Air Canada's second-quarter 2026 MD&A; the Competition Bureau's 2025 airline market study; the IATA Worldwide Airport Slot Guidelines; EUROCONTROL's think paper on punctuality; the Airbus A321 airport planning manual; and the current public sources cited in the text.

MIT OpenCourseWare material is adapted under the Creative Commons BY-NC-SA 4.0 licence, and this book is shared under the same licence: it may be copied and adapted for non-commercial purposes, with attribution, under the same terms. Quotations from other sources are brief and attributed, and every figure has been redrawn.

Laurentian, the airline of the worked examples, is fictional. Its network uses real airports and aircraft types, but every demand, fare and cost figure attached to it is illustrative and invented for teaching. Nothing in this book is an Air Canada forecast, cost, target or policy. Public facts are dated October 2026 unless stated otherwise.

Set in TeX Gyre Pagella and Inter with XeLaTeX. Figures drawn with Matplotlib and TikZ. The land on the cover is sampled from NOAA's GLOBE one-kilometre land mask.

Contents

How to Use This Book **xiii**
What you need to know already . . . xiii
How the book is organized . . . xiii
Reading paths . . . xiv
Conventions . . . xiv
The running case . . . xiv
About the sources . . . xv

I The Airline as a System **1**

1 The Planning Problem **3**
1.1 The planning decisions and their horizons . . . 4
1.2 Decomposition and feedback . . . 7
1.3 Why airline planning is hard . . . 11
1.4 The network planning analyst . . . 13
1.5 Incremental network contribution . . . 17
1.6 Exercise A: is the full-looking route the best use? . . . 20
1.7 Laurentian, the running case . . . 23
1.8 Air Canada in October 2026 . . . 25
1.9 Summary . . . 27
1.10 Further reading . . . 28

2 Measuring an Airline **29**
2.1 Four objects: movements, legs, itineraries and tickets 30
2.2 Passengers per day each way . . . 32
2.3 Capacity, traffic and load factor . . . 33
2.4 Yield, unit revenue and unit cost . . . 37
2.5 Stage length . . . 39
2.6 Break-even load factor . . . 43
2.7 Reading Air Canada's second-quarter 2026 MD&A . 45
2.8 Summary . . . 50
2.9 Further reading . . . 51

3 Costs and Productivity **53**
3.1 How airlines classify costs . . . 54
3.2 Cost drivers and a flight cost model . . . 55
3.3 Why unit cost falls with stage length . . . 58
3.4 Aircraft size: trip cost versus seat cost . . . 61
3.5 Utilization and productivity . . . 63
3.6 Which costs count: decision horizons . . . 68

3.7 Laurentian's cost model at work 70
3.8 Air Canada's cost structure in 2026 72
3.9 Summary 76
3.10 Further reading 77

II Demand **79**

4 Air Travel Demand **81**
4.1 Markets, itineraries and legs 82
4.2 What drives demand 84
4.3 Price elasticity 86
4.4 Time, frequency and generalized cost 90
4.5 Seasonality and the weekly cycle 95
4.6 Market size, market share and the S-curve 96
4.7 Quality of service and the traffic a new nonstop creates 101
4.8 Summary 106
4.9 Further reading 108

5 Choice Models I: Logit **109**
5.1 Random utility 110
5.2 From extreme-value errors to the logit formula . . . 112
5.3 Substitution and the independence of irrelevant alternatives 117
5.4 Elasticities 119
5.5 Willingness to pay and the value of time 121
5.6 Estimation by maximum likelihood 123
5.7 The logsum and consumer surplus 128
5.8 From probabilities to market shares 132
5.9 Summary 134
5.10 Further reading 135

6 Choice Models II: Nested, Mixed and Itinerary Choice **137**
6.1 Correlated alternatives and the GEV family 138
6.2 Nested logit 140
6.3 Mixed logit 145
6.4 The Boeing decision window model 149
6.5 The PODS passenger choice model 154
6.6 Itinerary choice in network planning tools 158
6.7 Summary 161
6.8 Further reading 162

7 Prices Are Not Random: Endogeneity **163**
7.1 Why airline fares are endogenous 164
7.2 The direction of the bias 167

7.3 A route priced by revenue management 170
7.4 Instrumental variables and two-stage least squares . 172
7.5 Instruments for airline demand 176
7.6 Logit demand with endogenous fares: Berry and BLP 180
7.7 Control functions 185
7.8 Natural experiments and difference-in-differences . 187
7.9 Before you use someone else's elasticity 191
7.10 Summary 193
7.11 Further reading 194

8 Forecasting Demand 195
8.1 What to forecast, and at which level 196
8.2 Benchmarks and exponential smoothing 197
8.3 ARIMA and dynamic regression 200
8.4 Measuring accuracy honestly 202
8.5 Forecasts that add up 205
8.6 Booking curves and pickup 208
8.7 Markets without history 212
8.8 Uncertainty, scenarios and the decision 215
8.9 Summary 218
8.10 Further reading 219

9 Spill and Recapture 221
9.1 Demand, load and spill 222
9.2 A probability model for one flight 223
9.3 The expected spill formula 225
9.4 Load factor, demand factor and spill rate 228
9.5 Unconstraining: recovering demand from censored loads 231
9.6 What is a seat worth? Capacity under uncertainty . 234
9.7 Recapture 236
9.8 Spill when revenue management controls the flight 237
9.9 Summary 239
9.10 Further reading 240

III Designing the Network 241

10 Network Structure and Connectivity 243
10.1 Point to point and hub and spoke 244
10.2 The combinatorics of connectivity 246
10.3 Economies of density, scope and scale 249
10.4 Hub banks: how connections are made 252
10.5 Connecting traffic, the sixth freedom and thin markets 254
10.6 What one more spoke adds: Seville at YUL 257

10.7 Alliances, codeshares, interline and joint ventures . 259
10.8 Freedoms of the air, bilateral agreements and Blue Sky 263
10.9 Market structure and competition in Canada 265
10.10 Hub premiums 268
10.11 Summary 271
10.12 Further reading 272

11 Fleet Planning 273
11.1 The fleet planning problem 274
11.2 Top-down and bottom-up models 276
11.3 Evaluating an aircraft 280
11.4 Payload-range diagrams 282
11.5 The A321XLR on Montreal–Seville 285
11.6 Financing and commonality 289
11.7 Turnaround time and utilization 291
11.8 A Laurentian fleet decision 294
11.9 Summary 297
11.10 Further reading 299

12 Route Evaluation 301
12.1 Three views of route economics 302
12.2 Revenue allocation and prorate 304
12.3 Feed, network contribution and displacement 307
12.4 Costs a route decision can avoid 309
12.5 The opportunity cost of aircraft time 310
12.6 The handbook's exercises 312
12.7 Laurentian YUL–SVQ: a full seasonal evaluation . . 316
12.8 Scenarios, sensitivity and the threshold that reverses the decision 320
12.9 Ramp-up after launch 322
12.10 The one-page decision memo 324
12.11 Summary 326
12.12 Further reading 327

13 Frequency and Schedule Design 329
13.1 From a route to a timetable 330
13.2 Frequency, schedule delay and total trip time 331
13.3 Frequency versus gauge 335
13.4 Choosing departure times 337
13.5 Hub banks and connections 339
13.6 Rotations, clocks and the fleet a timetable needs . . 345
13.7 Seasons, SSIM and slots 349
13.8 Schedule design as an optimization problem 352

13.9 When is a schedule change worth it? 355
13.10 Summary 358
13.11 Further reading 359

14 Fleet Assignment 361
14.1 The problem and its place in planning 362
14.2 Time-space networks 363
14.3 The fleet assignment model 367
14.4 Spill costs by flight and type 368
14.5 Solving the model 371
14.6 A Laurentian bank solved 373
14.7 Network effects: itinerary-based fleet assignment . . 378
14.8 Re-fleeting close to departure: demand driven dispatch 384
14.9 Fleet assignment in the planning chain 387
14.10 Summary 388
14.11 Further reading 390

15 Aircraft Routing and Crew Scheduling 391
15.1 From fleet types to tails and crews 392
15.2 Maintenance routing 394
15.3 Crew pairings and what they cost 398
15.4 Crew pairing as set partitioning 401
15.5 Column generation and the pricing problem 404
15.6 Branch-and-price 407
15.7 Rostering and bidlines 409
15.8 Routing and crews together 409
15.9 Fatigue rules in Canada 412
15.10 What breaks when a planner moves a flight 413
15.11 Summary 416
15.12 Further reading 417

16 Robustness and Operations Recovery 419
16.1 Where delays come from 420
16.2 Slack and propagated delay 422
16.3 What a minute of slack is worth 424
16.4 Buffers and utilization 427
16.5 Robust routing and retiming 429
16.6 Operations control and recovery 434
16.7 Measuring what passengers experience 439
16.8 Disruption in the business case 441
16.9 Summary 445
16.10 Further reading 446

IV Revenue **449**

17 Pricing and Fare Structures **451**
17.1 One cabin, many prices 452
17.2 Price discrimination in theory 453
17.3 Fences and self-selection 457
17.4 Fare products and restrictions 461
17.5 From restricted to simplified fares 464
17.6 Branded fares and bundles 466
17.7 Laurentian's YUL–SVQ fare ladder 468
17.8 Pricing, revenue management and competition . . . 471
17.9 Summary 474
17.10 Further reading 475

18 Seat Inventory Control on a Single Flight **477**
18.1 The seat allocation problem 478
18.2 The revenue management system 479
18.3 Booking classes, nesting and protection levels 481
18.4 Two classes: Littlewood's rule 483
18.5 Many classes: EMSR-a and EMSR-b 486
18.6 Dynamic programming and the optimality of nested protection levels 490
18.7 Booking curves and controls over time 495
18.8 What revenue management is worth 497
18.9 The network planner's takeaway 498
18.10 Summary 499
18.11 Further reading 500

19 Overbooking **503**
19.1 The overbooking problem 504
19.2 How many passengers show up? 506
19.3 Rules of thumb 508
19.4 The cost-based authorization level 510
19.5 Overbooking the Seville flight 514
19.6 Service-level criteria 518
19.7 Volunteers and involuntary denied boardings 519
19.8 Overbooking over the booking horizon 523
19.9 Summary 524
19.10 Further reading 526

20 Network Revenue Management **527**
20.1 Why leg control fails on a network 528
20.2 Origin-destination control: value buckets 531
20.3 Displacement and DAVN 532
20.4 Bid-price control 533

20.5 The deterministic LP and its dual 536
20.6 Laurentian's network solved 538
20.7 Uncertainty: probabilistic bid prices and simulation 541
20.8 What the PODS simulations found 546
20.9 Bid prices as the value of capacity 548
20.10 Summary 551
20.11 Further reading 552

21 Willingness to Pay, Sell-Up and Competition 553
21.1 Willingness to pay 554
21.2 Sell-up and buy-down 555
21.3 Q-forecasting 560
21.4 The spiral-down effect 562
21.5 Fare adjustment 566
21.6 Choice-based revenue management 570
21.7 Dynamic pricing 573
21.8 Willingness to pay and competition 577
21.9 Summary 579
21.10 Further reading 580

22 Where Planning Meets Revenue Management 581
22.1 Two functions, two horizons 582
22.2 Bid prices: the value of a seat, departure by departure 584
22.3 Eight valuations of the same seats 586
22.4 RM-aware spill 590
22.5 The network contribution of a new leg 595
22.6 Consistent forecasts 598
22.7 How revenue management reacts to a schedule change 600
22.8 Questions for the revenue management team 601
22.9 Summary 605
22.10 Further reading 606

V Practice 607

23 The Industry in 2026 609
23.1 Pricing and revenue management technology 610
23.2 Network planning practice and tools 616
23.3 Supply constraints: engines, deliveries and the long-range narrowbody 617
23.4 Carbon rules that enter route economics 622
23.5 The Canadian market in 2026 626
23.6 Air Canada in the autumn of 2026 629
23.7 What changed since 2006, and what did not 632

23.8 Summary 634
23.9 Further reading 635

24 The Analyst's Toolkit 637
24.1 Four objects and unit discipline 638
24.2 The data an analyst will meet 639
24.3 The data contract 644
24.4 A reproducible workflow 646
24.5 Clocks, seasons and the schedule card 648
24.6 Sensitivity matrices and decision thresholds 653
24.7 The one-page decision memo 656
24.8 Audit checks 658
24.9 A reading plan and the first weeks 660
24.10 Summary 663
24.11 Further reading 663

A Probability and Statistics Toolkit 665
A.1 Random variables, expectation and conditioning . . 665
A.2 The normal distribution 668
A.3 The loss function and truncated moments 668
A.4 Counting distributions 671
A.5 The Gumbel distribution and logit 674
A.6 Limit theorems and the normal approximation . . . 676
A.7 Estimation and inference 677
A.8 Censoring and truncation 681
A.9 Prediction intervals 682
A.10 Summary of distributions 683

B Optimization Toolkit 685
B.1 Linear programming 686
B.2 Duality 688
B.3 Sensitivity analysis 691
B.4 Integer programming 692
B.5 Network flows 698
B.6 Dynamic programming 700
B.7 Lagrangian relaxation 701
B.8 Dantzig–Wolfe decomposition and column generation 703
B.9 Solvers and modelling languages 706
B.10 Where the tools are used 707

C Glossary 709

D Notation 723
D.1 General mathematics 723
D.2 Measures, costs and capacity 724

D.3 Demand, frequency and spill 724
D.4 Choice models 725
D.5 Schedule and fleet 725
D.6 Revenue management 726
D.7 Letters with two meanings 726

E Sources and Reading List **727**
E.1 The reading library 727
E.2 Reading the older material 730
E.3 Where to go next 731

References **733**

Index **755**

How to Use This Book

This book is a self-contained course in the economics and mathematics of airline network planning, written for an analyst starting in a network planning department. It explains every concept from first principles, derives the results it uses, and works each model through numerically on one running example, so that you can reproduce the arithmetic with a spreadsheet or a few lines of Python. Revenue management appears throughout, but always from a planner's seat: how it fills the aircraft you schedule, how it changes the value of a seat, and what its outputs tell you about where capacity is scarce.

What you need to know already

The book assumes the background of a graduate student in economics: calculus, probability at the level of expectations and the normal distribution, regression and maximum likelihood, and the idea of a linear program and its dual. Nothing about airlines is assumed. Appendices A and B collect the probability and optimization results the chapters lean on, and Appendix C defines every acronym you will hear in your first month.

How the book is organized

Part I, The airline as a system, sets out the sequence of planning decisions, from fleet to operations, and the measures airlines use to describe themselves (ASK, RPK, yield, unit revenue and unit cost), then the structure of airline costs.

Part II, Demand, builds demand from the ground up: what drives air travel, how travellers choose among itineraries (logit, nested and mixed logit, with full derivations), why prices observed in airline data cannot be treated as random, how to forecast, and how to handle demand that capacity has capped (spill and recapture).

Part III, Designing the network, is the core of the planner's job: network structure and connectivity, fleet planning, route evaluation, frequency and schedule design, fleet assignment, aircraft routing and crew, and the robustness of the plan once it meets reality.

Part IV, Revenue, covers pricing and fare structures, seat inventory control on a single flight, overbooking, network revenue man-

agement with bid prices, willingness to pay and competition, and a closing chapter on the interface between planning and revenue management.

Part V, Practice, describes the industry as it stands in 2026 and gives a working toolkit: data sources, systems, workflows and the checks that keep an analysis honest.

Reading paths

If you are starting the job next week, read Chapters 1, 2, 4, 9, 10, 12, 13, 22 and 24 first. If you want the analytical depth that distinguishes a strong analyst, add Chapters 5 to 8, 14 and 20. If your desk is closer to operations and feasibility, read Chapters 11, 15 and 16. The chapters cross-reference each other, so you can enter anywhere and follow the links back.

Conventions

Every chapter opens with a short list of what it covers and the documents of the reading library it draws on. Key terms are set in **bold** where they are defined. Six kinds of boxes structure the text:

- **Definition** boxes state precise meanings.
- **Result** boxes state formulas or propositions; most are followed by a proof or derivation.
- **Worked example** boxes solve a model with numbers, usually on the running case.
- **In practice** boxes describe how the idea is used by airlines today.
- **Pitfall** boxes flag common errors.
- **Key idea** boxes isolate the one sentence to remember.

Definitions, results, examples, figures, tables and equations are numbered by chapter, so Result 9.2 is the second result of Chapter 9. Money is in Canadian dollars unless stated. Distances are in statute miles (mi) or kilometres (km) and, for aircraft performance, nautical miles (nm). The symbols used across chapters are collected in Appendix D.

The running case

Most worked examples use **Laurentian**, a fictional airline with its hub at Montreal-Trudeau (YUL). Laurentian's geography and aircraft are real; its demand, fares and costs are invented, chosen to be

plausible and to make each model's behaviour visible. Do not read any Laurentian number as a fact about Air Canada or any other airline.

Table 1 Laurentian's fleet. Trip cost is block hours times the hourly cost, plus a departure cost (CAD 2,000 to 7,000 by type and region) and a cost per passenger (CAD 20 domestic, 30 transborder, 50 transatlantic). All figures are illustrative.

Aircraft	Seats	Business	Premium economy	Economy	CAD per block hour
A220-300	137	12		125	8,000
737-8	169	16		153	9,000
A321XLR	182	14		168	11,000
787-9	298	30	21	247	15,500

Laurentian's most important route in this book is YUL–SVQ, Montreal to Seville, flown with the A321XLR, a narrowbody with transatlantic range. Its schedule pattern mirrors a real service that was announced publicly in October 2026 for launch in September 2027: an evening departure from Montreal arriving the next morning in Seville, three times a week at launch and four times a week from October. That public announcement is the only real element; everything Laurentian earns and spends on the route is invented. For network examples Laurentian also runs a small bank at YUL that connects two feeder flights, from Quebec City (YQB) and Boston (BOS), to two transatlantic flights, to Paris (CDG) and Seville (SVQ). Its four legs and sixteen fare products are small enough to solve by hand and rich enough to show every network effect in Chapters 12, 14 and 20.

About the sources

The reading library mixes teaching material from 2003 and 2006 with documents dated 2025 and 2026. The older material teaches durable methods: spill models, fleet assignment, bid prices and choice models have not changed in substance. Its fares, market shares, software descriptions and parameter values are history, and the text says so wherever it uses them. Current facts are dated and sourced. When the book makes a claim about practice in 2026, it cites a public source; when a claim could not be verified, it is labelled as such or left out.

PART I

The Airline as a System

An airline sells journeys but produces flights, and it commits to aircraft years before it knows who will fly. Part I sets out the sequence of decisions that turns a fleet into a flown timetable, the measures airlines use to describe capacity, traffic, revenue and cost, and the structure of the costs that every network decision has to cover.

CHAPTER

1

The Planning Problem

IN THIS CHAPTER

- The decisions that turn a fleet into a flown timetable, and how far ahead of the day of operation each one is taken.
- Why airlines split planning into sequential stages, what that costs, and the feedback loops that hold the stages together.
- Six reasons the problem is hard: perishable seats, network effects, uncertainty, lumpy aircraft, competition and regulation.
- The network planning analyst's decisions, deliverables and the five questions behind every recommendation.
- Incremental network contribution, defined precisely, and the handbook's Exercise A solved step by step.
- Laurentian, the fictional airline of the running case, and the public context of Air Canada in October 2026.

DRAWS ON Belobaba, Introduction to the airline planning process; Airline schedule development; Route planning and evaluation (MIT 16.75J, 2006); Barnhart, Airline schedule planning (MIT 1.206J, 2003); handbook; Air Canada Q2 2026 MD&A

Suppose that in your first week the manager of your desk asks a short question: should Laurentian, the fictional Montreal airline of this book, add a fourth weekly departure to Seville next October, or should that aircraft time go somewhere else? The question names one route, but the answer runs through the whole airline. The extra departure needs an aircraft that is free on the right days, a crew that can legally fly it, a slot and a gate in Montreal and Seville, and enough customers who would not otherwise have flown on another Laurentian flight. It changes which connections the Montreal hub can offer, what revenue management will charge on the other three departures, and what the same aircraft would have earned elsewhere. A planner who answers "the route is 85 percent full, so yes" has not answered the question.

This chapter maps the territory that question lives in. It describes the sequence of planning decisions, from fleet orders a decade ahead

to recovery on the day of operation; explains why airlines break the problem into stages and how the stages talk to each other; and pins down what a network planning analyst is trying to optimize, using the framework of the preparation handbook written for this job (Network Planning Handbook 2026). The central object is the incremental network contribution of a change, which compares two whole networks rather than two route profit statements. Two running examples are introduced: the handbook's Exercise A, solved in full, and Laurentian's small hub network at Montreal-Trudeau (YUL), which reappears in most later chapters.

The rest of Part I supplies the vocabulary. Chapter 2 defines the measures that airlines use to describe themselves (ASK, RPK, load factor, yield, unit revenue and unit cost) and reads Air Canada's latest quarterly report with them; Chapter 3 explains what drives costs and which ones a given decision can avoid. Demand, network design and revenue management, the substance of each planning stage, fill Parts II to IV.

1.1 The planning decisions and their horizons

Belobaba's introduction to airline planning lists five decisions, each with its own question (Belobaba 2006c). **Fleet planning** decides which aircraft to acquire or retire, how many and when. **Route evaluation** decides which network structure to operate and which city pairs to serve. **Schedule development** decides how often, at what times and with which aircraft each route is flown. **Pricing** decides which fare products, with which restrictions, to offer in each origin-destination market. **Revenue management** decides how many bookings to accept for each fare product on each flight. The planning chart that the lecture takes from Barnhart adds the decisions further down the line that make a schedule operable: crew scheduling, airport resource management, sales and distribution, and operations control on the day (Belobaba 2006c; Barnhart 2003).

Schedule development is itself four linked problems (Belobaba 2006f). **Frequency planning** sets the number of departures on each route. **Timetable development** sets departure and arrival times, including the timing of connections at hubs. **Fleet assignment** chooses the aircraft type for each flight. **Aircraft rotation planning** chains flights into sequences that each aircraft can fly, so that the number of aircraft arriving at every station matches the number departing. Throughout the book a **flight leg** (or segment, or sector) is one nonstop operation from A to B, a **flight** is one or more legs flown under one flight number, and an **itinerary** is the combination of legs a pas-

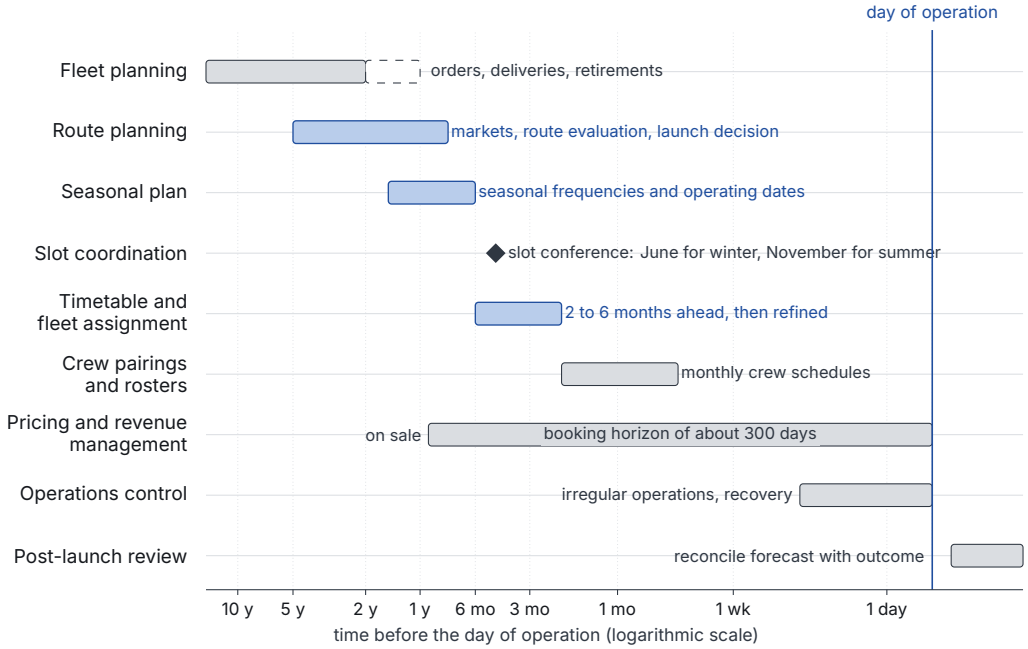


Figure 1.1 Planning decisions and their typical horizons, on a logarithmic scale of time before the day of operation. Bars are indicative: practice varies by airline, and every decision is revisited as its date approaches. Cobalt marks the decisions at the centre of a network planner's work. Sources: Belobaba's 2006 lectures on schedule development and revenue management; the IATA Worldwide Airport Slot Guidelines for slot conferences.

senger takes to complete a journey; Chapter 2 makes these objects precise.

What separates the decisions is time. Early decisions are large, slow to reverse and taken with little information; late decisions are detailed, quick and taken within the box drawn by everything decided before. Figure 1.1 places each decision on a time axis that runs from a decade before the day of operation to the review that follows it.

Fleet planning works years ahead because aircraft are expensive, long-lived and delivered slowly. The 2006 lecture quoted prices of USD 40 to 60 million for a narrowbody and over USD 200 million for a long-range widebody, depreciation lasting 10 to 15 years and aircraft flown economically for more than 30 years; the prices are historical, the orders of magnitude are not (Belobaba 2006c). Air Canada's own commitments show the lead times. The 787-10s it ordered in September 2023 begin to arrive in 2026, the eight A350-1000s announced in February 2026 are scheduled from 2030, and its 30 A321XLRs arrive between 2026 and 2029 (Air Canada 2026e). A

fleet plan is therefore a staged commitment: the number and type of aircraft in service in any future season is mostly settled several years before that season is planned in detail.

Route planning takes a fleet as given and selects the markets to serve. The economics turn on forecast demand and revenue, the share the airline can expect against competitors, the network effects of the new flight on the rest of the system and the opportunity cost of the aircraft; the practical side turns on range, operating cost, rotations, bilateral traffic rights and slots (Belobaba 2006c, 2006i). Public announcements give a lower bound on the horizon. Air Canada announced Montreal to Seville on 14 October 2026 for a first flight on 8 September 2027, almost eleven months later, and the analysis behind such an announcement is done well before it (Air Canada 2026g).

Schedule development begins a year or more ahead with frequency plans; timetables and aircraft rotations are fixed two to six months ahead, and revisions continue until departure (Belobaba 2006f). The industry runs on two scheduling seasons, a northern summer that starts on the last Sunday in March and a winter that starts on the last Sunday in October (IATA, ACI, and WWACG 2025). At airports where capacity is coordinated, airlines request slots at the IATA Slot Conference, held in June for the following winter and in November for the following summer (IATA, ACI, and WWACG 2025). **Crew scheduling** then builds pairings (sequences of legs a crew can legally fly from its base and back) and assembles them into monthly bidlines or rosters (Barnhart 2003).

Pricing and revenue management take over once a schedule is on sale. Belobaba's description of a large airline's inventory problem assumes about 300 days of bookings before departure, during which booking limits are revised continuously (Belobaba 2006l). **Operations control** owns the last days and the day itself: it monitors irregularities and implements recovery plans when weather, air traffic control, maintenance problems or diversions break the plan (Belobaba 2006j). Finally, a **post-launch review** compares what happened with what the approved proposal forecast, so that the next forecast is better (Network Planning Handbook 2026).

Two properties of this timeline matter for everything that follows. First, each stage inherits constraints from the earlier ones: a timetable cannot use aircraft the fleet plan did not buy, and revenue management cannot sell seats the timetable did not schedule. Second, the cost of changing a decision rises as the day of operation approaches, while the information available to make it improves. The handbook draws the practical conclusion: once a schedule is published it cre-

ates commitments to customers and recovery costs, so the economic threshold for cancelling a flight that is already on sale differs from the threshold for a route that has not yet been sold (Network Planning Handbook 2026). A planner always asks not only what the best schedule is, but what can still be changed and at what cost.

1.2 Decomposition and feedback

Why not choose the fleet, the routes, the timetable, the aircraft assignments, the crew schedules and the prices together, in one model? Because the joint problem is far too large. Consider a single day of 400 flights. With four aircraft types there are $4^{400} \approx 10^{241}$ ways to assign types to flights; allowing each departure to move within plus or minus 30 minutes in five-minute steps gives $13^{400} \approx 10^{446}$ timetables; choosing both gives about 10^{686} combinations, before crews, maintenance or prices enter. Most of these combinations are infeasible, which is exactly the difficulty: feasibility depends on aircraft flows, connection times, maintenance access and crew rules that couple every flight to the others. In 2006, fleet assignment for 2,000 to 3,500 flights and 10 to 14 fleets could be solved in minutes, while timetable optimization was still out of reach (Belobaba 2006f, 2006c).

Airlines therefore **decompose** the problem. The classical sequence for schedule planning has four stages: schedule design selects the flight legs and their times; fleet assignment chooses a type for each leg to maximize contribution; aircraft routing builds tail sequences that respect maintenance; crew scheduling covers the legs with legal pairings and rosters (Barnhart 2003). Each stage solves a well-defined optimization problem, taking the output of the previous stage as fixed data. The decomposition works because each stage captures the decisions that matter most to it, but it is a heuristic, and the price is a set of feedback loops through which a later stage sends problems back upstream (Figure 1.2).

The loops are where most of a planner's iteration happens. A fleet assignment that spills too many passengers, or needs more aircraft of one type than exist, sends the timetable back for different frequencies or times (loop 1). A routing that cannot bring every aircraft to a maintenance base within its interval forces a swap of types (loop 2). Crew duty limits and the cost of overnights can make a timetable that looks efficient for aircraft expensive or illegal for crews (loop 3). Delays propagate through tight connections and turns, which argues for buffers that cost utilization (loop 4, Chapter 16). A schedule that needs more aircraft than the fleet plan provides, or leaves

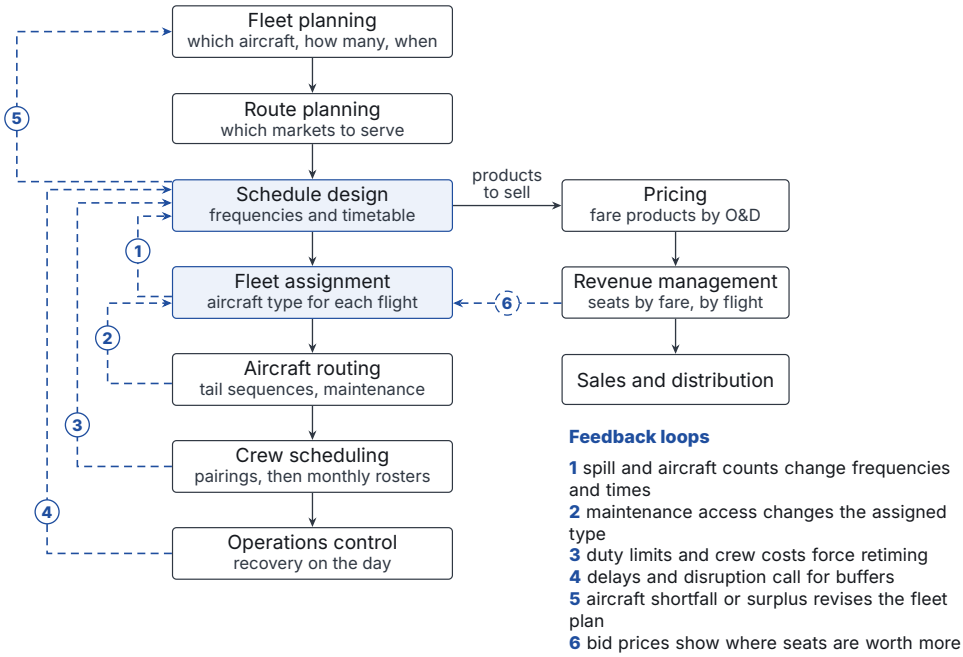


Figure 1.2 The sequential decomposition of airline planning (solid arrows) and the feedback loops that tie it together (dashed, numbered). Schedule design and fleet assignment, shaded, are the core of network planning. Adapted from the planning charts in Belobaba's 2006 introduction to the planning process and Barnhart's 2003 course on airline schedule planning.

aircraft idle, revises the fleet plan (loop 5). Finally, revenue management's bid prices show which legs are short of seats, which tells fleet assignment and schedule design where capacity is worth more (loop 6, Chapter 22).

The formal reason the loops matter is simple, and worth stating once.

RESULT 1.1 Sequential planning is feasible but not optimal

Let the planning decisions be $x = (x_1, \dots, x_m)$, chosen from a feasible set $\mathcal{X}$, with network profit $\Pi(x)$. Suppose a sequential procedure fixes $\hat{x}_1$, then $\hat{x}_2$ given $\hat{x}_1$, and so on, each stage maximizing its own objective subject to the earlier choices, and suppose the result $\hat{x}$ is feasible. Then

$$\Pi(\hat{x}) \le \max_{x \in \mathcal{X}} \Pi(x),$$

and the difference is the **value of integration**. If instead the stages are revisited in turn, each maximizing Π over its own decisions with the others held fixed, profit never decreases from one pass to the next, but the procedure can stop at a plan that no single stage can improve and that is still not the best plan.

Proof. The first claim holds because $\hat{x} \in \mathcal{X}$, so its profit cannot exceed the maximum over $\mathcal{X}$. For the second, a pass that changes only x_k chooses the best x_k given the others, and the current x_k is one of the candidates, so profit cannot fall. Profit is bounded above by the maximum, so the sequence of profits converges. A plan where no single block of decisions can be improved alone can still be improved by changing two blocks together, as the example that follows shows, so the limit need not be the maximum. ∎

A two-stage illustration with invented numbers shows the trap. Schedule design prefers a 07:00 departure because it earns CAD 2,000 more revenue than 08:00. Fleet assignment then finds that the only aircraft free at 07:00 is a larger type that costs CAD 3,500 more per trip, while at 08:00 the smaller aircraft would have been available. The sequential plan loses CAD 1,500 against the plan that moves both decisions together, and neither stage on its own would discover it. Integrated models that choose the timetable and the fleet assignment jointly exist for exactly this reason (Lohatepanont and Barnhart 2004), as do models that carry passenger itineraries into fleet assignment (Barnhart, Kniker, and Lohatepanont 2002) and that combine maintenance routing with crew pairing (Cohn and Barnhart 2003). Belobaba listed the obstacles to fully joint planning as inconsistent data, missing models of competitive behaviour and organizational coordination between departments (Belobaba 2006c). They still explain why airlines plan in stages and iterate.

WORKED EXAMPLE 1.1 One route touches every stage

Laurentian's Montreal to Seville pattern mirrors the timetable that Air Canada has published for its own service, as worked through in the handbook (Network Planning Handbook 2026): depart YUL at 20:30, arrive in Seville (SVQ) at 09:30 the next morning, leave Seville at 12:45 and arrive in Montreal at 14:40 the same day, three times a week at launch and four times a week from October, with a 182-seat A321XLR. Follow one round trip through the stages.

Timetable and time zones. On a date when Montreal is at UTC minus 4 hours and Seville at UTC plus 2, the 20:30 departure is 00:30 UTC and the 09:30 arrival is 07:30 UTC, a block time of 7 hours 0 minutes eastbound. The return leaves at 10:45 UTC and arrives at 18:40 UTC, 7 hours 55 minutes westbound, against the prevailing winds. The aircraft spends 3 hours 15 minutes on the ground in Seville and is back in Montreal 18 hours 10 minutes after it left, with 5 hours 50 minutes before another 20:30 departure. The six-hour offset does not hold all season. North America and Europe change their clocks on different dates, so for two or three weeks in March and about one week at the turn of October and November the offset is five hours; if the local times were kept, the eastbound block would have to be 8 hours and the westbound block 6 hours 55 minutes, an hour less than the flight needs. A timetable must be checked date by date, not once.

Fleet assignment. The route needs a type with transatlantic range and about the right number of seats; in Laurentian's fleet only the A321XLR and the 787-9 qualify, since the A220-300 and the 737-8 would be at the edge of their range westbound, and the 787-9 has 298 seats against Laurentian's illustrative mean eastbound demand of 150 to 175 per departure in the shoulder and peak seasons (Chapter 9 values the difference).

Aircraft routing. Three weekly round trips occupy the aircraft for $3 \times 18\frac{1}{6} = 54.5$ hours a week, 32.4 percent of the week, and four occupy it for 72.7 hours, 43.3 percent. Each round trip contains 14.9 block hours. The rest of the week the aircraft must fly something else that starts and ends at times compatible with this pattern, and it must reach maintenance; the 5 hours 50 minutes at base between rotations is a natural window for routine checks, but only if the base can do them then.

Crew scheduling. The aircraft turns in Seville in 3 hours 15 minutes, but the crew cannot: from report time before the 20:30 departure to arrival back in Montreal is more than 18 hours, far beyond a workable duty period, so in practice the outbound crew lays over in Seville and another crew flies the return. Crew planners check such a pattern against the fatigue rules (Transport Canada 2024); the planner's job is to know that each layover adds hotel nights and crew days, which are costs the timetable creates.

Revenue management. Once the schedule is on sale, every later change touches customers who have already booked, and the cost of changing the timetable rises with each booking.

The lesson: a decision that looks like one line in a route plan (four weekly instead of three) is simultaneously a fleet, routing, crew, airport and inventory decision. The planner does not solve all of these alone, but has to see them coming.

1.3 Why airline planning is hard

Many industries plan capacity under uncertainty. Airline planning combines six features that are each manageable on their own and difficult together.

> **DEFINITION 1.1 Perishable capacity**
>
> Capacity is **perishable** when it is produced at a fixed time whether or not it is sold, cannot be stored, and is worth nothing once that time has passed. A seat on a departure is perishable capacity: the seat is flown, empty or full, and an empty seat cannot be sold after the doors close.

Perishable seats. Because capacity is set before demand is known and cannot be carried over, an airline faces two problems at once. It must choose capacity under uncertainty, accepting that some flights will leave seats empty and others will turn customers away (Chapter 9). And once a flight is scheduled, almost all of its cost is fixed with respect to the number of passengers. On Laurentian's illustrative cost model, a YUL–SVQ departure with 150 passengers costs CAD 88,000, of which only CAD 7,500 (8.5 percent) varies with the number of passengers; the average cost is CAD 587 per passenger and the marginal cost CAD 50. Any customer willing to pay more

than the marginal cost and the value of the seat to someone else is worth carrying, which is the economic root of differentiated fares and of revenue management (Chapters 17 and 18).

Network effects. Demand is defined by origin-destination market, but supply is produced as flight legs, and each leg supplies seats to many markets at once (Belobaba 2006d). Belobaba's 2006 example of a single Boston to Houston flight showed bookings for over twenty origin-destination markets, with the local Boston-Houston market accounting for only about 30 percent of the bookings among the top 23 markets (values read from the lecture's chart; historical). On a hub network every planning decision on one leg changes the itineraries available in many markets, and the value of a flight cannot be read from that flight alone. Section 1.5 turns this into a measurement rule.

Uncertainty. Demand for a given departure varies from week to week, typically with a coefficient of variation between 0.2 and 0.4 (Belobaba 2006b), as Chapter 9 shows; the forecast of a new market is worse. Input prices move too: in the second quarter of 2026 Air Canada paid 50.8 percent more per litre of fuel than a year earlier (Air Canada 2026e), and it had suspended its full-year guidance on 30 April 2026 because jet fuel costs were too uncertain to forecast (The Canadian Press 2026). Weather disruptions lowered the same quarter's completed capacity below the company's guidance (Air Canada 2026e). Competitors' responses and regulatory decisions add further uncertainty that no forecast removes.

Indivisibility. Capacity comes in lumps. Laurentian's aircraft have 137, 169, 182 or 298 seats and nothing in between; frequencies are integers; an extra departure needs a whole aircraft for a whole rotation. Marginal analysis therefore has to be done with discrete alternatives, and small changes in demand can make the best plan jump from one configuration to another (Chapters 11 and 14).

Competition. Market share depends on frequency, timing, fares and product relative to competitors, and the relationship between frequency share and market share is S-shaped rather than proportional, which gives airlines an incentive to match each other's frequencies (Belobaba 2006d). Concentration is a live policy issue in Canada: the Competition Bureau found that Air Canada and WestJet carried between 56 and 78 percent of domestic passengers at major Canadian airports in 2023 (Competition Bureau Canada 2025). Every planning decision is a move in a game whose other players react (Chapters 4 and 10).

Regulation and access. International routes need traffic rights under bilateral agreements; coordinated airports allocate slots under

the IATA guidelines (IATA, ACI, and WWACG 2025); airports impose operating restrictions, and the handbook warns against summarizing Montreal's as "YUL closes at night", since the airport describes a 24-hour operation with category-specific restrictions and exceptions (Network Planning Handbook 2026); crews work under fatigue rules (Transport Canada 2024); and joint ventures between airlines need government approval, as the Air Canada and United transborder agreement does again in 2026 (Air Canada 2026h). Each of these can make an economically attractive plan infeasible.

A seat is many products at once: it serves many markets, it expires at departure, and it comes in a block with every other seat on the aircraft.

1.4 The network planning analyst

The handbook describes the analyst's intellectual core as demand estimation, substitution, constrained optimization, opportunity cost and counterfactual analysis, and the learning curve as translating those ideas into sellable itineraries, aircraft rotations and seasonal schedule decisions (Network Planning Handbook 2026). A good planner can show both that a change makes economic sense and that it can be operated. The work comes at several decision levels, summarized in Table 1.1.

Table 1.1 Decision levels of network planning (adapted from the handbook, 2026). The levels overlap in time; who owns which decision must be learned inside each airline.

Decision level	Typical question	Evidence to produce
Multi-year network and fleet	Which regions and aircraft capabilities deserve scarce capital?	Scenarios, fleet availability, demand outlook, airport and regulatory feasibility
Seasonal plan	Which routes, months, days and frequencies should operate?	Demand and contribution by period, competitors' timetables, aircraft-days required
Schedule refinement	What if we retime, upgauge or move a frequency?	Changed itineraries, revenue displacement, airport peaks, feasible rotations
In-season adjustment	Should capacity be added, cut or redeployed?	Booking outlook from RM, avoidable costs, customer commitments, feasible alternatives
Post-launch review	Why did results differ from the proposal?	A reconciled bridge of volume, fare, cost and operational variances

Across all levels, the handbook frames one question.

DEFINITION 1.2 The core decision

Among the schedules and aircraft allocations that can actually be operated, choose the one that adds the most value to the airline's whole network, measured against the best alternative that could realistically be chosen instead (Network Planning Handbook 2026).

Three ideas in the definition carry most of its weight. **Feasible** means the plan respects aircraft, crew, airport, maintenance and regulatory constraints. **Incremental** means value is measured as the difference between the network with the change and the network without it. **Alternative** means the comparison is not with doing nothing but with the best other use of the same scarce resources. The handbook turns the definition into five questions that every recommendation should answer (Network Planning Handbook 2026):

1. **The counterfactual.** What happens without the change? Specify a dated baseline schedule, assumptions about competitors and a cost scenario.
2. **The journeys.** Which passenger journeys change? Count local, connecting, partner, diverted, stimulated and lost demand.
3. **The increment.** What is new to the airline, as opposed to traffic that merely moves between its own services?
4. **The binding resources.** Which resources become scarce? Aircraft, seats on feeder flights, crews, slots, gates, maintenance access and cash can all bind.
5. **The alternative.** What is the best feasible alternative, compared over the same dates and the same constrained resources, with downside cases and a clear threshold for the decision?

Each question corresponds to a term in the contribution equation of the next section: the first fixes the baseline, the second and third define the revenue change, the fourth identifies the costs and displacement elsewhere, and the fifth sets the hurdle. The handbook adds a sixth habit: a recommendation should say what would make it wrong, for example by recommending schedule A unless revised feeder availability would cut its contribution below that of schedule B. A recommendation stated this way can be acted on and tested.

The deliverables follow from the questions: a schedule card for each proposal (both directions, local and UTC times, block times, turns, aircraft and seats by cabin, connections and the implied rotation), a contribution bridge from gross revenue to incremental net-

work contribution, a sensitivity analysis that reports the assumption that would reverse the ranking, and a short decision memo with the recommendation, its conditions and the owners of unresolved inputs (Network Planning Handbook 2026). Chapters 12 and 24 return to these documents.

> **PITFALL**
>
> Answering a network question with a load factor. A route can be 85 percent full and still be the wrong use of its aircraft, because its passengers may have flown on other services of the same airline anyway, because its connecting passengers may displace more valuable ones on feeder flights, or because the same aircraft would earn more elsewhere. Load factor describes a flight; it does not value a decision (Chapter 2).

IN PRACTICE **Network planning in 2026: organization and tools**

At Air Canada, network planning sits on the commercial side: the Chief Commercial Officer's portfolio covers global network development and performance together with yield management and pricing strategy, sales, cargo, and alliance and joint-venture management (Air Canada 2025). Commercial planning systems still follow the classical chain of this chapter: Lufthansa Systems describes its NetLine/Plan product as evaluating O&D-based profitability with linked models of demand and market share, spill and recapture, and cost and revenue (a vendor description) (Lufthansa Systems 2026). Lufthansa Systems has confirmed separately that Air Canada uses its NetLine/Ops ++ product for operations control (Lufthansa Systems 2025a). Research continues on the feedback loops of Figure 1.2: the tentative 2026 program of the AGIFORS scheduling and strategic planning study group lists work by HEC Montréal, KLM and the University of Twente on integrated network planning and scheduling with bank structures, and work by Air Canada on aircraft routing under uncertainty (AGIFORS 2026). Treat all of these as public signals; the tools, data and decision rights at your own desk must be learned on the job.

1.5 Incremental network contribution

A route profit statement reports how a flight performed under the airline's accounting rules. A planning decision needs something else: how much better off the whole airline is with a change than without it. The two answers differ whenever the change moves passengers between services, consumes seats on other flights, or takes an aircraft from another use. The handbook gives a clean comparison equation (Network Planning Handbook 2026), which we write formally.

DEFINITION 1.3 Incremental network contribution

For a feasible schedule s, let $R(s)$ be the passenger revenue the airline retains (net of taxes, partner settlements and other deductions), summed over every journey it carries; $O(s)$ the other revenue the schedule generates, such as cargo and ancillaries, net of their own incremental costs; and $\mathrm{AC}(s)$ the operating costs that the schedule causes and could avoid, summed over every flight it operates. The **network contribution** of s is $\mathrm{NC}(s) = R(s) + O(s) - \mathrm{AC}(s)$, and the **incremental network contribution** of s against a baseline b is

$$\begin{aligned}\Delta\mathrm{NC}(s) = {} & [R(s) + O(s) - \mathrm{AC}(s)] \\ & - [R(b) + O(b) - \mathrm{AC}(b)].\end{aligned} \tag{1.1}$$

Incremental fixed costs (station opening, launch marketing) and longer-run capital requirements are then deducted as the decision horizon requires.

Three features of the definition do the work. Revenue is counted by **journey**, not by leg, so a connecting passenger's fare is counted once. Costs are **avoidable** costs at the horizon of the decision: in a seasonal plan, aircraft ownership is largely fixed but aircraft time is scarce; over a fleet plan, almost everything is variable. Belobaba's route-evaluation lecture makes the same point with a table of which profit measure suits which decision horizon, from variable profit with network contribution for short-term scheduling to fully allocated profit for judging the viability of a hub, on the grounds that over a long enough horizon every cost is variable (Belobaba 2006i). And both terms are evaluated **across the whole network**, which is what makes the quantity hard to estimate and easy to state.

For a single new flight added to an existing network, (1.1) decomposes naturally. Split the journeys of the new schedule into those that use the new flight and those that do not, and compare each group with the baseline:

$$\begin{aligned}\Delta\mathrm{NC} = {} & \left[\textstyle\sum_j f_j\, n_j^{\mathrm{new}} + \Delta O - \mathrm{AC}_{\mathrm{new}}\right] \\ & - \Delta\mathrm{AC}_{\mathrm{other}} - D_{\mathrm{other}} - \textstyle\sum_j f_j\, n_j^{\mathrm{moved}}.\end{aligned} \tag{1.2}$$

Here n_j^{new} is the number of journeys of fare product j, at fare f_j, carried on the new flight, ΔO the other revenue it adds and $\mathrm{AC}_{\mathrm{new}}$ its avoidable cost; $\Delta\mathrm{AC}_{\mathrm{other}}$ is the extra cost of carrying its connecting passengers on other flights, D_{other} the contribution of the passengers

they displace there, and n_j^{moved} the journeys that switch to the new flight from other services of the same airline, whose revenue is not new. The bracket is what a route profit statement measures; the three deductions are what turn it into a network decision.

Two consequences of the definition prevent the most common errors.

RESULT 1.2 Prorates move profit between legs, not into the network

Suppose the fare f_j of each journey of product j is allocated to the legs l of its itinerary with shares $\omega_{jl} \geq 0$ that sum to one over the itinerary. Then the allocated leg revenues add up to the network revenue for every choice of shares, so any allocation rule leaves $\mathrm{NC}(s)$ and $\Delta\mathrm{NC}(s)$ unchanged, while the profit reported for an individual leg depends on the rule.

Proof. Summing allocated revenue over legs and exchanging the order of summation, $\sum_l \sum_j \omega_{jl} f_j n_j = \sum_j f_j n_j \sum_l \omega_{jl} = \sum_j f_j n_j$, since the shares of each journey sum to one. Costs are incurred by flights and need no allocation. ∎

The prorate debate (mileage-based, fare-ratio based, or a joint-venture formula) therefore matters for reporting and for incentives, but not for the network decision, as long as the analysis adds the same quantities back together. What breaks the invariance is double counting: allocating a share of a connecting fare to the new leg and also counting the full fare on the feeder, or adding a separate "feed value" to a simulation that already includes connecting traffic (Network Planning Handbook 2026).

RESULT 1.3 Charge for a scarce aircraft once

Let A be a set of mutually exclusive, feasible uses of the same scarce resource (for example one aircraft over a season), each evaluated against the same baseline. Choosing the use with the largest $\Delta\mathrm{NC}$ is equivalent to accepting a candidate s if and only if

$$\Delta\mathrm{NC}(s) - \max_{a \in A,\, a \neq s} \Delta\mathrm{NC}(a) \geq 0,$$

that is, to charging s the value of the best alternative as an opportunity cost. Charging that opportunity cost and then also comparing the result with the alternative counts the alternative twice.

Proof. The two rules select the same element of A: s maximizes $\Delta\mathrm{NC}$ over A exactly when its value is at least that of every other element, which is the displayed condition. The double-counted rule accepts s only if $\Delta\mathrm{NC}(s) - \Delta\mathrm{NC}(a) \geq \Delta\mathrm{NC}(a)$, that is, if $\Delta\mathrm{NC}(s) \geq 2\,\Delta\mathrm{NC}(a)$, which rejects candidates that are strictly better than the alternative whenever $\Delta\mathrm{NC}(a) > 0$. ∎

In practice the handbook allows either approach, explicit comparison of alternatives or a shadow price on aircraft time in a screening model, provided the analyst says how the price was estimated and does not use both at once (Network Planning Handbook 2026). A shadow price per block hour is the same idea in marginal form: it charges each candidate the price times its block hours, and it is consistent with Result 1.3 when that charge equals the value of what those hours would otherwise do. Chapters 14 and 20 show how linear programming produces such shadow prices.

1.6 Exercise A: is the full-looking route the best use?

The handbook's first exercise puts the definition to work on invented numbers (Network Planning Handbook 2026). It is worth working slowly, because every step corresponds to one of the five questions.

WORKED EXAMPLE 1.2 **Exercise A from the handbook**

A candidate flight, evaluated for one departure (all amounts in CAD, all invented for teaching), carries 110 local customers paying CAD 220 each and 40 connecting customers paying CAD 450 each for their whole itineraries. Cargo and ancillaries contribute CAD 3,000 net. The flight's avoidable operating cost is CAD 32,000. Carrying the 40 connecting customers on existing feeder flights adds CAD 2,000 of operating cost and displaces CAD 4,000 of other contribution on those flights. Passengers who switch to the new flight from other services of the same airline take CAD 3,500 of revenue with them. A feasible alternative use of the same aircraft, evaluated independently against the same baseline, is worth CAD 4,500.

Step 1: revenue by journey. Ticket revenue is $110 \times 220 = 24{,}200$ from local passengers plus $40 \times 450 = 18{,}000$ from connecting passengers, CAD 42,200 in all. The connecting fares are counted in full here, once, because each covers a whole journey that exists only if the new flight operates.

Step 2: the flight itself. Adding the CAD 3,000 of cargo and ancillaries and deducting the CAD 32,000 of avoidable cost leaves CAD 13,200. This is the flight's own result when it keeps every connecting fare, the number a leg profit statement with full-fare allocation would show.

Step 3: effects on other flights. The connecting customers ride feeder flights too. Their extra cost (CAD 2,000) and the contribution they displace (CAD 4,000) reduce the total to CAD 7,200.

Step 4: diversion. CAD 3,500 of the revenue counted in Step 1 would have been earned anyway on other services; deducting it gives

$$\Delta \mathrm{NC} = 42{,}200 + 3{,}000 - 32{,}000 - 2{,}000 - 4{,}000 - 3{,}500 = 3{,}700.$$

Step 5: the alternative. The same aircraft could earn CAD 4,500 elsewhere, so the candidate is CAD 800 worse than the alternative. Under Result 1.3, its opportunity-cost-adjusted value is $3{,}700 - 4{,}500 = -800$, and the recommendation is to use the aircraft elsewhere.

Figure 1.3 draws the same bridge. The flight carries 150 customers; on a 169-seat Laurentian 737-8 that would be an 89 per-

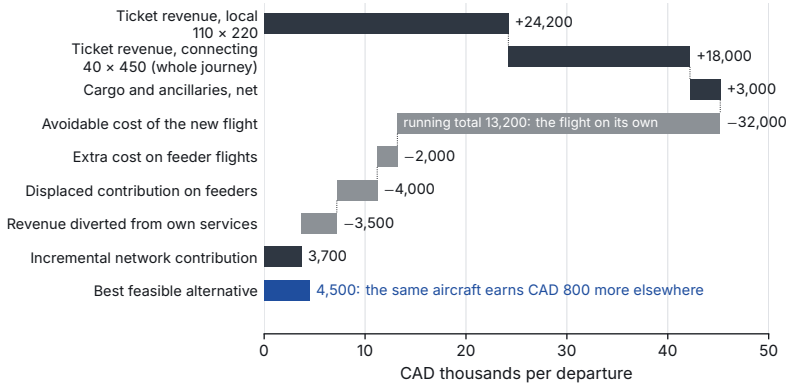


Figure 1.3 Exercise A as a contribution bridge, per departure. The flight's own result, with every connecting fare credited to it, is CAD 13,200; the effects on feeder flights and the diversion of the airline's own customers reduce the network gain to CAD 3,700, below the CAD 4,500 the same aircraft earns in its best alternative use. Numbers are the handbook's, invented for teaching.

cent load factor, and the flight looks profitable on its own (CAD 13,200). Neither fact answers the question.

The most useful output of the exercise is not the CAD 800 gap but the list of assumptions that would close it, because those are the inputs to check first. With every other number fixed, the candidate would match the alternative if diversion were CAD 2,700 instead of CAD 3,500, if displacement on the feeders were CAD 3,200 instead of CAD 4,000, or if each connecting customer paid CAD 20 more (CAD 800 across 40 customers). Each threshold assumes that nothing else moves, which is rarely true: a higher connecting fare would lose some of those customers and change the mix, so CAD 20 is a threshold for discussion, not a pricing recommendation. The handbook also suggests a test that costs nothing in aircraft: whether a different departure time would raise the candidate's contribution without increasing the number of aircraft it needs (Network Planning Handbook 2026).

Result 1.2 can be seen at work in the same numbers. If the 40 connecting fares were prorated, say 60 percent to the candidate and 40 percent to the feeders, the candidate's leg statement would show $24{,}200 + 0.6 \times 18{,}000 + 3{,}000 - 32{,}000 = 6{,}000$ and the feeders would gain $0.4 \times 18{,}000 - 2{,}000 - 4{,}000 = 1{,}200$; after the CAD 3,500 of diversion the network total is again CAD 3,700. The allocation changed the story each leg tells, not the decision.

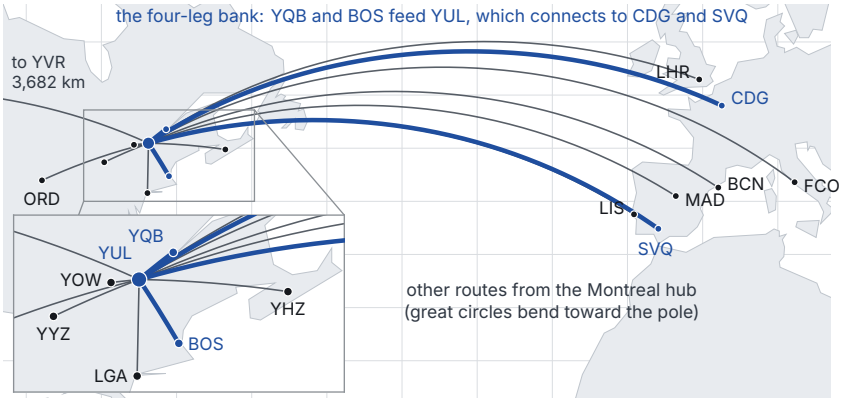


Figure 1.4 Laurentian's network as used in this book: great-circle routes from the Montreal hub to the airports of the running case, drawn on an equidistant cylindrical projection with standard parallel 45 degrees north, on which great circles bend toward the pole. The four-leg bank is in cobalt; the inset enlarges the hub and its North American spokes. Vancouver lies off the map to the west. Coastlines from Natural Earth (public domain).

> **PITFALL**
>
> Comparing with doing nothing. A positive incremental contribution says only that the change beats the baseline. When the change uses a scarce resource, the comparison that matters is with the best feasible alternative use of that resource, evaluated on the same dates and against the same baseline.

1.7 Laurentian, the running case

Laurentian is a fictional airline with its hub at Montreal-Trudeau. Its geography, airports and aircraft types are real; its demand, fares and costs are invented, chosen to be plausible and to make each model's behaviour visible, and none of them is an Air Canada figure. The fleet has four types, the 137-seat A220-300, the 169-seat 737-8, the 182-seat A321XLR and the 298-seat 787-9, with the cost model described in the preface and developed in Chapter 3. Figure 1.4 shows the network used in the book's examples.

Two pieces of the network carry most of the book's examples. The first is the route **YUL–SVQ**, flown with the A321XLR on the evening pattern of Example 1.1; it is used for spill (Chapter 9), route evaluation (Chapter 12) and single-flight revenue management (Chapter 18). The second is a **bank** at Montreal: two feeder flights, from Quebec City (YQB) and Boston (BOS), arrive in time to connect to two transatlantic departures, to Paris (CDG) and Seville. Its four

legs (Table 1.2) and sixteen fare products, a flexible and a restricted fare in each of eight origin-destination markets, are small enough to solve by hand and rich enough to show the network effects of Chapters 14 and 20.

Table 1.2 The four legs of Laurentian's bank. BH: block hours; km: great-circle distance; Demand: mean unconstrained demand per departure, summed over the O&D markets that use the leg; DF: demand factor, demand divided by seats; Conn.: the share of demand that continues on, or arrives from, another leg of the bank. All demand figures are illustrative.

Leg	Aircraft	Seats	BH	km	Demand	DF	Conn.
YQB–YUL	A220-300	137	0.9	233	141	1.03	48%
BOS–YUL	A220-300	137	1.4	409	153	1.12	54%
YUL–CDG	787-9	298	6.9	5,524	317	1.06	32%
YUL–SVQ	A321XLR	182	7.0	5,556	161	0.88	30%

WORKED EXAMPLE 1.3 The bank at a glance

Add up the mean demand of the sixteen fare products. The eight markets generate 621 journeys per bank: 470 local journeys on a single leg and 151 connecting journeys (24.3 percent) that use two legs each. The legs therefore see $470+2\times151 = 772$ boardings, 1.24 per journey, against 754 seats in total. Three of the four legs have more mean demand than seats (Figure 1.5), so spill, the choice of which passengers to carry, and the value of extra seats will matter on every one of them.

At mean demand and the fares of the case, the bank's journeys are worth CAD 377,639. Connecting journeys are 24.3 percent of journeys but 31.6 percent of that revenue, because their average fare is CAD 790 against CAD 550 for local journeys. Per seat used, the comparison reverses. A restricted Boston to Paris journey pays CAD 599 for two seats, CAD 299.50 per leg, while a restricted Montreal to Paris journey pays CAD 549 for one seat on the same transatlantic flight.

Which passenger should get the last seat on YUL–CDG? The answer depends on what the Boston passenger's seat on the feeder would otherwise earn, and on whether either passenger would fly Laurentian anyway on another day. That is the displacement problem of Chapters 12 and 20, and the reason the incremental contribution of Section 1.5 is measured across the network.

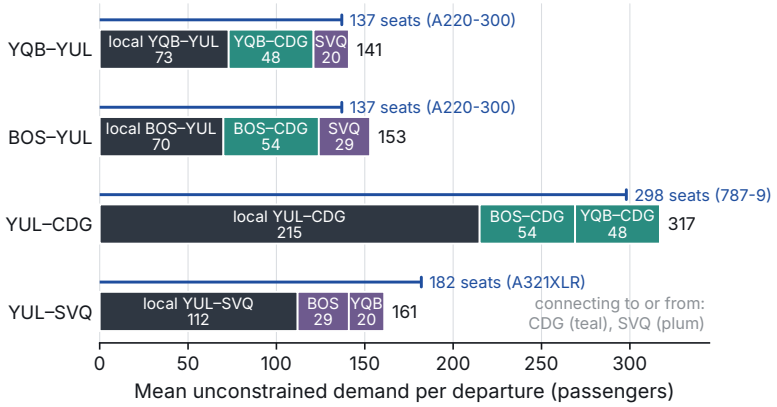


Figure 1.5 What each leg of Laurentian's bank carries: mean unconstrained demand per departure by O&D market, local in slate and connecting in colour, against the seats of the assigned aircraft (cobalt lines). Connecting traffic is coloured by the long-haul flight it uses. All figures are illustrative.

1.8 Air Canada in October 2026

The facts in this section are public, dated and cited; they describe the context of the job, not its internal workings. They were collected on 27 October 2026, two days before Air Canada's third-quarter results, scheduled for release ahead of a call on 29 October (Network Planning Handbook 2026). Check that release before relying on anything below.

The company. Air Canada describes itself as the largest provider of scheduled passenger services in the Canadian domestic market, the Canada-United States transborder market and the international market to and from Canada. Regional flying under the Air Canada Express brand, bought from Jazz Aviation under a capacity purchase agreement and from PAL Airlines under a commercial agreement, provides feed to the mainline and Rouge networks; Air Canada Rouge is the leisure carrier; Air Canada is a founding member of Star Alliance and owns the Aeroplan loyalty program; Air Canada Cargo also flies dedicated 767 freighters (Air Canada 2026e). Its operating fleet, including aircraft flown by regional partners, numbered 357 at 30 June 2026 (Air Canada 2026e).

Hubs and network. Air Canada develops three international hubs, at Toronto Pearson, Montreal-Trudeau and Vancouver (Air Canada 2025). The handbook suggests treating the familiar shorthand (Montreal for Europe, Toronto as the largest mixed network, Vancouver for the Pacific) as a hypothesis to test against actual schedules rather than as a fact (Network Planning Handbook 2026).

Traffic between two other countries that connects over these hubs, called sixth-freedom traffic, is a stated growth theme: sixth-freedom revenue grew 9 percent year over year in the second quarter of 2026, led by the US-Atlantic and US-Pacific corridors, and premium cabins earned about 30 percent of passenger revenue (Air Canada 2026e).

A difficult 2026. Fuel prices rose sharply in the first half of 2026. Air Canada suspended its full-year guidance on 30 April (The Canadian Press 2026) and reported an operating loss of CAD 215 million for the second quarter after CAD 388 million of labour-related and legal charges (Air Canada 2026e). It reinstated guidance in August, with capacity growth of 2.25 to 3.25 percent in available seat miles for the year (Air Canada 2026d). Chapter 2 reads the quarter's operating statistics in detail.

Fleet transition. The second-quarter report shows a fleet in motion: 737 MAX 8 aircraft moving from the mainline to Rouge, A220-300 deliveries continuing (46 of 65 firm orders delivered by 30 June 2026), the first 787-10s due in 2026, and A350-1000s from 2030 (Air Canada 2026e). The A321XLR is the planning story of the year. Air Canada received its first on 24 April 2026 (Air Canada 2026a); the type entered scheduled service on 9 June 2026 in a 182-seat layout with 14 lie-flat Signature Class seats and 168 economy seats (Network Planning Handbook 2026; Air Canada 2026g). The program covers 30 aircraft, with 2 delivered at 30 June 2026, 9 planned by the end of 2026, 19 by the end of 2027 and the last in 2029 (Air Canada 2026e). Worldwide, airlines had taken delivery of about 50 A321XLRs by October 2026, in layouts ranging from 138 to 239 seats (Cirium 2026b). An announced program is not the number of aircraft available to any particular desk in any particular month (Network Planning Handbook 2026).

Montreal to Seville. Announced on 14 October 2026, the route starts on 8 September 2027 with three weekly A321XLR flights, rising to four weekly from 1 October 2027, departing Montreal at 20:30; Air Canada describes it as the only non-stop service between North America and Seville (Air Canada 2026g). The published plan includes autumn-to-spring operating periods, which makes the route a live example of an aircraft's capability meeting seasonal demand and hub connections (Network Planning Handbook 2026). It is the model for Laurentian's YUL–SVQ case, whose economics are invented.

The United transborder agreement. On 16 October 2026 Air Canada and United announced new transborder routes for summer 2027, including Montreal to Indianapolis, and an agreement to

renew and expand their transborder joint business agreement, subject to regulatory approval (Air Canada 2026h). Transport Canada opened a public consultation on the expansion, which would bring into the agreement city pairs that had so far been carved out, among them Montreal to Chicago and Montreal to Washington (Travelweek 2026a). For a Montreal planner the agreement matters because it shapes the transborder feed into the hub; the handbook's advice is to keep existing cooperation separate from changes still awaiting approval (Network Planning Handbook 2026).

> **PITFALL**
>
> Reading company-wide signals as route facts. Premium revenue shares, sixth-freedom growth and a fleet program describe the airline as a whole. None of them establishes the profitability of any route, the demand for any proposed service, or the aircraft available to a given schedule.

1.9 Summary

- Airline planning is a sequence of decisions on different horizons: fleet planning years ahead, route planning and seasonal frequency plans a year or more ahead, timetables and fleet assignment two to six months ahead, crew schedules monthly, revenue management throughout the booking period, and operations control on the day.
- Each stage takes earlier decisions as fixed, and the cost of changing a decision rises as departure approaches; a published schedule creates commitments.
- Airlines decompose planning into schedule design, fleet assignment, aircraft routing and crew scheduling because the joint problem is too large. Sequential planning yields feasible plans whose shortfall from the best plan is the value of integration; feedback loops recover part of it.
- Planning is hard because seats are perishable, legs serve many markets, demand and costs are uncertain, aircraft are indivisible, competitors react and regulation constrains access.
- The analyst's core decision is the feasible schedule and aircraft allocation with the largest incremental network value, measured against the best realistic alternative; the handbook's five questions (counterfactual, journeys, increment, binding resources, alternative) structure every recommendation.

- Incremental network contribution counts revenue by journey and avoidable cost by flight, across the whole network, against a dated baseline.
- Revenue prorates change the profit reported for a leg but not the network total; scarce aircraft must be charged for once, either by comparing alternatives or through a shadow price.
- In Exercise A, a 150-passenger flight that looks profitable on its own adds CAD 3,700 to the network and loses to an alternative worth CAD 4,500. Load factor did not answer the question.

1.10 Further reading

Belobaba's introduction to the planning process (2006c) and his lecture on schedule development (2006f) are short and give the vocabulary of this chapter; his route-evaluation lecture (2006i) develops the incremental logic further and is the starting point of Chapter 12. Barnhart's course notes (2003) treat each stage of the decomposition as an optimization problem. The handbook (2026) is the source of Exercise A, the decision levels and the five questions; read its framework section before your first week. Barnhart and Cohn (2004) survey airline schedule planning and the case for integration, and Lohatepanont and Barnhart (2004) show what integrating schedule design and fleet assignment buys. For a textbook treatment of the planning process, see the chapter on the airline planning process in Belobaba, Odoni and Barnhart (2015, ch. 7).

CHAPTER

2

Measuring an Airline

IN THIS CHAPTER

- The four objects that airline data mix up: aircraft movements, flight legs, passenger itineraries and tickets, and why demand lives in origin-destination markets while supply lives on legs.
- Passengers per day each way, and what it does and does not say about the load of a new flight.
- Capacity and traffic in seat-kilometres and seat-miles, and why the average of load factors is the wrong load factor.
- Yield, unit revenue and unit cost, the identity that ties them together, and its use as a reconciliation check.
- Why yield and unit cost fall with distance, and what a stage-length adjustment can and cannot do.
- The break-even load factor: derivation, a more exact version, and its limits.
- A full reading of Air Canada's second-quarter 2026 operating statistics, including a figure that does not reconcile.

DRAWS ON Belobaba, Introduction to the airline planning process; Airline economics review; Operating costs and productivity measures (MIT 16.75J, 2006); handbook; Air Canada Q2 2026 MD&A

Airlines describe themselves in ratios. Air Canada's report for the second quarter of 2026 says that unit passenger revenue rose 10.5 percent on capacity up 0.3 percent, that the load factor gained 2.6 points and that unit cost rose 23.9 percent (Air Canada 2026e); an investor presentation compares cost per seat-mile with a competitor's; a route review reports that a flight ran at 84 percent. Each of these numbers is a ratio with a numerator, a denominator and a scope, and each hides weights. A network planning analyst needs to read them fluently, recompute them from their parts, and know which questions they can answer.

This chapter builds the vocabulary from the bottom up. It starts with the objects that airline data count (movements, legs, itineraries,

tickets) and the distinction between origin-destination demand and leg supply, which runs through the whole book. It then defines capacity, traffic, load factor, yield, unit revenue and unit cost, derives the identities that link them, and explains why both yield and unit cost fall with distance. The break-even load factor closes the analytical part. The last section reads Air Canada's report for the second quarter of 2026 as a worked example, with real numbers, every identity checked, and a figure that does not add up.

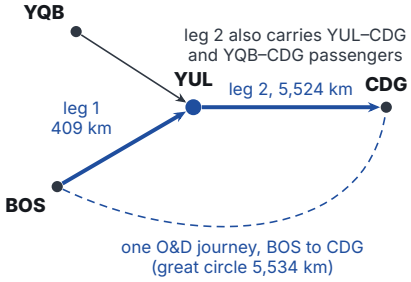
Measures describe an airline; they do not decide anything. Chapter 1 argued that planning decisions should be judged by incremental network contribution, and nothing in this chapter changes that. But every contribution estimate is assembled from these measures, every forecast is reconciled against them, and every conversation with finance, revenue management or senior management uses them. Chapter 3 takes the cost side further.

2.1 Four objects: movements, legs, itineraries and tickets

The handbook names the analyst's most important data skill as keeping four objects separate: an aircraft movement, a flight leg, a passenger itinerary and a ticket (Network Planning Handbook 2026). Summing rows without knowing which object a row represents is an easy way to double count capacity or demand.

> **DEFINITION 2.1 The objects airline data count**
>
> An **aircraft movement** is one takeoff or one landing of one aircraft. A **flight leg** (also segment or sector) is one non-stop operation from an origin to a destination with scheduled times; it is what consumes an aircraft, a crew and a slot. A **flight** is one or more consecutive legs flown under a single flight number. An **origin-destination (O&D) market** is a directional pair of points between which passengers want to travel, and a **journey** is one passenger's trip from origin to destination in one direction. An **itinerary** (or path) is the sequence of legs a journey uses. A **ticket** is the commercial document that may cover several journeys, each leg of which is a **coupon**.

Demand is defined by O&D market, but supply is produced as legs (Belobaba 2006d). A single leg supplies seats jointly to every market whose itineraries use it, and a single market can be served by many itineraries over different hubs and airlines. Belobaba draws



One passenger, BOS to CDG via YUL	
O&D journeys	1
itineraries (paths)	1
legs flown, seats used	2
flight coupons	2
passengers carried (IATA, by flight number)	2
RPK, at flown distance	5,932
If the ticket includes the return	
tickets	1
O&D journeys	2
flight coupons	4

Figure 2.1 One passenger travelling from Boston to Paris over Montreal, and how each unit of measurement counts her. The dashed curve is the O&D journey, measured along the great circle; the cobalt arrows are the two legs she actually flies. Distances from the case data.

the consequences: there is no well-defined supply of seats to any one market, so whether a market is in equilibrium, or whether the airline's service in it is profitable, cannot be answered without assumptions about how costs and revenues are allocated (Belobaba 2006d). Chapter 12 returns to the allocation problem; here we need only the accounting.

Write $\mathcal{L}$ for the set of legs and $\mathcal{J}$ for the set of O&D fare products, with $a_{lj} = 1$ if product j uses leg l and 0 otherwise (the leg-ODF incidence used throughout the book). If n_j journeys of product j are carried, the number of passengers on leg l is

$$N_l = \sum_{j \in \mathcal{J}} a_{lj}\, n_j, \qquad \sum_{l \in \mathcal{L}} N_l = \sum_{j \in \mathcal{J}} \Big(\sum_{l} a_{lj} \Big) n_j \;\geq\; \sum_{j \in \mathcal{J}} n_j,$$

with equality only if every journey is non-stop. Leg counts exceed journey counts by exactly the number of extra legs that connecting journeys use.

WORKED EXAMPLE 2.1 Boston to Paris over Montreal

A Laurentian passenger buys a restricted one-way ticket from Boston to Paris at CAD 599 and flies BOS–YUL (409 km) and YUL–CDG (5,524 km), as in Figure 2.1. She is one O&D journey in the BOS–CDG market and one itinerary, but two leg passengers and two coupons, and she occupies one seat on each of two aircraft. Her revenue passenger-kilometres, measured along the legs she flew (408.7 and 5,523.5 km), are 5,932, which is 7.2 percent more than the 5,534 km great-circle distance from Boston to Paris.

Each count answers a different question. The O&D count (one) is what a demand forecast for the Boston-Paris market predicts. The leg counts (one on each leg) are what fill the seats of each flight. Her fare divided by her flown kilometres, 599/5,932 = 10.1 cents per RPK, is what enters a yield statistic; divided by the O&D distance it would be 10.8 cents. A local passenger on YUL–CDG paying the restricted fare of CAD 549 yields 9.9 cents per RPK on the same aircraft. If her ticket included the return trip, it would be one ticket with two journeys and four coupons.

Across Laurentian's whole bank (Section 1.7), mean demand of 621 journeys per bank becomes 772 leg boardings, because the 151 connecting journeys are counted on two legs each. A planner who adds up leg loads and calls the total "passengers" overstates demand by 24 percent.

Public reporting adds a fifth convention. Air Canada counts revenue passengers carried on a flight-number basis, following the IATA definition, rather than by journey or by leg (Air Canada 2026e). A connecting passenger on two flight numbers counts twice; a passenger staying on board through an intermediate stop under one flight number counts once. None of the four objects above is exactly what such a statistic counts, which is why the handbook asks analysts to record, for every table, whether its rows are tickets, passengers, coupons, legs or flights (Network Planning Handbook 2026).

> **PITFALL**
>
> Counting codeshares as flights. A flight sold under several airlines' codes appears several times in schedule data, once per marketing carrier. Capacity counts must use operating flights only; demand counts must not add the same passenger under each code. The handbook lists both checks among the routine audits of any analysis (Network Planning Handbook 2026).

2.2 Passengers per day each way

Market sizes are usually quoted as a daily, directional rate, because flights are daily and directional.

DEFINITION 2.2 Passengers per day each way

The **passengers per day each way** (**PDEW**) of an O&D market over a period of T days is the number of journeys in the market in the period, summed over both directions, divided by $2T$. When the two directions differ, the directional rates (journeys in one direction divided by T) should be reported separately.

The handbook's example: 73,000 journeys in one market over 365 days, both directions together, is $73{,}000/(2 \times 365) = 100$ PDEW, if demand is symmetric (Network Planning Handbook 2026). A common error is to forget one of the two divisions, which overstates the market by a factor of two or of 365.

WORKED EXAMPLE 2.2 From PDEW to passengers on a departure

Take the handbook's market of 100 PDEW and suppose, for illustration, that a new non-stop four times a week could capture 45 percent of it. In one direction the market produces $100 \times 7 = 700$ journeys a week, of which the airline would carry $0.45 \times 700 = 315$.

If every customer were willing to travel on one of the four operating days, each departure would carry $315/4 = 78.75$ passengers. If instead only the customers whose preferred day is an operating day travel, the airline carries $315 \times 4/7 = 180$ journeys a week, 45 per departure. The truth lies between, depending on how flexible customers are about the day, and the handbook calls the difference a missed-day penalty (Network Planning Handbook 2026).

A market of 100 PDEW therefore does not mean 100 passengers on each new flight. Before it says anything about a departure, the rate must pass through seasonality and day-of-week patterns, the itinerary choices of customers among all competing services, the days actually operated, capacity limits and any stimulation of new demand by the new service. Chapters 4 and 8 build each of these layers.

2.3 Capacity, traffic and load factor

Airline output is measured in seat-distance, because a seat flown 5,000 km is a different product from a seat flown 300 km.

DEFINITION 2.3 Capacity, traffic and load factor

For a set of flight legs l with C_l seats offered, N_l revenue passengers carried and length d_l, **available seat-kilometres** and **revenue passenger-kilometres** are

$$\text{ASK} = \sum_l C_l\, d_l, \qquad \text{RPK} = \sum_l N_l\, d_l,$$

with d_l in kilometres; with d_l in statute miles they are **available seat-miles (ASM)** and **revenue passenger-miles (RPM)**. The **load factor** is

$$\text{LF} = \text{RPK}/\text{ASK} = \text{RPM}/\text{ASM}.$$

For a single leg it reduces to passengers divided by seats.

Belobaba's example makes the arithmetic concrete: a 200-seat aircraft flying 1,000 miles with 140 passengers produces 200,000 ASM and 140,000 RPM, a load factor of 70 percent (Belobaba 2006c). One mile is exactly 1.609344 km, so ASK = 1.609344× ASM and the same holds for traffic. Load factor is unaffected by the unit, but every measure expressed per unit of distance is not: a cost of 7.5 cents per ASM is $7.5/1.609344 = 4.66$ cents per ASK. Air Canada reports in miles (Air Canada 2026e), while IATA's industry statistics are in kilometres (IATA 2026d). The handbook's advice is never to mix the two within one calculation (Network Planning Handbook 2026).

The load factor of a network is a ratio of sums, not an average of ratios, and the difference matters.

RESULT 2.1 The system load factor is an ASK-weighted average

For legs $l = 1, \dots, n$ with load factors $\mathrm{LF}_l = N_l / C_l$, the system load factor is

$$\mathrm{LF} = \sum_l w_l \, \mathrm{LF}_l, \qquad w_l = \frac{C_l \, d_l}{\sum_k C_k \, d_k}.$$

The simple average $\overline{\mathrm{LF}} = \frac{1}{n} \sum_l \mathrm{LF}_l$ differs from it by

$$\mathrm{LF} - \overline{\mathrm{LF}} = n \, \mathrm{Cov}(w, \mathrm{LF}),$$

where Cov is the covariance across the n legs (with divisor n) between the weights and the load factors.

Proof. $\mathrm{RPK} = \sum_l N_l d_l = \sum_l \mathrm{LF}_l \, C_l d_l$, so dividing by ASK gives the weighted average. For the second claim, the weights average $1/n$, so $\sum_l w_l \mathrm{LF}_l - \frac{1}{n} \sum_l \mathrm{LF}_l = \sum_l (w_l - \frac{1}{n})(\mathrm{LF}_l - \overline{\mathrm{LF}})$, which is n times the covariance. ∎

The simple average is biased whenever the legs with most capacity have systematically different load factors from the others. In many networks the long-haul legs, which carry most of the ASKs, are also the fullest, so the simple average understates the system figure. A third ratio, passengers divided by seats summed over legs, weights each leg by its seats and ignores distance; in the example below it sits in between.

WORKED EXAMPLE 2.3 One bank, three load factors

On one illustrative day Laurentian's bank carries 82 passengers on YQB–YUL (137 seats, 233 km), 97 on BOS–YUL (137 seats, 409 km), 280 on YUL–CDG (298 seats, 5,524 km) and 150 on YUL–SVQ (182 seats, 5,556 km). The leg load factors are 59.9, 70.8, 94.0 and 82.4 percent.

- The simple average of the four load factors is 76.8 percent.
- Passengers divided by seats is $609/754 = 80.8$ percent.
- RPK divided by ASK is 88.8 percent.

The ASK weights explain the gap: the two transatlantic legs produce 60.0 and 36.8 percent of the day's ASKs, the two feeders 1.2 and 2.0 percent. Figure 2.2 draws each leg as a bar whose

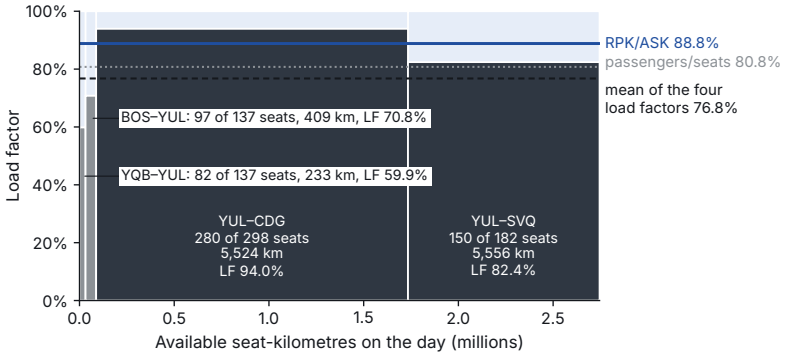


Figure 2.2 Load factors on Laurentian's bank on one illustrative day. Each bar's width is the leg's available seat-kilometres and its height the leg's load factor, so the network load factor, RPK/ASK, is the area of the bars divided by the total width. The two feeder legs contribute almost nothing to it.

width is its ASK and whose height is its load factor, so the system load factor is the area of the bars divided by the total width; the feeders barely register. As Result 2.1 predicts, $n\,\mathrm{Cov}(w, \mathrm{LF}) = 4 \times 0.0302 = 0.121$, exactly the gap of 12.1 points between the unrounded 88.84 and 76.76 percent.

None of the three numbers is wrong; they answer different questions. RPK/ASK measures how much of the capacity produced was sold and is the figure that enters unit revenue. Passengers per seat describes boardings. The simple average describes nothing in particular.

PITFALL

Averaging load factors. Adding the load factors of flights of different sizes and lengths and dividing by their number gives a number that matches neither the network's capacity utilization nor its boardings. Always recompute from RPK and ASK, or from passengers and seats if boardings are what you need, and say which one you used (Network Planning Handbook 2026).

2.4 Yield, unit revenue and unit cost

Revenue and cost enter through three more ratios. Write p for yield and ρ for revenue other than passenger revenue, per ASK.

DEFINITION 2.4 Yield, unit revenue and unit cost

For a given period and scope, the **yield** is passenger revenue per revenue passenger-kilometre, p = passenger revenue/RPK. **Passenger revenue per ASK (PRASK)** is passenger revenue divided by ASK. **Revenue per ASK (RASK)**, also called total revenue per ASK, is total operating revenue (passenger, cargo and other) divided by ASK, so RASK = PRASK + ρ. **Cost per ASK (CASK)** is operating expense divided by ASK. The mile-based versions are written PRASM, RASM (or TRASM) and CASM.

Definitions of revenue and cost scope vary between airlines and reports, so a ratio is meaningful only with its scope stated: which revenues (cargo, ancillaries, loyalty, tour packages), which costs, which carriers (regional partners included or not). Air Canada, for example, reports an **adjusted CASM** that excludes fuel, the cost of ground packages sold by its tour operator, freighter costs and items it considers unusual; it is a non-GAAP measure without a standardized meaning (Air Canada 2026e). The handbook flags two common confusions: calling PRASK "yield", and comparing RASK figures built on different revenue streams (Network Planning Handbook 2026).

The ratios are tied together by identities that follow from their definitions (Figure 2.3).

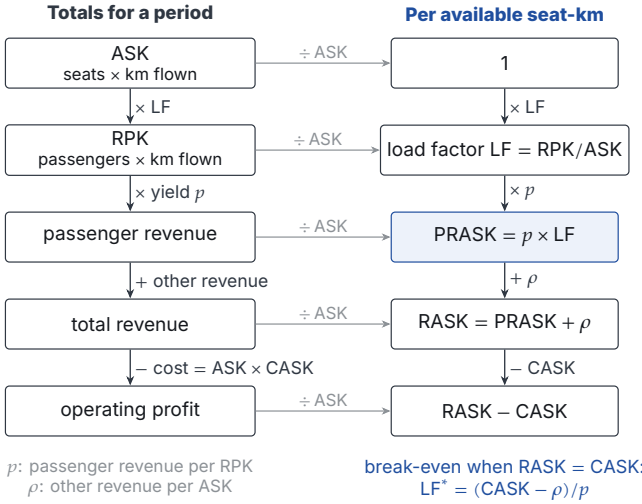


Figure 2.3 The cascade of airline measures. Totals (left) become unit measures (right) on division by ASK. Passenger revenue per ASK is the product of yield and load factor, and the operating margin per ASK is unit revenue minus unit cost; setting it to zero gives the break-even load factor of Section 2.6.

RESULT 2.2 Identities linking the unit measures

For a common scope and period,

$$\text{PRASK} = p \times \text{LF}, \qquad \text{RASK} = p\,\text{LF} + \rho, \tag{2.1}$$

and operating profit is

$$\Pi = \text{RPK} \cdot p + \text{ASK} \cdot \rho - \text{ASK} \cdot \text{CASK} = \text{ASK}\,(\text{RASK} - \text{CASK}).$$

Between two periods, with g denoting the proportional change of each quantity,

$$\ln(1+g_{\text{PRASK}}) = \ln(1+g_p) + \ln(1+g_{\text{RPK}}) - \ln(1+g_{\text{ASK}}). \tag{2.2}$$

Proof. Multiplying and dividing by RPK,

$$\text{PRASK} = \frac{\text{passenger revenue}}{\text{ASK}} = \frac{\text{passenger revenue}}{\text{RPK}} \cdot \frac{\text{RPK}}{\text{ASK}} = p \cdot \text{LF}.$$

Adding other revenue per ASK gives RASK, and multiplying RASK − CASK by ASK gives revenue minus cost. Taking loga-

rithms of PRASK $= p \cdot$ RPK/ASK in both periods and subtracting gives (2.2). ■

Belobaba writes the same profit identity as RPM times yield minus ASM times unit cost and warns against reading any single term as a measure of success: a high yield is no use if the load factor is too low, a low unit cost is of little value if revenues are weak, and a high load factor may simply reflect very low fares (Belobaba 2006d). The handbook adds that the identity is a reconciliation check, not a causal model (Network Planning Handbook 2026). In its example, a yield of CAD 0.20 per passenger-mile and a load factor of 0.85 give a PRASM of 17 cents. The check is useful precisely because the three quantities are usually reported separately and rounded: if they do not multiply out, one of them was computed on a different scope, or is wrong.

Belobaba's 200-seat example completes the set (historical, in dollars): with revenue of 16,000 dollars and operating expense of 15,000 dollars, yield is 11.4 cents per RPM, unit cost 7.5 cents per ASM and unit revenue 8.0 cents per ASM (Belobaba 2006c). The identity holds, $11.43 \times 0.70 = 8.0$, and the flight earns $200{,}000 \times (8.0 - 7.5)$ cents, or 1,000 dollars. In kilometres the same flight has a yield of 7.10 cents per RPK, a CASK of 4.66 cents and a RASK of 4.97 cents. Every number changed except the load factor and the profit.

> **PITFALL**
>
> Yield per passenger versus yield per kilometre. Average fare (revenue per passenger) and yield (revenue per passenger-kilometre) move differently whenever the mix of trip lengths changes. A network that adds long flights can raise its average fare while its yield falls, without any change in pricing.

2.5 Stage length

Distance shapes both sides of the unit economics. Two averages describe it. The **average stage length** is the average non-stop leg distance, defined by Belobaba as aircraft miles flown divided by departures (Belobaba 2006d); Air Canada instead divides total ASMs by total seats dispatched, which weights each leg by its seats (Air Canada 2026e). The **average passenger trip length** is RPK divided by passengers, which is typically longer than the stage length be-

cause some passengers take more than one flight (Belobaba 2006d). Capacity then decomposes exactly into three factors:

$$\text{ASK} = \text{departures} \times \frac{\text{seats}}{\text{departure}} \times \frac{\text{ASK}}{\text{seats}},$$

frequency times average gauge times seat-weighted stage length, so the growth of capacity is the sum of the growth rates of the three, in logarithms.

Yield falls with distance because fares rise less than proportionally with distance. A simple model writes the fare of a journey of length d as $f(d) = a + b\,d$, with $a > 0$ representing costs and value that do not scale with distance (airport and terminal processes, the fixed effort of travelling) and b the price per additional kilometre. Then the yield is

$$p(d) = \frac{f(d)}{d} = b + \frac{a}{d},$$

which falls steeply at short distances and flattens towards b. Belobaba's 2006 chart of yield against distance in the 50 largest O&D markets shows exactly this shape, from 20 to over 60 cents per mile in markets shorter than 300 miles to about 7 to 12 cents beyond 1,500 miles (Belobaba 2006d). Fitting the model by least squares to the fares implied by the chart gives $f(d) \approx 64 + 0.065\,d$ dollars, with d in miles, the curve in the left panel of Figure 2.4; the dispersion around it reflects differences in competition and in the mix of business and leisure travel between markets. The numbers are historical; the shape is general.

Unit cost falls with distance for a parallel reason. Write the cost of one departure of an aircraft with C seats flying d kilometres as $\text{TC}(d) = F + v\,d$, where F collects the costs incurred once per departure (landing and terminal charges, ground handling, the fuel and time of taxi, takeoff and climb, maintenance driven by cycles) and v the costs that grow with distance flown. Then

$$\text{CASK}(d) = \frac{\text{TC}(d)}{C\,d} = \frac{v}{C} + \frac{F}{C\,d}.$$

Chapter 3 builds F and the cost per kilometre from block-hour and departure costs and shows why aircraft size matters as much as distance. Here we need only the shape and what it implies for comparisons.

RESULT 2.3 The stage-length elasticity of unit cost

Under $\mathrm{TC}(d) = F + v\,d$, the elasticity of CASK with respect to stage length is

$$\eta(d) = \frac{d\,\mathrm{CASK}'(d)}{\mathrm{CASK}(d)} = -\frac{F}{F + v\,d},$$

minus the share of distance-independent costs in the cost of a departure. It is close to -1 for very short legs, rises towards 0 as legs lengthen, and is therefore not a constant.

Proof. Differentiating, $\mathrm{CASK}'(d) = -F/(Cd^2)$, so

$$\frac{d\,\mathrm{CASK}'(d)}{\mathrm{CASK}(d)} = -\frac{F/(Cd)}{(v + F/d)/C} = -\frac{F}{F + v\,d}.$$

■

A **stage-length adjustment** restates one airline's (or one route's) CASK at a reference stage length d_0, so that unit costs of operations with different average distances can be compared. If a cost model is available, the adjustment moves along the model's curve:

$$\mathrm{CASK}_{\mathrm{adj}} = \mathrm{CASK}(d)\,\frac{v + F/d_0}{v + F/d}. \tag{2.3}$$

A simpler alternative when no cost model is at hand is a constant-elasticity rule, $\mathrm{CASK}_{\mathrm{adj}} = \mathrm{CASK}(d)\,(d/d_0)^{\theta}$, with θ estimated, for example, as minus the slope of a regression of log CASK on log stage length across a sample of airlines. By Result 2.3 such a rule is only a local approximation, and the right exponent for adjusting from d to d_0 is the arc elasticity between the two, which depends on both.

WORKED EXAMPLE 2.4 Which carrier is cheaper?

Two narrowbody operations with the same aircraft are compared in the right panel of Figure 2.4, using the illustrative cost model $F = 9{,}000$, $v = 12$, $C = 170$. Carrier A flies an average stage of 800 km at a CASK of 13.7 cents, exactly on the model's curve. Carrier B flies 2,000 km at 10.3 cents, 6 percent above the curve. On the raw numbers B looks 24.8 percent cheaper.

Moving B along its own cost curve to 800 km with (2.3) gives $10.29 \times 13.68/9.71 = 14.5$ cents: at A's stage length, B is 6.0 percent more expensive, which is the true difference built into the

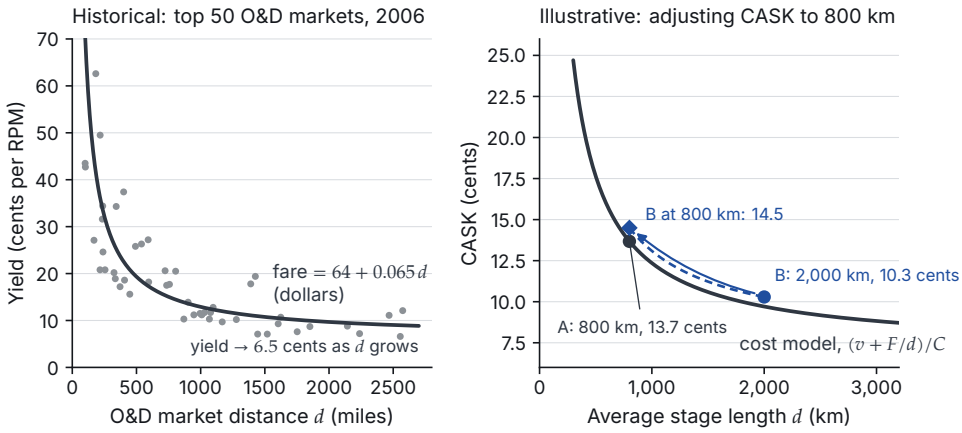


Figure 2.4 Left: yields in the 50 largest O&D markets, read from the chart in Belobaba's 2006 airline economics lecture, with the least-squares fare taper $f = a + b\,d$ (historical data, cents per revenue passenger-mile). Right: an illustrative cost model $\mathrm{CASK} = (v + F/d)/C$ with F = CAD 9,000, v = CAD 12 per km and $C = 170$ seats, two carriers A and B, and B's unit cost moved along its own cost curve to A's stage length.

example. Two constant-elasticity adjustments give different answers. With $\theta = 0.3$, B's adjusted CASK is $10.29 \times 2.5^{0.3} = 13.5$ cents, 1.0 percent cheaper than A; with $\theta = 0.5$ it is $10.29 \times 2.5^{0.5} = 16.3$ cents, 18.9 percent more expensive. The arc elasticity of the model between 800 and 2,000 km is -0.374, and $\theta = 0.374$ is the only exponent that reproduces the correct 6.0 percent.

The adjustment decided the answer. With the wrong exponent the analyst would conclude either that B is cheaper or that it is much more expensive than it is.

The example shows the first caution: stage-length adjustment is a model, and the conclusion can depend on its exponent. The second caution is that distance is only one of the things that differ between operations. Longer stages usually come with larger aircraft, different cabin mixes, different labour costs and different airports, and an adjustment for distance alone attributes all of these differences to the carrier's efficiency. Belobaba's 2003 data showed low-fare carriers with the lowest CASM at every stage length, a gap that no distance adjustment removes (Belobaba 2006e). For a network decision the adjustment is beside the point: the relevant cost of a route is its own avoidable cost, computed from its own block hours, departures and passengers (Chapter 3). The handbook warns specifically against

applying a system average cost mechanically to an incremental route (Network Planning Handbook 2026).

2.6 Break-even load factor

At what load factor does an operation cover its costs? Setting the operating margin per ASK of Result 2.2 to zero gives the answer the handbook uses (Network Planning Handbook 2026).

RESULT 2.4 Break-even load factor

If yield p, other revenue per ASK ρ and unit cost CASK do not depend on the load factor, operating profit is zero at

$$\mathrm{LF}^{*} = \frac{\mathrm{CASK} - \rho}{p}.$$

If instead unit cost is $\mathrm{CASK} = \kappa_0 + \kappa_1\,\mathrm{LF}$, where κ_0 is cost per ASK that does not vary with the number of passengers and κ_1 is passenger-related cost per RPK, and passengers also bring other revenue α per RPK, the break-even load factor is

$$\mathrm{LF}^{*} = \frac{\kappa_0 - \rho}{p + \alpha - \kappa_1},$$

where ρ now covers only the other revenue that does not depend on passengers. The simple formula, evaluated with costs and per-passenger revenue measured at an observed load factor LF_0, overstates this break-even by $\frac{\kappa_1 - \alpha}{p}\,(\mathrm{LF}_0 - \mathrm{LF}^{*})$.

Proof. Profit per ASK is $p\,\mathrm{LF} + \rho - \mathrm{CASK}$ in the first case; setting it to zero gives the first formula. In the second case profit per ASK is $(p + \alpha - \kappa_1)\,\mathrm{LF} + \rho - \kappa_0$, which is zero at the second formula. Write $N = \kappa_0 - \rho$ and $\delta = \kappa_1 - \alpha$. The simple formula with costs and other revenue measured at LF_0 gives $(N + \delta\,\mathrm{LF}_0)/p$, and

$$\frac{N + \delta\,\mathrm{LF}_0}{p} - \frac{N}{p - \delta} = \frac{\delta}{p}\Big(\mathrm{LF}_0 - \frac{N}{p - \delta}\Big) = \frac{\delta}{p}\,(\mathrm{LF}_0 - \mathrm{LF}^{*}).$$

■

The correction matters little at the system level, where passenger-related costs are a small part of unit cost, but it can matter on a single route evaluated far from break-even. More important are the assumptions the formula makes, which the next example tests.

WORKED EXAMPLE 2.5 Break-even on YUL–SVQ

Use the handbook's illustrative Seville departure in Laurentian form (Network Planning Handbook 2026): a 182-seat A321XLR flies 5,556 km with 150 passengers paying an average net fare of CAD 620, ancillaries of CAD 30 per passenger, cargo contribution of CAD 3,500 and an avoidable cost of CAD 88,000, made of CAD 80,500 that does not depend on the number of passengers and CAD 50 per passenger.

Unit measures. ASK $= 182 \times 5{,}556 = 1{,}011{,}192$ and RPK $= 150 \times 5{,}556 = 833{,}400$, so the load factor is 82.4 percent, the yield $93{,}000/833{,}400 = 11.16$ cents per RPK and PRASK $= 11.16 \times 0.824 = 9.20$ cents. Ancillaries and cargo add CAD 8,000, or 0.79 cents per ASK, for a RASK of 9.99 cents against a CASK of 8.70 cents. The margin of 1.29 cents per ASK is the handbook's flight contribution of CAD 13,000.

Break-even. The handbook's formula with these numbers gives $(8.70 - 0.79)/11.16 = 70.9$ percent. Counting exactly, each passenger adds CAD $620 + 30 - 50 = 600$ and the departure needs $(80{,}500 - 3{,}500)/600 = 128.3$ passengers, a break-even load factor of 70.5 percent. The difference, 0.4 points, is the overstatement of Result 2.4: $\frac{50-30}{620}(0.824 - 0.705) = 0.004$.

When yield moves. Now compare, as the handbook does, 170 passengers at an average fare of CAD 530. Passenger revenue falls from CAD 93,000 to CAD 90,100 even though the load factor rises to 93.4 percent. Each passenger now adds CAD 510, so the break-even load rises to $77{,}000/510 = 151$ passengers, a load factor of 83.0 percent, and the contribution falls to CAD 9,700. The fuller flight earns less.

The example shows the formula's central limitation. It treats yield as fixed, but load factor and yield are chosen together: revenue management fills an aircraft by opening cheaper fare classes, so a higher load factor usually comes with a lower yield, and the break-even target moves with it. The handbook adds two further limits: the fare mix changes with demand and inventory availability, and connecting passengers have network effects that a single route's ratio ignores; the break-even load factor is a diagnostic and should not replace itinerary-level analysis (Network Planning Handbook 2026). It is also an average: it says at what load the operation covers its costs on average, not whether one more passenger, or one more seat,

is worth having. Those are marginal questions, answered with spill and bid prices (Chapters 9 and 20).

Load factor is an outcome of capacity, demand and pricing together, not an objective. The break-even load factor tells you how much room there is between them, under the assumption that fares do not move.

2.7 Reading Air Canada's second-quarter 2026 MD&A

Air Canada's management's discussion and analysis (MD&A) for the second quarter of 2026, published on 11 August 2026, reports the quarter's results with the measures of this chapter (Air Canada 2026e). Its operating statistics include the regional carriers that fly for Air Canada under capacity purchase agreements, distances are in miles, and its regional breakdown (Canada, US transborder, Atlantic, Pacific and Other) gives revenues and percentage changes but not regional capacity levels. Table 2.1 transcribes the figures used below.

Table 2.1 Air Canada, operating statistics and results for the second quarter of 2026 and 2025, as printed in the MD&A (pp. 1 and 8), including regional partners flying under capacity purchase agreements. Changes are those reported; pp means percentage points.

Statistic	Q2 2026	Q2 2025	Change
Revenue passenger miles, RPM (millions)	23,597	22,796	+3.4%
Available seat miles, ASM (millions)	26,942	26,860	+0.3%
Passenger load factor	87.5%	84.9%	+2.6 pp
Yield (cents per RPM)	23.7	22.1	+7.2%
PRASM (cents)	20.7	18.7	+10.5%
TRASM (cents)	23.3	21.0	+10.9%
CASM (cents)	24.1	19.4	+23.9%
Adjusted CASM (cents)	15.5	14.4	+7.4%
Seats dispatched (thousands)	14,644	14,478	+1.2%
Aircraft frequencies (thousands)	100.3	98.5	+1.8%
Average stage length (miles)	1,840	1,855	−0.8%
Revenue passengers carried (thousands)	12,032	11,551	+4.2%
Fuel cost per litre (cents)	132.7	88.0	+50.8%
Passenger revenue (CAD millions)	5,576	5,031	+11%
Cargo revenue (CAD millions)	325	253	+29%
Other revenue (CAD millions)	365	348	+5%

Statistic	Q2 2026	Q2 2025	Change
Operating expenses (CAD millions)	6,481	5,214	+24%
of which aircraft fuel	1,713	1,148	+49%
Operating income (CAD millions)	−215	418	

WORKED EXAMPLE 2.6 **Recompute, reconcile, interpret**

Step 1: recompute every ratio from its parts. From the levels in Table 2.1, the Q2 2025 ratios all reproduce exactly. The load factor is 84.9 percent (22,796/26,860), the yield 22.1 cents (5,031/22,796), PRASM 18.7, TRASM 21.0, CASM 19.41 and adjusted CASM 14.40 cents, and the seat-weighted stage length 1,855 miles (26,860 million ASMs over 14.478 million seats). For Q2 2026, PRASM (5,576/26,942 = 20.70 cents), TRASM (23.26), CASM (24.06; the MD&A's own two-decimal table prints 24.05), adjusted CASM (4,167/26,942 = 15.47 cents, using the adjusted expense in the MD&A's reconciliation) and stage length (1,840 miles) also reproduce. Two do not: 23,597/26,942 = 87.58 percent, which rounds to 87.6, not the printed 87.5, and 5,576/23,597 = 23.63 cents, which rounds to 23.6, not 23.7.

Step 2: find what is inconsistent. The discrepancies are too large to be rounding of the inputs. Two other printed figures point the same way: traffic growth of 3.4 percent (23,597/22,796 is 3.5 percent) and yield growth of 7.2 percent (the printed levels imply 7.1). The printed load factor gain of 2.6 points (the levels imply 2.7) adds no independent evidence, because the MD&A's percentage-point changes appear to be differences of the rounded percentages: 87.5 minus 84.9, just as the first-half gain of 3.4 points is 86.9 minus 83.5 although the unrounded first-half levels differ by 3.3. All the printed ratios are consistent with Q2 2026 traffic between about 23,560 and 23,580 million RPM, and none with the printed 23,597. Either the traffic level or the ratios contain an error, or they were computed from different data; the document does not say which. The first-half figures are internally consistent, which suggests that the problem lies in the quarterly figures. An analyst would note the discrepancy, use the ratios with a caveat, and check the comparative figures in the next report. The lesson is general: check the identities before using a figure.

Step 3: check the identity and decompose growth. With the printed ratios, $23.7 \times 0.875 = 20.74$ cents, consistent with the printed PRASM of 20.7 within rounding; with the printed levels, $23.63 \times 0.8758 = 20.70$ exactly. By (2.2), the 10.5 percent rise in PRASM splits into about 7 points of yield and 3 points of load factor: Air Canada sold its seats both fuller and dearer. Capacity barely moved, and it too decomposes: 1.8 percent more departures were offset by an average gauge 0.7 percent smaller (146.0 seats per departure against 147.0) and a stage length 0.8 percent shorter, which sum, in logarithms, to the 0.3 percent growth in ASMs.

Step 4: margin and break-even. TRASM minus CASM is $23.26 - 24.06 = -0.80$ cents, and -0.80 cents on 26,942 million ASMs is the reported operating loss of CAD 215 million. Other revenue per ASM is $(325 + 365)/26{,}942 = 2.56$ cents, so the break-even load factor of Result 2.4 is $(24.06 - 2.56)/23.63 = 91.0$ percent, above the 87.6 percent achieved (with 23,570 million RPM, the middle of the range implied by the printed ratios, 90.9 against 87.5 percent). A year earlier the same calculation gives 77.8 percent against 84.9 percent achieved. Figure 2.5 decomposes the 13.1-point rise. Fuel per ASM rose from 4.27 to 6.36 cents while fuel burned per ASM stayed at about 47.3 millilitres, so the whole increase came from price; it alone is worth 9.4 points of load factor at the 2025 yield. The CAD 388 million of pension, labour and legal charges add 6.5 points, other costs 5.1 points; higher cargo and other revenue take 1.5 points back and the 7 percent higher yield 6.4 points. Without the special charges the break-even load factor would have been 84.9 percent, below the load factor achieved, and the quarter would have shown an operating profit of about CAD 173 million.

Step 5: regions. By region, passenger revenue was 32.7 percent Atlantic, 27.0 percent Canada, 19.3 percent US transborder, 14.0 percent Pacific and 7.0 percent other international markets. Applying (2.2) to the regional percentage changes splits each region's PRASM change into yield and load factor (Figure 2.6). Transborder capacity fell 5.7 percent while traffic held, and PRASM rose 18.5 percent, two thirds of it from yield; the MD&A attributes the yield gain to higher fares and lower industry capacity in the market (Air Canada 2026e). Pacific capacity grew 7.9 percent, the load factor was flat and PRASM

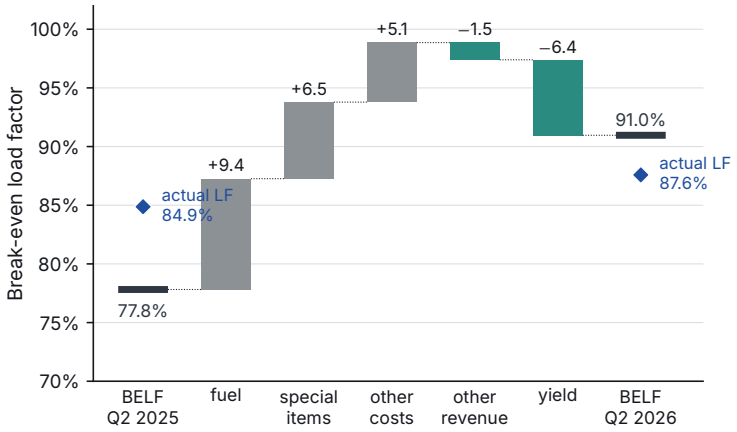


Figure 2.5 Air Canada's break-even load factor, Q2 2025 to Q2 2026, computed from the MD&A as (CASM minus other revenue per ASM) divided by yield, with the actual load factors (cobalt). The steps are applied in the order shown; a different order would split the total differently. Data: Air Canada Q2 2026 MD&A.

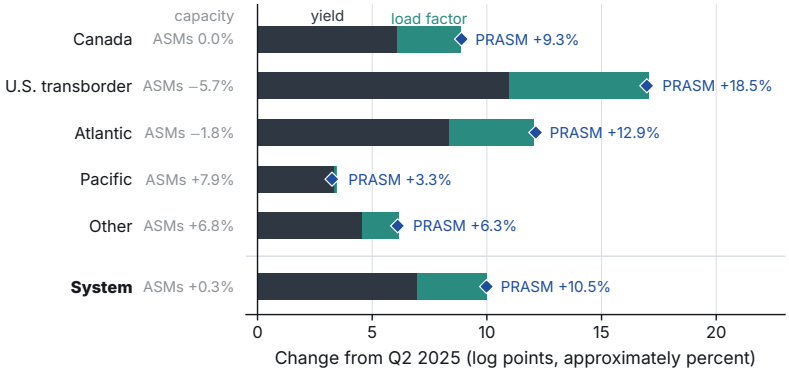


Figure 2.6 The change in Air Canada's passenger revenue per ASM between Q2 2025 and Q2 2026, by region, split into the contributions of yield and load factor (log points, which approximate percentage changes), with the reported PRASM change (cobalt) and the change in capacity. Data: Air Canada Q2 2026 MD&A, p. 9.

rose only 3.3 percent. The pattern is consistent with capacity restraint supporting unit revenue, but the data cannot establish cause: the same quarter saw higher fares in every region, partly in response to the fuel shock.

What can an analyst conclude from such a reading, and what not? The measures support statements about the airline as a whole: unit

revenue rose through fares more than through load factor, unit cost rose faster because of fuel and one-time charges, and the margin of safety between break-even and actual load factor disappeared in the quarter. They do not support statements about routes. The regional figures aggregate many routes and give no capacity levels, so regional load factors and yields cannot be recovered; trying to back out the Canadian load factor from the rounded changes (traffic up 2.8 percent, capacity flat, load factor up 2.4 points) places its 2025 level anywhere between about 81 and 91 percent. Regional yields would not be comparable anyway, since the Atlantic and Pacific regions fly far longer stages than the domestic network (Section 2.5), and none of the figures separates fare changes from changes in cabin mix, point of sale or surcharges. The handbook makes the same point about this report: its premium and sixth-freedom signals describe the company, not the profitability of any route (Network Planning Handbook 2026).

Two further measurement cautions arise from the same table. First, dividing passengers carried by seats dispatched gives 82.2 percent (12,032/14,644), five points below the RPM-based load factor. Part of the gap is the weighting of Result 2.1, since the long flights that dominate ASMs tend to be fuller; part is the flight-number counting of passengers, which counts a through passenger once on a multi-leg flight. The MD&A does not allow the two effects to be separated. Second, comparisons with other airlines need the same care as stage-length adjustments: definitions of revenue scope, regional partners and adjusted costs differ, and the MD&A itself cautions that its non-GAAP measures may not be comparable with those of other companies (Air Canada 2026e).

IN PRACTICE **Unit measures in 2026 reporting and guidance**

Airlines and their industry body speak to investors in exactly these units. When Air Canada reinstated its 2026 guidance in August, it expressed it as growth in ASMs (2.25 to 3.25 percent for the year) and in adjusted CASM (5 to 6 percent), alongside adjusted EBITDA and free cash flow, with explicit fuel price and exchange-rate assumptions (Air Canada 2026d). IATA's June 2026 industry outlook likewise forecast 2026 growth of 2.1 percent in RPKs and 1.6 percent in ASKs, with departures down 0.6 percent, and put fuel at 31.4 percent of industry costs (IATA 2026d); by the capacity decomposition of Section 2.5, more ASKs on fewer departures means larger aircraft, longer stages or both. Air Canada also re-

ports TRASM, includes its capacity purchase partners in its operating statistics and publishes adjusted CASM as a non-GAAP measure (Air Canada 2026e). Before comparing any of these figures across airlines, read each company's definitions.

2.8 Summary

- Keep four objects apart: aircraft movements, legs, itineraries and tickets. Demand lives in O&D markets, supply on legs; leg counts exceed journey counts by the extra legs connecting journeys use.
- PDEW divides journeys by both the number of days and the number of directions. A market's PDEW is not the load of a new flight.
- ASK and RPK weight seats and passengers by distance. The network load factor is RPK/ASK, an ASK-weighted average of leg load factors; the simple average of load factors differs from it by n times the covariance between weights and load factors.
- Yield is passenger revenue per RPK, PRASK passenger revenue per ASK, RASK total revenue per ASK, CASK cost per ASK. PRASK equals yield times load factor, and operating profit equals ASK times (RASK minus CASK).
- Yield and unit cost both fall with distance, because fares and costs both contain components incurred once per trip. The stage-length elasticity of CASK is minus the share of per-departure costs and is not constant, so a constant-elasticity adjustment can change the conclusion of a comparison.
- The break-even load factor is (CASK minus other revenue per ASK) divided by yield. It assumes a fixed yield and ignores network effects; it is a diagnostic, not a target.
- Air Canada's Q2 2026 statistics reproduce from their parts except for one traffic figure, which is inconsistent with four other printed figures. Check identities before using a number.
- The same statistics show unit revenue up 10.5 percent, mostly through yield, and a break-even load factor that rose from 77.8 to 91.0 percent, above the load factor achieved, mainly because of fuel and one-time charges. They say nothing about individual routes.

Table 2.2 Key formulas of the chapter.

Quantity	Formula
Leg passengers	$N_l = \sum_j a_{lj} n_j$
PDEW	journeys in both directions $/(2T)$
Capacity and traffic	$\text{ASK} = \sum_l C_l d_l$, $\text{RPK} = \sum_l N_l d_l$
Load factor	$\text{LF} = \text{RPK}/\text{ASK} = \sum_l w_l \text{LF}_l$, $w_l \propto C_l d_l$
Unit revenue	$\text{PRASK} = p\,\text{LF}$, $\text{RASK} = p\,\text{LF} + \rho$
Operating profit	$\Pi = \text{ASK}\,(\text{RASK} - \text{CASK})$
Growth decomposition	$\Delta\text{PRASK} = \Delta p + \Delta\text{RPK} - \Delta\text{ASK}$, with Δ the change in the logarithm
Capacity decomposition	ASK = departures × seats per departure × ASK per seat
Distance and unit cost	$\text{CASK}(d) = (v + F/d)/C$, elasticity $-F/(F + vd)$
Break-even load factor	$\text{LF}^* = (\text{CASK} - \rho)/p$, or $(\kappa_0 - \rho)/(p + \alpha - \kappa_1)$

2.9 Further reading

Belobaba's lectures on the planning process and on airline economics (2006c, 2006d) define the measures and the dichotomy between O&D demand and leg supply in a few slides each; his lecture on operating costs (2006e) shows the historical relationships between stage length, unit cost and productivity that Chapter 3 develops. The handbook's table of vocabulary and common errors (2026) is worth keeping at hand during your first weeks. Read the glossary and the non-GAAP section of Air Canada's MD&A (2026e) before any results call, and the next quarter's report to see whether the traffic figure discussed in this chapter is restated. For a textbook treatment, see the chapters on airline economics and on operating costs and productivity in Belobaba, Odoni and Barnhart (2015, chs. 3 and 6).

CHAPTER

3

Costs and Productivity

IN THIS CHAPTER

- How airlines classify operating costs, by what they buy and by what the money is for, and which activity drives each category.
- A cost model for one flight built from block hours, departures and passengers, and the block-time function that ties it to distance.
- Why cost per seat-kilometre falls with stage length, derived from that model, and the elasticity it implies.
- Trip cost versus seat cost: why a larger aircraft is cheaper per seat but riskier per trip.
- Utilization and productivity: how aircraft time spreads ownership cost, and why the same aircraft costs one airline twice what it costs another.
- Which costs count for which decision: fully allocated, avoidable and incremental cost by decision horizon.
- Laurentian's trip costs and break-even loads, and Air Canada's reported cost structure in the second quarter of 2026.

DRAWS ON Belobaba, Operating costs and productivity measures (MIT 16.75J, 2006); handbook; Air Canada Q2 2026 MD&A

Every proposal a network planner writes ends in a comparison of revenue with cost. Revenue gets most of the attention: demand, fares, spill, connections. Cost looks easier because the airline's accounts report it to the dollar. That impression is misleading. The accounts say what the airline spent last quarter on fuel, wages and airport fees. They do not say what one more departure to Seville would cost, or what the airline would save by cancelling the Tuesday flight to Boston. Those are different numbers, and the gap between them settles a large share of network decisions.

This chapter builds the cost side from first principles. It starts with the categories airlines use to report costs and the activity that drives each one, then turns the drivers into a cost model for a single flight. Two consequences of that model shape every network: unit cost falls with stage length, and larger aircraft cost less per seat but more per

trip. We then study utilization and productivity, which explain why one aircraft type can have very different costs at two airlines, and the question a planner must put to every cost figure: avoidable over which horizon? The chapter closes with Laurentian's cost model at work and with Air Canada's reported costs. The unit measures used here (ASK, CASK, yield, load factor) were defined in Chapter 2; the cost model feeds route evaluation (Chapter 12), fleet planning (Chapter 11) and fleet assignment (Chapter 14).

3.1 How airlines classify costs

Airline cost data come in two shapes, and the lecture material in the reading library introduces both through the reporting system of the US Department of Transportation, Form 41, under which large US carriers file traffic, financial and operating cost data, mostly quarterly (Belobaba 2006e). Aircraft operating expenses are reported by aircraft type and region; other expenses are reported for the airline as a whole because they are hard to attach to a type.

DEFINITION 3.1 Objective and functional cost categories

An **objective** classification groups costs by what is bought: labour, fuel, aircraft ownership, professional services, landing fees, food and beverage, maintenance material, commissions, insurance, communication, advertising and other items. A **functional** classification groups costs by the activity they serve:

- **aircraft operating**: the costs of flying the aircraft, also called direct operating costs;
- **aircraft servicing**: handling the aircraft on the ground, including landing fees;
- **traffic servicing**: processing passengers, baggage and cargo at airports;
- **passenger service**: cabin crew, meals and other in-flight service;
- **reservations and sales**: reservation centres, ticket offices and agency commissions;
- **other** costs, such as advertising and general administration.

The two views answer different questions. The objective view tells the finance team where the money goes. In the third quarter of 2005,

US airlines spent about 3 cents per available seat mile on labour and about as much on fuel, each a quarter of the total, about 1.5 cents on aircraft ownership and close to 1 cent on professional services, with a large residual "other" of about 2.6 cents (Belobaba 2006e). The functional view tells a planner what causes a cost. A pilot's salary is labour in the objective view and aircraft operating cost in the functional view; a gate agent's salary is also labour, but it belongs to traffic servicing and moves with passengers, not with flying hours. Planning needs the functional view.

The lecture adds a coarser grouping used by Boeing and MIT for comparisons. **Flight operating costs** cover everything needed to fly the aircraft (pilots, fuel, maintenance, ownership) and were historically about half of an airline's costs; **ground operating costs** (servicing passengers and aircraft at stations, landing fees, reservations and sales) about 30 percent; and **system operating costs** (marketing, administration, overhead, in-flight service, ground equipment) about 20 percent (Belobaba 2006e). These shares are historical rules of thumb, useful only for orientation.

Flight operating costs are reported per block hour and per aircraft type, split into four parts: crew (pilot wages and benefits), fuel (the easiest cost to allocate and the most clearly variable), maintenance (direct labour and parts for airframe and engines, plus an overhead "burden" for hangars and spare parts) and ownership (depreciation, lease payments, insurance). For the Boeing 757-200 in 2003, the US average was USD 869 per block hour for crew, 904 for fuel, 875 for maintenance and 767 for ownership, a total of USD 3,415 per block hour at an average stage length of 1,267 miles and 9.7 block hours of flying per aircraft per day (Belobaba 2006e). The same aircraft type cost USD 2,924 per block hour at one airline and USD 4,308 at another. Stage length, utilization, labour contracts, wage rates and fleet age all enter that number, which is why a cost per block hour borrowed from another airline, or from another year, is a starting point at best.

3.2 Cost drivers and a flight cost model

A planner needs the cost of flights that do not exist yet. Accounts cannot give it directly, but they can give rates: how much each category costs per unit of an activity that the schedule determines.

DEFINITION 3.2 Cost driver

The **cost driver** of a cost category is the activity whose volume determines it. Five drivers matter for network decisions: **block hours** (the time from leaving the departure gate to arriving at the destination gate), **departures** (each takeoff and landing, also called a cycle), **passengers** carried, **distance** flown, and **calendar time** (costs incurred per day, month or year whatever the flying). Some costs are driven by revenue (commissions, card fees); overhead has no natural driver and is allocated by a rule.

The lecture attaches one driver to each functional category and reports average rates for US carriers in 2003, reproduced in Table 3.1 (Belobaba 2006e). The rates are historical; the structure is permanent.

Table 3.1 Functional cost categories, their activity drivers and the 2003 US average rates (adapted from Belobaba, 2006; historical values), with their treatment in Laurentian's cost model.

Functional category	Driver	2003 US average	Laurentian's model
Aircraft operating	block hours	USD 3,415 per block hour (757-200)	c_{BH} by type
Aircraft servicing	departures	USD 1,135 per departure	c_{dep} by type and region
Traffic servicing	passengers enplaned	USD 18 per passenger	in c_{pax}
Passenger service	revenue passenger miles	USD 0.015 per RPM	in c_{pax} by region
Reservations and sales	revenue	10% of revenue	netted from fares
Overhead	total cost	12% of operating expense	excluded

Collapsing the table into three drivers gives the cost model used throughout this book. For a departure with block time BH carrying N passengers, the **trip cost** is

$$\mathrm{TC} = c_{\mathrm{BH}}\,\mathrm{BH} + c_{\mathrm{dep}} + c_{\mathrm{pax}}\,N, \tag{3.1}$$

where c_{BH} is the cost per block hour of the aircraft type (fuel, crew, the variable part of maintenance, navigation charges), c_{dep} the cost

per departure (landing and terminal charges, handling, station costs per turn) and c_{pax} the cost per passenger (catering, distribution, passenger fees). The first two terms do not depend on how many people board; we call them the **capacity cost** of the departure. Laurentian's running case uses exactly this structure with the parameters of Table 3.2. Ownership and overhead are left out on purpose: the model measures the cost a departure causes at the seasonal planning horizon, a choice explained in Section 3.6.

Table 3.2 Laurentian's illustrative cost parameters (CAD) from the running case, with departure costs for domestic and transborder flights and for international flights. Passenger costs are CAD 20 (domestic), 30 (transborder) and 50 (transatlantic) per passenger. Ranges are approximate manufacturers' headline figures, not mission capability (Chapter 11).

Type	Seats	Per block hour	Per departure, dom. / intl	Range (nm)
A220-300	137	8,000	2,000 / 3,000	3,400
737-8	169	9,000	2,500 / 3,500	3,500
A321XLR	182	11,000	2,500 / 3,500	4,700
787-9	298	15,500	5,000 / 7,000	7,500

WORKED EXAMPLE 3.1 Costing a flight with the lecture's drivers

Use the 2003 rates of Table 3.1 to cost a 1,000-mile flight of a 185-seat 757-200 with an assumed block time of 2.6 hours, carrying 140 passengers at an average fare of USD 150. All numbers are historical and in US dollars.

Aircraft operating cost is $2.6 \times 3{,}415 = 8{,}879$; servicing 1,135; traffic servicing $140 \times 18 = 2{,}520$; passenger service $140 \times 1{,}000 \times 0.015 = 2{,}100$; reservations and sales $10\% \times 21{,}000 = 2{,}100$. These sum to 16,734. If overhead is 12 percent of total cost, the total is $16{,}734/0.88 = 19{,}016$, so the flight earns $21{,}000 - 19{,}016 = 1{,}984$. Cost per seat mile is $19{,}016/185{,}000 = 10.28$ cents, yield is 15.0 cents and the load factor is 75.7 percent.

What load factor breaks even? The shortcut of Chapter 2, CASM divided by yield, gives $10.28/15.0 = 68.5$ percent. It is wrong for this purpose, because the CASM of 10.28 cents was computed at 140 passengers and includes passenger-driven costs that would disappear with the passengers. Separate the two kinds of cost, allocating overhead in proportion as the 12 percent rule implies. The flight itself costs $(8{,}879 + 1{,}135)/0.88 =$

11,380; each passenger costs $(18 + 15 + 15)/0.88 = 54.5$ (traffic servicing, passenger service on 1,000 miles, and 10 percent of the fare). Break-even needs $11{,}380/(150 - 54.5) = 119.2$ passengers, a load factor of 64.4 percent.

The planner's lesson: before computing any break-even, sort the cost into what the departure causes and what each passenger causes.

3.3 Why unit cost falls with stage length

Part of every flight's cost does not depend on how far it goes. The aircraft pays landing and handling charges once per departure; it taxis out, climbs, descends and taxis in once per departure, burning block hours that carry passengers no closer to their destination. On a long flight these fixed amounts are spread over many seat-kilometres; on a short flight they dominate. To make this precise we need block time as a function of distance.

$$\mathrm{BH}(d) = \tau_0 + \frac{d}{s}, \tag{3.2}$$

where d is the stage length in kilometres, τ_0 the fixed time allowance per flight (taxi out and in, the slow part of climb and approach) and s the effective cruise speed in kilometres per block hour. For Laurentian we use $\tau_0 = 0.65$ hours and $s = 850$ km/h. This line is a smooth approximation, not a timetable. It gives 0.92 hours for Quebec City–Montreal (233 km, against the running case's scheduled 0.9), 1.13 for Boston–Montreal (409 km, scheduled 1.4), 4.98 for Montreal–Vancouver (3,682 km, 5.2), 7.15 for Montreal–Paris (5,524 km, 6.9) and 7.19 for Montreal–Seville (5,556 km, scheduled 7.0 eastbound and 7.9 westbound). The errors run from 1 to 16 minutes, except on the westbound Seville flight, which is scheduled 44 minutes longer than the line because real block times differ by direction with the winds, a point taken up in Chapter 11. The line serves the general results of this section and fleet-level illustrations; worked examples on the running case's routes use their scheduled block times.

RESULT 3.1 Unit cost and stage length

Let an aircraft with C seats have cost c_{BH} per block hour and c_{dep} per departure, with block time (3.2). Its capacity cost per available seat-kilometre on a stage of d km is

$$\mathrm{CASK}(d) = \frac{a}{d} + b, \qquad a = \frac{c_{\mathrm{dep}} + c_{\mathrm{BH}}\tau_0}{C}, \qquad b = \frac{c_{\mathrm{BH}}}{Cs}. \quad (3.3)$$

CASK falls with distance toward the cruise cost b, and its elasticity with respect to stage length equals minus the share of capacity cost that does not vary with distance:

$$\frac{\mathrm{d}\ln \mathrm{CASK}}{\mathrm{d}\ln d} = -\frac{F}{F + v\,d}, \qquad F = c_{\mathrm{dep}} + c_{\mathrm{BH}}\tau_0, \qquad v = \frac{c_{\mathrm{BH}}}{s},$$

where v is the capacity cost per kilometre flown, as in Chapter 2.

Proof. The capacity cost of the departure is $c_{\mathrm{BH}}(\tau_0 + d/s) + c_{\mathrm{dep}} = F + vd$ and it produces Cd seat-kilometres, so $\mathrm{CASK} = (F + vd)/(Cd) = F/(Cd) + v/C$, which is (3.3) because $v/C = c_{\mathrm{BH}}/(Cs) = b$. Since $\ln \mathrm{CASK} = \ln(F + vd) - \ln C - \ln d$, the derivative is $vd/(F + vd) - 1 = -F/(F + vd)$. ∎

Passenger costs add a term of the same shape. With $N = \mathrm{LF} \cdot C$ passengers, the passenger cost per ASK is $c_{\mathrm{pax}}\mathrm{LF}/d$, again falling as $1/d$ because a passenger is checked in and fed once per flight, not once per kilometre.

Figure 3.1 draws (3.3) for Laurentian's fleet. The per-seat cost of a departure, a, is remarkably similar across types: CAD 52.6 for the A220-300, 49.4 for the 737-8, 53.0 for the A321XLR and 50.6 for the 787-9, because departure costs and block-hour costs both scale roughly with aircraft size. The types differ mainly in the cruise cost b: 6.12 cents per ASK for the 787-9, 6.27 for the 737-8, 6.87 for the A220-300 and 7.11 for the A321XLR. The XLR is the most expensive aircraft per seat-kilometre in the fleet. Its block-hour cost is high for its seat count, the price of a low-density cabin and of the fuel and structure that give it transatlantic range. Whether that price is worth paying is a question about trip cost and demand, not unit cost (Section 3.4).

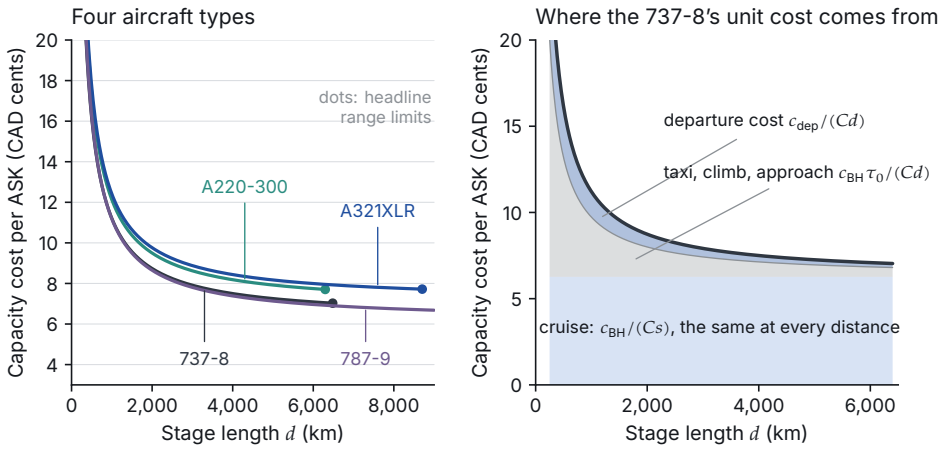


Figure 3.1 Left: capacity cost per ASK of Laurentian's four aircraft types as stage length grows (domestic departure costs, block time $0.65 + d/850$ hours); dots mark the headline range limits. Right: the 737-8's unit cost split into the cruise cost b, which does not depend on distance, and the two distance-independent costs per departure, spread over d.

WORKED EXAMPLE 3.2 **One aircraft, two stage lengths**

Fly Laurentian's 737-8 from Montreal to Toronto (507 km) and from Montreal to Vancouver (3,682 km). Block times from (3.2) are 1.246 and 4.982 hours (the running case schedules Vancouver at 5.2 hours; the smooth line keeps the comparison between distances clean). Capacity costs are

$$\mathrm{TC}_{\mathrm{YYZ}} = 9{,}000 \times 1.246 + 2{,}500 = 13{,}714,$$
$$\mathrm{TC}_{\mathrm{YVR}} = 9{,}000 \times 4.982 + 2{,}500 = 47{,}338,$$

and dividing by 169×507 and $169 \times 3{,}682$ seat-kilometres gives CASKs of 16.0 and 7.6 cents. The Vancouver flight is 7.3 times as long, costs 3.5 times as much, and has less than half the unit cost. The elasticity of Result 3.1 is -0.61 at 500 km and -0.18 at the Vancouver distance: at short range, 10 percent more distance lowers unit cost by about 6 percent; at long range, by under 2 percent.

The lesson is that unit cost is partly geometry. A short-haul route is not inefficient because its CASK is high, and a network that lengthens its average stage lowers its CASK without becoming any more efficient. Comparisons of CASK across routes, fleets or airlines must first remove the stage-length ef-

fect; Chapter 2 shows the standard adjustment, whose local exponent is the fixed-cost share $F/(F + vd)$ derived here.

The lecture's cross-section of US carriers in 2003 shows both forces at once (Belobaba 2006e). Within the group of legacy carriers and within the group of low-fare carriers, unit cost falls with average stage length; between the groups, the low-fare carriers sit 3 to 6 cents per ASM lower at every stage length (JetBlue reported 6.1 cents at 1,279 miles where United reported 11.0 cents at 1,259 miles). Geometry explains the slope, and the rest of this chapter explains most of the gap.

3.4 Aircraft size: trip cost versus seat cost

Larger aircraft are cheaper per seat. Two pilots fly a 298-seat 787-9 as they fly a 137-seat A220; a departure needs one slot, one gate and one turn whatever the gauge; fuel burn grows less than proportionally with size. But a larger aircraft is more expensive per trip, and the trip cost is paid whether or not the extra seats fill.

DEFINITION 3.3 Trip cost, seat cost and cost per passenger

For one departure, the **trip cost** TC is the cost of operating it, the **seat cost** TC/C is the cost per seat offered, and the **cost per passenger** $\mathrm{TC}/\mathbb{E}[Y]$ is the cost per passenger carried, where $\mathbb{E}[Y] = \mathbb{E}[\min(D, C)]$ is the expected load of Chapter 9.

The handbook's example makes the distinction concrete. Aircraft A costs CAD 80,000 per trip and has 180 seats; aircraft B costs CAD 105,000 and has 280 seats (Network Planning Handbook 2026). B is cheaper per seat (CAD 375 against 444), A is cheaper per trip, and the handbook stresses that neither is preferable until revenue, demand and the opportunity cost of the aircraft are known. A useful way to see the pair is through an affine trip cost, $\mathrm{TC}(C) = \alpha + \beta C$: here $\beta = (105{,}000 - 80{,}000)/(280 - 180) = 250$ dollars per additional seat and $\alpha = 35{,}000$. The seat cost $\beta + \alpha/C$ falls with size because the fixed part α is shared by more seats. The question a planner must answer is whether the seats that B adds at CAD 250 each will be filled often enough.

RESULT 3.2 When the larger aircraft pays

Let each passenger contribute r (net fare less passenger costs) and let demand D have any distribution. Replacing aircraft A (C_A seats, trip cost TC_A) by a larger aircraft B (C_B seats, TC_B) raises expected contribution if and only if

$$\frac{1}{C_B - C_A}\int_{C_A}^{C_B} \Pr(D > x)\,\mathrm{d}x \;\geq\; \frac{\mathrm{TC}_B - \mathrm{TC}_A}{(C_B - C_A)\,r} = \frac{\beta}{r}.$$

The added seats must be filled, on average, with a probability at least equal to their marginal cost divided by the contribution per passenger.

Proof. Chapter 9 shows that $\mathbb{E}[\min(D, C)] = \int_0^C \Pr(D > x)\,\mathrm{d}x$, because seat number x is occupied exactly when demand exceeds x. The change in expected contribution is therefore

$$\begin{aligned} &r(\mathbb{E}[\min(D, C_B)] - \mathbb{E}[\min(D, C_A)]) - (\mathrm{TC}_B - \mathrm{TC}_A) \\ &\quad = r\int_{C_A}^{C_B} \Pr(D > x)\,\mathrm{d}x - (\mathrm{TC}_B - \mathrm{TC}_A). \end{aligned}$$

It is non-negative exactly when the stated inequality holds; divide by $r(C_B - C_A) > 0$. ∎

The result is the discrete counterpart of the newsvendor rule of Chapter 9: there the last seat is worth adding if it fills with probability at least c/r; here a block of seats is worth adding if it fills, on average, with probability at least β/r.

WORKED EXAMPLE 3.3 The handbook's two aircraft with uncertain demand

Let demand per departure be normal with mean μ and $k = 0.30$, and let each passenger contribute $r = 400$ dollars. At $\mu = 200$, the spill formula of Chapter 9 gives expected loads of 164.7 on A and 197.5 on B. B's extra 100 seats carry 32.7 more passengers per departure, an average fill probability of 0.327. Result 3.2 requires $\beta/r = 250/400 = 0.625$, so A is better; the arithmetic agrees: $400 \times 32.7 - 25{,}000 = -11{,}920$ dollars per departure for B.

Costs per passenger tell the same story (Figure 3.2). At $\mu = 200$ they are CAD 486 on A and 532 on B, although B is CAD 69

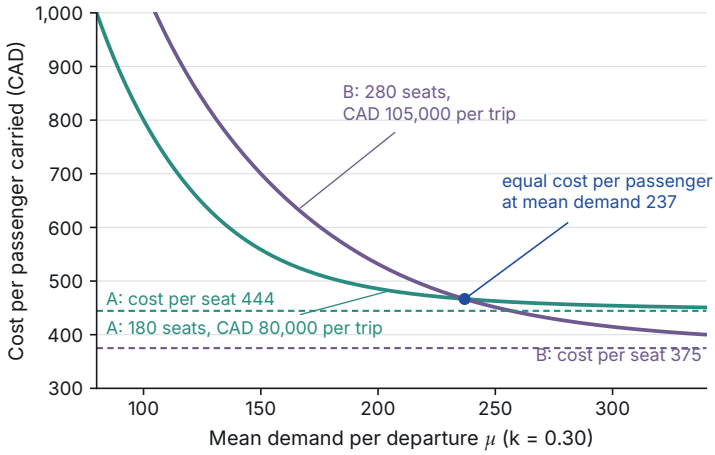


Figure 3.2 Cost per passenger carried for the handbook's aircraft A (180 seats, CAD 80,000 per trip) and B (280 seats, CAD 105,000) as mean demand grows, with normal demand and $k = 0.30$. The dashed lines are the costs per seat. B is always cheaper per seat but cheaper per passenger only when mean demand exceeds about 237.

cheaper per seat. The curves cross at a mean demand of 237, a demand factor of 1.32 on A's seats. In contribution terms, B starts to win at $\mu = 256$ when $r = 400$, at $\mu = 215$ when $r = 600$, and only at $\mu = 330$ when $r = 300$: the more each passenger is worth, the less often the extra seats need to fill.

Seat cost is a property of the aircraft. Cost per passenger and contribution are properties of the aircraft in a market. Laurentian's own fleet shows why the distinction matters: at a common seven-hour block (international departure costs), the seat costs are CAD 431 for the A220-300, 393 for the 737-8, 442 for the A321XLR and 388 for the 787-9. The XLR is the dearest seat in the fleet and the 787-9 the cheapest, yet for a market with 150 passengers a day the XLR's CAD 80,500 trip will beat the 787-9's CAD 115,500 by a wide margin. Chapter 11 turns this observation into a fleet decision.

3.5 Utilization and productivity

Some costs accrue with the calendar, not with flying. An aircraft's lease payment or depreciation is the same in a month when it flies 400 hours as in a month when it flies 250; so are much of the pay of salaried staff, rent and insurance. The more an airline flies its as-

sets, the thinner these costs are spread. The lecture summarizes the measures airlines use (Belobaba 2006e).

DEFINITION 3.4 Utilization and productivity

Aircraft utilization U is the number of block hours an aircraft flies per day. **Aircraft productivity** is the capacity it produces per day,

$$\frac{\text{ASK}}{\text{aircraft-day}} = n \times d \times C = U \times s_b \times C,$$

where n is the number of departures per day, d the average stage length, C the seats and $s_b = d/\text{BH}(d)$ the block speed. **Employee productivity** is measured as ASK per employee, ASK per dollar of labour cost, revenue per employee and revenue per dollar of labour cost.

If an aircraft's ownership cost is O dollars a year, its ownership cost per block hour is $O/(365\,U)$. The relation is a hyperbola: raising utilization by 10 percent cuts ownership cost per block hour by $1 - 1/1.1 = 9.1$ percent, and the gain shrinks as utilization rises. How much an aircraft can fly depends on its schedule.

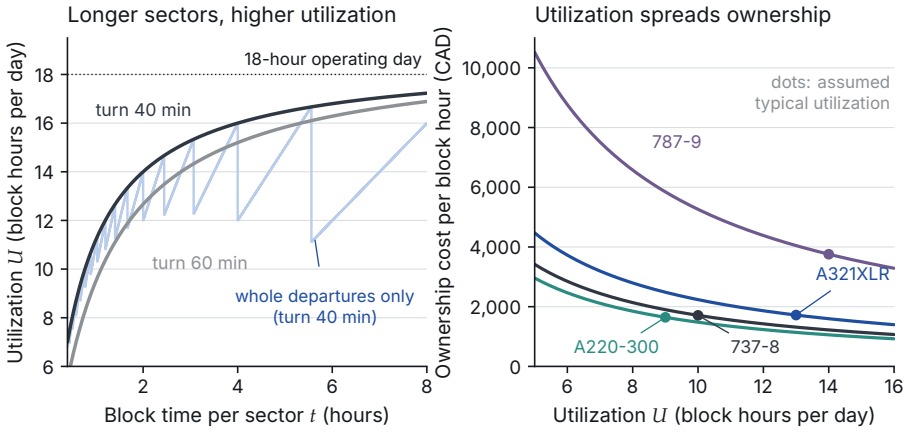


Figure 3.3 Left: utilization of an aircraft flying identical sectors within an 18-hour day, for turns of 40 and 60 minutes; the pale sawtooth counts whole departures only, for the 40-minute turn. Right: ownership cost per block hour against utilization for Laurentian's types, with illustrative monthly ownership costs of CAD 0.45 million (A220-300), 0.52 million (737-8), 0.68 million (A321XLR) and 1.60 million (787-9).

RESULT 3.3 Utilization of a shuttling aircraft

An aircraft that flies sectors of block time t separated by turns of length g, within an operating day of H hours, can fly $n = \lfloor (H+g)/(t+g) \rfloor$ sectors, so that

$$U = t\left\lfloor \frac{H+g}{t+g} \right\rfloor \approx \frac{(H+g)\,t}{t+g}. \tag{3.4}$$

The approximation is increasing and concave in t and decreasing in g: longer sectors and shorter turns raise utilization.

Proof. n sectors need n block times and $n-1$ turns, so $nt + (n-1)g \le H$, that is $n \le (H+g)/(t+g)$, and $U = nt$. Treating n as continuous gives the approximation, whose derivative with respect to t is $(H+g)g/(t+g)^2 > 0$ and decreasing in t; its derivative with respect to g is $-(H+g)t/(t+g)^2 + t/(t+g) = -t(H-t)/(t+g)^2 < 0$ for $t < H$. ∎

Figure 3.3 shows both halves of the argument. On the left, an aircraft flying one-hour sectors with 40-minute turns fits 11 sectors into the day, 11.0 block hours; with five-hour sectors it fits three, 15.0 block hours (the continuous approximation gives 11.2 and 16.5). This is the lecture's observation that average stage length and utilization move together, and it reinforces the stage-length effect of Re-

sult 3.1: longer flights spread departure costs over more kilometres and ownership costs over more hours. The sawtooth is a reminder that departures come in whole numbers: an aircraft that cannot quite fit an eighth sector in the day loses a full sector's worth of utilization, which is why a few minutes of turn time can matter more than they seem (Chapter 11 uses the manufacturer's turn-time charts to show this). On the right, a 787-9 owned for CAD 19.2 million a year costs CAD 3,757 per block hour at 14 hours a day and CAD 5,260 at 10. These ownership figures are illustrative assumptions for Laurentian, not market quotations.

Table 3.3 Boeing 737-500 productivity and flight operating costs at three US airlines (adapted from Belobaba, 2006; historical values from the early 2000s, the slides give no year). The last row is computed as cost per block hour times block hours per day, divided by ASMs per day.

	Continental	United	Southwest
Departures per day	3.9	4.3	8.2
Block hours per day	8.3	7.5	10.2
Average stage length (miles)	719	564	400
Seats	104	109	122
ASMs per aircraft-day	291,246	264,284	399,746
Crew per block hour (USD)	510	927	388
Fuel per block hour	430	487	537
Maintenance per block hour	651	1,048	251
Ownership per block hour	698	510	350
Total per block hour	2,291	2,974	1,526
Flight operating cost per ASM (cents)	6.53	8.44	3.89

WORKED EXAMPLE 3.4 Same aircraft, three airlines

The lecture compares the Boeing 737-500 at three US airlines (Table 3.3; the slides give no year, but the charts around them are for 2003). The aircraft is the same; the economics are not.

Southwest's cost per seat mile is less than half of United's. Decompose the ratio $8.44/3.89 = 2.17$. Cost per block hour is $2{,}974/1{,}526 = 1.95$ times higher at United, and Southwest produces 1.11 times as many seat miles per block hour ($399{,}746/10.2 = 39{,}191$ against $264{,}284/7.5 = 35{,}238$); indeed $1.95 \times 1.11 = 2.17$. The second factor is almost entirely seats (122 against 109): the two airlines' block speeds are nearly identical, 322 and 323 miles per block hour, despite Southwest's much

shorter stages. The first factor comes from crew rates (USD 927 against 388), maintenance (1,048 against 251) and ownership. Utilization explains most of the ownership gap: with equal annual ownership per aircraft, the ratio of ownership costs per block hour would be 7.5/10.2 = 0.74; the observed ratio is 350/510 = 0.69.

The lesson for a planner: a cost per block hour is a property of an airline's whole operation (labour contracts, maintenance practice, fleet age, seating density and, above all, how hard the aircraft is worked), not of the aircraft type.

The lecture's industry series make the same point over time (Belobaba 2006e). US major airlines produced 630,700 ASMs per aircraft-day in 2000, fell to about 588,000 in 2002 and 2003 as departures per aircraft dropped from 4.34 to 3.83 a day, and recovered to 633,500 in 2004, mostly because average stage length rose from 882 to 981 miles. Employee productivity rose about 30 percent between 2002 and 2005, largely through job cuts at legacy carriers. One finding runs against the theory: network carriers, despite longer stages and larger aircraft, produced fewer ASMs per employee than low-fare carriers. The lecture states the finding without explaining it; a plausible reason is that hubs, international operations and higher service levels need more people per seat.

Air Canada's second-quarter 2026 statistics allow the same calculations (Air Canada 2026e). Its operating fleet of 357 aircraft at the end of June (mainline, Rouge and the Express aircraft flown under capacity purchase agreements) flew 100,300 departures in the 91-day quarter with 14.6 million seats and produced 26,942 million ASMs. Six of the 357 are freighters, which produce no ASMs; using the period-end fleet of 351 passenger aircraft, that is 146 seats per departure, 3.14 departures per aircraft-day, an average stage length of 1,840 miles and about 843,000 ASMs per aircraft-day. Dividing ASMs by the 37,000 full-time-equivalent employees gives 728,000 ASMs per employee in the quarter, about 2.9 million a year. That last ratio needs a caution that the MD&A itself supplies: the operating statistics include the flying of third-party carriers under capacity purchase agreements, while the employee count covers Air Canada and its subsidiaries only, so the ratio overstates the labour productivity of the flying Air Canada staffs. Productivity ratios, like unit costs, must have matching numerators and denominators.

3.6 Which costs count: decision horizons

A cost figure answers a question only once the question says what is being decided and over what period. Cancelling a flight next Tuesday saves its fuel and landing fees, but not the pilots' salaries, which are fixed by monthly schedules, nor the aircraft lease. Dropping a route from next summer's schedule saves crew costs too, because staffing is planned for the season. Shrinking the fleet saves the lease, but only when the lease can be ended.

DEFINITION 3.5 Fully allocated, avoidable and incremental cost

For a decision with horizon h:

- the **fully allocated cost** of a flight is its share of all the airline's operating costs, including overhead assigned by an accounting rule;
- its **avoidable cost** is the cost the airline would not incur during h if the flight were not operated;
- the **incremental cost** of a decision is the change in the airline's total cost that it causes, wherever the change falls, including costs on other flights;
- a **committed** cost is already contracted and cannot be recovered within h;
- the **opportunity cost** of a scarce resource (an aircraft, a slot, a crew) is the value of its best alternative use.

Table 3.4 adapts the handbook's classification of cost buckets (Network Planning Handbook 2026) to three horizons: the operational horizon of days, the seasonal schedule, and the fleet plan of several years.

Table 3.4 Which costs a flight can avoid at each horizon (adapted from the handbook's cost buckets). "Partly" marks costs with a variable and a fixed component, such as hotel and per diem costs within an otherwise fixed crew roster.

Cost bucket	Main drivers	Days	Season	Years
Fuel	block hours, weight, winds, routing	yes	yes	yes
Crew	duty time, pairings, bases, agreements	partly	yes	yes
Airport and navigation	movements, weight, passengers, distance	yes	yes	yes
Maintenance	flight hours, cycles, scheduled checks	partly	mostly	yes
Passenger service and distribution	passengers, cabin, channel	yes	yes	yes
Station and launch	staff, contracts, marketing	no	partly	yes
Aircraft ownership or lease	fleet commitments, lease terms	no	no	when leases end
Corporate overhead	shared functions	no	no	partly

RESULT 3.4 Avoidable cost grows with the horizon

Let $\mathcal{A}(h)$ be the set of cost buckets a flight can avoid within horizon h. If $h_1 < h_2$ then $\mathcal{A}(h_1) \subseteq \mathcal{A}(h_2)$, so the avoidable cost of a flight is non-decreasing in the horizon. A flight should be operated, at horizon h, if its contribution over avoidable cost is positive and at least as large as the contribution of the best feasible alternative use of the same aircraft, crew and slots.

Proof. Any commitment that can be undone within h_1 can be undone within the longer h_2, which gives the nesting and hence the monotonicity of avoidable cost. Operating the flight changes the airline's total profit over h by its revenue minus its avoidable cost; choosing it over an alternative use of the same scarce resources changes profit by the difference of the two contributions. Both must be non-negative. ∎

The handbook puts the consequence in two sentences worth memorizing: an accounting loss is not sufficient reason to cancel a flight whose costs are mostly unavoidable and whose contribution is pos-

itive, and a positive contribution is not sufficient reason to launch one if a better use of the aircraft exists or if longer-run costs go uncovered (Network Planning Handbook 2026). Laurentian's trip-cost model (equation (3.1)) sits at the seasonal horizon: fuel, crew, airport, maintenance and passenger costs vary with the flying, while ownership and overhead do not. The opportunity cost of the aircraft, which is often the largest cost a new route imposes, is not in the model at all; it enters through the comparison of alternatives, as displacement and network contribution in Chapter 12.

> **PITFALL**
>
> Applying system-average CASK to an incremental route. Suppose Laurentian's finance team reports a fully allocated system CASK of 15.0 cents (an illustrative figure at the network's average stage length). Applied to one departure of the 182-seat A321XLR to Seville (5,556 km), it gives a cost of $0.150 \times 182 \times 5{,}556$, or CAD 151,700, and the flight appears to lose CAD 50,700 with 150 passengers paying CAD 650 each including ancillaries plus CAD 3,500 of cargo. Its avoidable trip cost is CAD 88,000, and it contributes CAD 13,000. Applied to the 137-seat A220 on Quebec City–Montreal (233 km), the same average gives CAD 4,800 against an avoidable cost of CAD 11,600 with 120 passengers: now it understates. The system average is wrong in both directions, for two reasons: it carries costs the flight cannot avoid, and it embeds the network's average stage length, which neither route has.

3.7 Laurentian's cost model at work

With the trip-cost model of equation (3.1), the break-even load of a departure follows directly. Let $\mathrm{TC}_0 = c_{\mathrm{BH}}\mathrm{BH} + c_{\mathrm{dep}}$ be its capacity cost, the trip cost with no passengers on board, f the average net fare allocated to the leg, a the ancillary revenue per passenger and K other contribution such as cargo.

RESULT 3.5 Break-even load from the trip-cost model

The departure breaks even with

$$N^* = \frac{\mathrm{TC}_0 - K}{f + a - c_{\mathrm{pax}}}$$

passengers, a **break-even load factor** of N^*/C. Equivalently, it is the capacity cost per ASK net of cargo divided by the passenger margin per RPK.

The proof is one line: contribution $(f+a-c_{\mathrm{pax}})N+K-\mathrm{TC}_0$ is zero at N^*. Dividing numerator and denominator by Cd turns the ratio into a ratio of unit measures, which is the precise version of the CASK-over-yield shortcut of Chapter 2: precise because passenger-driven costs are netted from the yield instead of being folded into CASK.

Table 3.5 Block hours (BH), capacity cost (CAD), unit cost and break-even load factor on five Laurentian routes (illustrative running-case data; average net fare in CAD). Capacity cost is $c_{\mathrm{BH}}\mathrm{BH} + c_{\mathrm{dep}}$; CASK is computed on capacity cost; the break-even load factor nets passenger costs from fares and, for Seville, adds ancillary and cargo revenue.

Route	Aircraft	BH	Capacity cost	CASK (¢)	Fare	Break-even LF
YQB–YUL	A220-300	0.9	9,200	28.8	201	37.1%
BOS–YUL	A220-300	1.4	13,200	23.6	246	44.6%
YUL–YVR	737-8	5.2	49,300	7.9	420	72.9%
YUL–CDG	787-9	6.9	113,950	6.9	675	61.2%
YUL–SVQ	A321XLR	7.0	80,500	8.0	620	70.5%

WORKED EXAMPLE 3.5 Trip costs and break-even loads on five Laurentian routes

Table 3.5 applies Result 3.5 to five Laurentian flights. Block times come from the running case (5.2 hours for Montreal–Vancouver as in Chapter 9, 7.0 hours for the eastbound Seville flight). Fares are the demand-weighted averages of the case's local fares, except for Vancouver, where CAD 420 matches the CAD 400 contribution per passenger used in Chapter 9, and for Seville, where the case gives an allocated net fare of CAD 620, ancillary revenue of CAD 30 per passenger and cargo contribution of CAD 3,500.

Take the Seville line step by step. Capacity cost is $11{,}000 \times 7.0 + 3{,}500 = 80{,}500$. Each passenger brings $620 + 30 - 50 = 600$. Net of cargo, the flight must cover $80{,}500 - 3{,}500 = 77{,}000$, so $N^* = 77{,}000/600 = 128.3$ passengers, a load factor of 70.5 percent. With 150 passengers the trip cost of equation (3.1) is $80{,}500 + 50 \times 150 = 88{,}000$, the case's avoidable cost, and the departure contributes $150 \times 650 + 3{,}500 - 88{,}000 = 13{,}000$.

The table spans break-even load factors from 37 to 73 percent. The short-haul figures are low because their fares are high per kilometre and the model leaves out the overhead that weighs most on short flights; the long-haul figures are high because fares rise much less than proportionally with distance while block-hour costs rise almost proportionally. Break-even loads are therefore not comparable across routes as performance measures, only as thresholds for each route.

Demand is uncertain, so the threshold must be compared with the expected load, not with mean demand. The case gives Seville demand per departure of $\mathcal{N}(175, 42^2)$ in the peak, $\mathcal{N}(150, 40^2)$ in the shoulder and $\mathcal{N}(110, 35^2)$ in winter. The spill formula of Chapter 9 gives expected load factors of 88.7, 79.8 and 60.3 percent, and contributions per eastbound departure of CAD 19,900, 10,100 and −11,200 (Figure 3.4). On these numbers a winter departure loses money before any network effect is counted.

The figure also shows why break-even load factors understate risk near capacity. Below about 130 passengers the two lines nearly coincide, because spill is rare. Above it, the expected-load line bends away from the deterministic one: every extra unit of mean demand adds less than one passenger, so the contribution of a strong peak season is smaller than a mean-demand calculation suggests, while the losses of a weak winter are not cushioned at all.

3.8 Air Canada's cost structure in 2026

Air Canada reports its costs in the objective style, by line of expense, together with two unit measures: CASM, operating expense per available seat mile, and adjusted CASM, which removes aircraft fuel, the ground package costs of Air Canada Vacations, freighter costs and special items (Air Canada 2026e). Table 3.6 reproduces

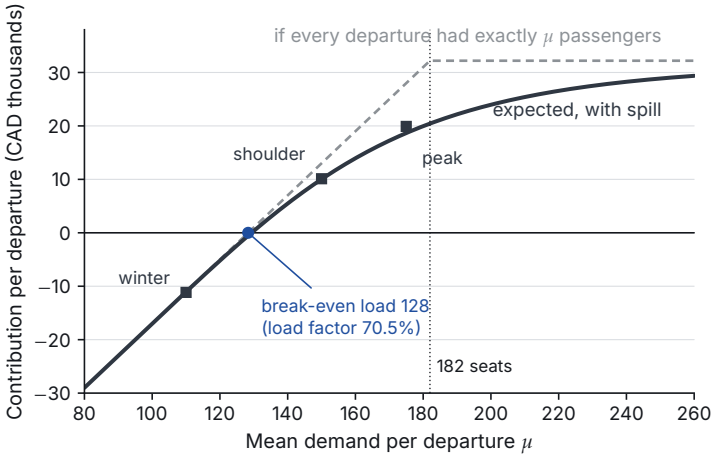


Figure 3.4 Expected contribution of an eastbound YUL–SVQ departure of Laurentian's A321XLR as mean demand grows (CAD 600 per passenger, CAD 77,000 capacity cost net of cargo, $k = 0.27$). The dashed line assumes every departure carries exactly its mean demand; the solid line accounts for spill. Squares mark the case's peak, shoulder and winter demand, each with its own standard deviation; the peak's lower k-factor (0.24) puts its square slightly above the curve.

the second quarter of 2026, converts each line to cents per ASM and adds the cost driver each line mainly follows.

Table 3.6 Air Canada's operating expenses in the second quarter of 2026 (CAD millions; 26,942 million ASMs, against 26,860 million a year earlier). Amounts and total CASM from the MD&A (Air Canada 2026e), with shortened line names (depreciation includes impairment); line-level cents per ASM computed here; the driver column is this book's interpretation.

Expense line	CAD M	Share	¢/ASM 2026	¢/ASM 2025	Main driver
Wages and benefits	1,740	26.8%	6.46	4.64	hours, departures, staff
Aircraft fuel	1,713	26.4%	6.36	4.27	block hours, price
Other	660	10.2%	2.45	2.25	handling, crews, rent
Depreciation and amortization	546	8.4%	2.03	1.83	fleet, time
Airport and navigation	446	6.9%	1.66	1.52	departures, weight
Aircraft maintenance	364	5.6%	1.35	1.27	hours, cycles
Sales and distribution	293	4.5%	1.09	1.02	passengers, revenue
Capacity purchase fees	232	3.6%	0.86	0.82	regional flying
Catering and onboard	180	2.8%	0.67	0.64	passengers
Ground package costs	164	2.5%	0.61	0.58	vacation sales
Information technology	143	2.2%	0.53	0.56	overhead
Total operating expense	6,481	100%	24.05	19.41	

Three lessons stand out. First, fuel and labour are each about a quarter of the cost base, as they were in the lecture's US data two decades earlier, and fuel is the volatile one: fuel expense rose 49 percent year over year while fuel consumption rose 0.2 percent, because the price per litre rose 51 percent, from 88.0 to 132.7 cents, despite a CAD 205 million hedging gain. Fuel's share of operating expense rose from 22.0 to 26.4 percent in a year. Second, unit cost moved for reasons that have nothing to do with the schedule: CASM rose 23.9 percent to 24.05 cents on capacity that grew 0.3 percent. Of the CAD 1,740 million in wages, CAD 360 million were one-time charges for pension amendments, lump-sum payments to pensioners and voluntary separation packages; without these and a CAD 28 million le-

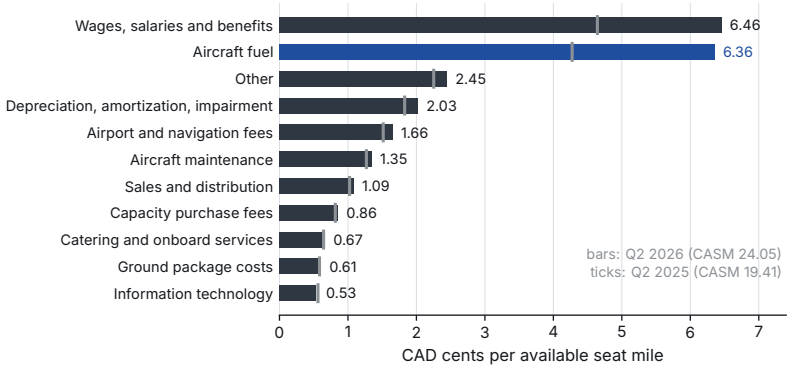


Figure 3.5 Air Canada's operating expense per available seat mile by line, second quarter of 2026 (bars) against the second quarter of 2025 (ticks), computed from the MD&A (Air Canada 2026e). Fuel is highlighted.

gal provision, CASM would have been 22.62 cents. Adjusted CASM, which also removes fuel, rose 7.4 percent to 15.47 cents, mainly on labour, maintenance, airport and navigation fees and selling costs (Air Canada 2026e). Third, several lines are not costs of Air Canada's own flying: the ground packages are the cost of holiday products sold by Air Canada Vacations, and the capacity purchase fees pay for regional flying bought from another airline. A system CASM blends all of these over an average stage length of 1,840 miles.

In kilometre terms, Air Canada's CASM was 14.9 cents per ASK and its adjusted CASM 9.6 cents. The temptation to compare these numbers with Laurentian's 8 cents of capacity cost on the Seville flight is precisely the pitfall of Section 3.6: the reported figure is fully allocated, system-wide and includes non-flying businesses; the Laurentian figure is the avoidable cost of one long flight. Neither is wrong, and they measure different things.

IN PRACTICE Cost inputs to a planning model in 2026

The fuel price was the dominant cost story of 2026. Air Canada attributes its 51 percent year-over-year increase in fuel price per litre in the second quarter to conflict in the Middle East and disruption to shipping through the Strait of Hormuz (Air Canada 2026e). IATA's June 2026 outlook assumed a fuel bill equal to about 31 percent of the industry's costs and an industry net margin of about 2 percent (IATA 2026d). When Air Canada reinstated its 2026 guidance in August, it assumed jet fuel at about CAD 1.38 per litre for

the third quarter, against about CAD 0.90 in its original plan, and adjusted CASM growth of 5 to 6 percent (Air Canada 2026d). For a planner, the consequences are practical. Fuel should enter route models as a scenario input, priced by period and burned by direction and season, not as a fixed share of cost; the handbook makes the same recommendation and adds that the definitions of avoidable cost at each horizon belong to Finance, so the first question to ask about any cost number is which costs it treats as avoidable (Network Planning Handbook 2026).

3.9 Summary

- Airlines report costs by object (what is bought) and by function (what it is for). Planning uses the functional view, because each function has an activity driver: block hours, departures, passengers, distance, calendar time or revenue.
- A flight's cost is modelled as $c_{\text{BH}}\text{BH} + c_{\text{dep}} + c_{\text{pax}}N$. Separating what the departure causes from what each passenger causes is the first step of any break-even or contribution calculation.
- With block time $\tau_0 + d/s$, capacity cost per ASK is $a/d + b$: it falls with stage length toward the cruise cost, with an elasticity equal to minus the distance-independent share of capacity cost.
- Larger aircraft cost less per seat and more per trip. A larger gauge pays only if its added seats fill, on average, with probability at least their marginal cost divided by the contribution per passenger.
- Utilization spreads ownership and other calendar costs; it rises with sector length and falls with turn time, in whole departures. The same aircraft type can cost one airline twice what it costs another.
- Avoidable cost grows with the decision horizon. System-average and fully allocated costs answer accounting questions, not incremental ones, and can err in both directions.
- Laurentian's break-even load factors range from 37 to 73 percent across its routes; the Seville flight breaks even at 128 passengers and loses money in a winter with mean demand of 110.
- Air Canada's 2026 cost base shows fuel and labour each near a quarter of costs, a 51 percent fuel price shock, and one-time charges that move CASM without any change in the schedule.

Table 3.7 Key formulas of the chapter.

Quantity	Formula
Trip cost	$\mathrm{TC} = c_{\mathrm{BH}}\mathrm{BH} + c_{\mathrm{dep}} + c_{\mathrm{pax}}N$
Block time	$\mathrm{BH}(d) = \tau_0 + d/s$
Unit cost and stage length	$\mathrm{CASK}(d) = a/d + b$, elasticity $-F/(F + vd)$
Upgauge condition	average $\Pr(D > x)$ over added seats $\geq \beta/r$
Aircraft productivity	ASK per aircraft-day $= n\,d\,C = U s_b C$
Utilization	$U = t\lfloor (H+g)/(t+g) \rfloor \approx (H+g)t/(t+g)$
Ownership per block hour	$O/(365\,U)$
Break-even load	$N^* = (\mathrm{TC}_0 - K)/(f + a - c_{\mathrm{pax}})$

3.10 Further reading

Belobaba's lecture on operating costs and productivity (Belobaba 2006e) is the source of the cost taxonomy, the activity drivers and the 737-500 comparison; its charts of unit cost against stage length are worth looking at with Result 3.1 in mind. Chapter 6 of Belobaba, Odoni and Barnhart (2015) treats airline operating costs and productivity at textbook length, and Doganis (2019, chs. 3–4) discusses airline costs and their drivers from a commercial perspective. For Air Canada, the MD&A (Air Canada 2026e) defines CASM and adjusted CASM and reconciles them. The handbook (Network Planning Handbook 2026) sets out the cost buckets by decision horizon and the warnings about system averages that Section 3.6 develops.

PART II

Demand

Every plan begins with an estimate of who wants to travel, between which places, when, and at what price. This part builds that estimate from first principles: the economics of air travel demand, the choice models that divide a market among itineraries, the econometric traps hidden in fare data, the methods that project demand forward, and the spill models that turn demand into passengers carried on a flight of fixed size.

CHAPTER

4

Air Travel Demand

IN THIS CHAPTER

- Why demand lives in origin-destination markets while capacity is sold one flight leg at a time.
- What makes a market large: income, population, price, travel time, frequency and quality.
- Price elasticity derived for linear and constant-elasticity demand, its effect on revenue, and why business and leisure travellers respond so differently.
- Total trip time, schedule delay and generalized cost: the economics of frequency.
- Seasonality and the weekly cycle, and why planners never average across them.
- Market size versus market share, and the S-curve that links frequency share to market share: model, derivation and estimation.
- The quality of service index and the traffic a new nonstop creates, worked through for Laurentian's Montreal–Seville launch.

DRAWS ON Belobaba, Airline economics review (MIT 16.75J, 2006); Belobaba, Introduction to PODS passenger choice model (MIT 16.75J, 2006); Boeing, Load factor analysis (1978); Boeing, Decision window path preference methodology (1993); handbook

A network planner never sells a seat, but every decision a planner makes is a forecast of who will buy one. Opening Montreal–Seville, adding a tenth daily flight to Toronto, or moving the Paris departure by an hour are bets on how many people want to travel, what they will pay, and which itinerary they will choose once the competition has reacted. This chapter supplies the economics behind those bets. It treats demand the way an industrial economist would, as a function of prices, product attributes and market structure, but it uses the units, quirks and rules of thumb of the airline business.

Two ideas organize the chapter. The first is that demand is generated by markets, pairs of places that people want to travel between,

while airlines supply flights, and one flight serves many markets at once. The second is that an airline's traffic in a market is the product of two things it influences differently: the size of the market, which responds to the price and the total time of travel, and the airline's share of it, which responds to frequency, timing and quality relative to competitors. The S-curve, the quality of service index and the stimulation created by a new nonstop all live at the junction of the two.

The rest of Part II refines each piece. Chapters 5 and 6 derive the itinerary choice models that this chapter uses in reduced form, Chapter 7 explains why elasticities cannot be read off airline data naively, Chapter 8 projects demand forward, and Chapter 9 turns demand into passengers carried when capacity is limited.

4.1 Markets, itineraries and legs

Ask how many people fly between Montreal and Paris and you must first decide what you are counting. Belobaba distinguishes three definitions (Belobaba 2006d). A **city-pair market** counts travel between two cities, whatever the airports used. An **airport-pair market** splits it by airport, which matters where a city has several, such as London or New York. A **region-pair market** widens the catchment to whole metropolitan regions, including secondary airports that compete for the same travellers. Laurentian's Montreal traffic is concentrated at one airport, so for most of its markets the three definitions coincide; for London or New York they do not.

DEFINITION 4.1 O&D market, itinerary and leg demand

An **origin-destination (O&D) market** is the set of people who wish to travel from origin A to destination B in a given period. Its size is counted in one-way passenger trips, often expressed as **passengers per day each way (PDEW)**. An **itinerary** (or path) is a sequence of flight legs, with times, that takes a passenger from A to B. **Leg demand** is the number of passengers who want a seat on a given flight leg, summed over every itinerary that uses it:

$$D_l = \sum_{j \in J} a_{lj} D_j, \tag{4.1}$$

where D_j is the demand for O&D product j and $a_{lj} = 1$ if product j uses leg l.

Each direction is a separate market: Montreal–Paris and Paris–Montreal have different travellers, seasons and points of sale, even if many trips are round trips that appear in both. PDEW also hides a trap. The handbook's example is a market of 73,000 annual passengers in both directions: dividing by 365 days and two directions gives 100 PDEW, yet a new daily flight should not expect 100 passengers on every departure, because demand varies by season and day of week and is shared among itineraries and carriers (Network Planning Handbook 2026).

Belobaba calls the gap between how demand arises and how capacity is produced the **dichotomy of demand and supply** (Belobaba 2006d). Demand is generated market by market; supply is a set of leg departures, each of which offers joint supply to many markets. In his example of a Boston–Houston flight, the local Boston–Houston market was the largest single group of passengers, yet most of the flight's bookings continued beyond Houston to more than twenty other destinations. Two consequences follow. The seats on a flight are not the supply to any one market, so one cannot say whether a market is in equilibrium or whether serving it is profitable without allocating costs and revenues by assumption. And one market is served by many competing itineraries, so measuring its total traffic requires ticket-level data, not flight counts.

Table 4.1 Mean unconstrained demand per departure in Laurentian's four-leg bank at YUL, by O&D market (rows) and leg (columns). Each leg's demand is a column sum, as in (4.1); each connecting market appears on two legs. All numbers are illustrative.

O&D market	YQB–YUL	BOS–YUL	YUL–CDG	YUL–SVQ
YQB–YUL	73			
BOS–YUL		70		
YUL–CDG			215	
YUL–SVQ				112
YQB–CDG	48		48	
YQB–SVQ	20			20
BOS–CDG		54	54	
BOS–SVQ		29		29
Leg demand	141	153	317	161
Seats	137	137	298	182

Table 4.1 shows the dichotomy on Laurentian's small bank network. The YUL–CDG leg has mean demand of 317 passengers per departure for 298 seats, but only 215 of them, 68 percent, are Montreal–Paris travellers; the rest connect from Quebec City and Boston. The two feeder legs are the opposite case: only about half of their demand is local. Remove the Boston flight and the Paris flight loses 54 passengers of demand, even though nothing changed in the Montreal–Paris market.

> **PITFALL**
>
> Reading a leg's load as the demand of its city pair. A full YUL–CDG flight says that the leg is constrained; it says nothing direct about the Montreal–Paris market, part of which flies other carriers and other paths, and much of the flight's own demand belongs to other markets.

4.2 What drives demand

Air travel is a derived demand: people fly to do business, to take a holiday or to see family, and the trip is worth taking only if the purpose is worth more than its total cost. The determinants fall into four groups (Belobaba 2006d):

- **market characteristics**: the population and income of the two regions, the business ties between them, the tourist appeal of each end, and migration links that generate visits to friends and relatives;

- **the price** of travel, meaning the fares actually available rather than a published fare grid;
- **the time** the journey takes, from door to door, including the wait imposed by an inconvenient schedule;
- **the product**: frequency and departure times, the number of stops and connections, aircraft, cabin, reliability and loyalty programs, each relative to what competitors and other modes offer.

The first group sets the scale of a market and changes slowly with the economy. Air travel grows faster than income in normal times: IATA's survey of the evidence reports income elasticities consistently above one, around 1.5 for short-haul routes in developed economies and higher for long-haul travel (Smyth and Pearce 2008). The other three groups are what airlines control, and they are the subject of this chapter.

Travellers differ in how they weigh these factors, and planners group them into segments. **Business** travellers make trips required by their work; the employer usually pays, the date and time are tied to a meeting, and the trip is rarely cancelled because of the fare. **Leisure** travellers pay their own money, choose whether, when and where to go, and will accept an inconvenient schedule for a lower fare. **Visiting friends and relatives (VFR)** travellers sit between the two: price sensitive like leisure travellers, but tied to a particular destination and often to fixed dates such as holidays and family events. A Montreal–Paris flight carries all three.

The workhorse model of market demand is multiplicative (Belobaba 2006d):

$$D = M\,P^{a}\,T^{b}, \tag{4.2}$$

where D is O&D demand per period, M a market-size constant that captures population and the interaction between the two cities, P the average fare, T the total trip time, and $a < 0$ and $b < 0$ the price and time elasticities. Taking logarithms gives $\ln D = \ln M + a \ln P + b \ln T$, which can be estimated by regression on a market's history of demand, fares and schedules, with the serious caveat that fares and frequencies respond to demand (Chapter 7).

The same structure extends to segments. In Belobaba's version, business demand for the coach cabin is

$$D_{cb} = M_b\,I_c\,P_c^{a_1}\,T_c^{b_1}\,P_f^{c_1},$$

where M_b sizes the business segment, I_c is an image factor for the cabin, a_1 and b_1 are the business price and time elasticities, and $c_1 > 0$ is the cross-elasticity with respect to the first-class fare P_f. Leisure

demand for coach and discount products has the same form with its own parameters. Segmenting matters because the elasticities differ, which is the subject of the next section.

4.3 Price elasticity

DEFINITION 4.2 Price elasticity of demand

The **price elasticity** of demand is the percentage change in demand caused by a one percent increase in price, other things equal:

$$\eta = \frac{\partial D}{\partial P}\frac{P}{D} = \frac{\partial \ln D}{\partial \ln P}.$$

For a discrete change from (P_0, D_0) to (P_1, D_1) the **arc elasticity** $\ln(D_1/D_0)/\ln(P_1/P_0)$ is the natural counterpart.

Demand is **inelastic** when $-1 < \eta < 0$ and **elastic** when $\eta < -1$. Two functional forms dominate practice, and they behave very differently away from the point where they are calibrated.

RESULT 4.1 Elasticity and revenue for the two standard demand curves

1. For linear demand $D = A - BP$ with $A, B > 0$, the elasticity is
$$\eta(P) = \frac{-BP}{A - BP},$$
which falls from 0 at $P = 0$ to $-\infty$ at the choke price A/B and equals -1 at the midpoint $P = A/(2B)$.
2. For constant-elasticity demand $D = MP^a$, the elasticity is a at every price.
3. For any demand curve, revenue $R = PD$ satisfies

$$\frac{\mathrm{d}R}{\mathrm{d}P} = D\,(1 + \eta), \qquad \frac{\mathrm{d}\ln R}{\mathrm{d}\ln P} = 1 + \eta, \tag{4.3}$$

so a small fare increase raises revenue if and only if $|\eta| < 1$. Linear demand therefore has its revenue maximum at the midpoint, while constant-elasticity demand has none: revenue is $R \propto P^{1+a}$, rising in price when $a > -1$ and falling when $a < -1$.

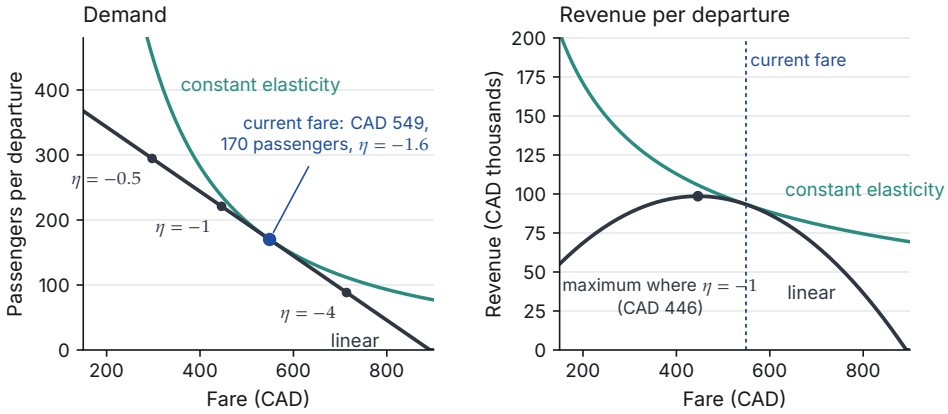


Figure 4.1 Linear and constant-elasticity demand calibrated to the same point, Laurentian's YUL–CDG restricted fare of CAD 549 with mean demand of 170 passengers per departure and elasticity -1.6. Left: the linear curve becomes more elastic as the fare rises; the constant-elasticity curve keeps $\eta = -1.6$ everywhere. Right: revenue per departure. The linear model says revenue would peak at a fare of CAD 446; the constant-elasticity model says every cut raises revenue, which is plausible only near the current fare.

Proof. For the linear curve, $\partial D/\partial P = -B$, so $\eta = -BP/D = -BP/(A - BP)$; setting $\eta = -1$ gives $BP = A - BP$. For the constant-elasticity curve, $\ln D = \ln M + a \ln P$ and the derivative of $\ln D$ with respect to $\ln P$ is a. For revenue, $\mathrm{d}R/\,\mathrm{d}P = D + P\,\mathrm{d}D/\,\mathrm{d}P = D\,(1+\eta)$, and dividing by $R/P = D$ gives the logarithmic form. ∎

Figure 4.1 is a warning about extrapolation. Both curves agree on the elasticity at the current fare, so for a small fare change they give the same answer. For large changes they disagree, and neither is right by construction: the linear curve has a choke price above which nobody flies, while the constant-elasticity curve with $|a| > 1$ implies that revenue grows without limit as the fare falls to zero. Treat an elasticity as a local property of a market at its current fares.

Belobaba's lecture uses the rule of thumb that a 10 percent fare increase changes demand by ten times the elasticity, in percent (Belobaba 2006d). That is a linear approximation; under constant elasticity the exact change is $1.1^{\eta} - 1$. The difference is small for small changes and grows with $|\eta|$, as the next example shows.

WORKED EXAMPLE 4.1 A 10 percent fare increase on YUL–CDG

Laurentian sells two products on YUL–CDG: a flexible fare of CAD 1,149 with mean demand of 45 passengers per departure, bought mostly by business travellers, and a restricted fare of CAD 549 with mean demand of 170, bought mostly by leisure travellers (Table 4.1 combines them). Use the lecture's illustrative elasticities: -0.8 for the business-dominated product and -1.6 for the leisure one.

Flexible fare. The fare rises to CAD 1,263.90. Under constant elasticity demand falls to $45 \times 1.1^{-0.8} = 41.70$ passengers, a decrease of 7.3 percent (the linear rule of thumb says 8 percent). Revenue per departure rises from CAD 51,705 to CAD 52,700 ($41.70 \times 1{,}263.90$), an increase of 1.9 percent, as (4.3) predicts for $|\eta| < 1$.

Restricted fare. The fare rises to CAD 603.90 and demand falls to $170 \times 1.1^{-1.6} = 145.96$, a decrease of 14.1 percent (16 percent by the rule of thumb). Revenue falls from CAD 93,330 to CAD 88,143, a loss of 5.6 percent.

The planner learns three things. Segment elasticities are the economic case for charging the two groups different prices, which is the subject of Chapter 17. The direction of the revenue effect of a small fare change depends only on whether $|\eta|$ is above or below one, but its size depends on the functional form: the linear model puts the leisure revenue loss at 7.6 percent rather than 5.6 percent. And the arithmetic assumes that capacity does not bind; on a flight that spills, a fare increase loses fewer carried passengers than it loses demand (Chapter 9).

Why segments differ Business travellers are less price elastic than leisure travellers for reasons that are easy to list: someone else pays, the trip has a fixed purpose and date, and there are few substitutes for being in Paris on Tuesday. Leisure travellers can change destination, date or mode, or stay home. Section 4.4 adds a sharper explanation: the fare is a smaller share of a business traveller's total cost of the trip, once the value of time is counted.

The published evidence agrees in direction and varies in magnitude. Belobaba's lecture, summarizing studies available in 2006, describes business demand as slightly inelastic, leisure demand as clearly elastic, and typical market elasticities between -0.8 and -2.0 (Belobaba 2006d); these values are historical. IATA's review of the

literature reports the median estimates of a survey prepared for Canada's Department of Finance by Gillen, Morrison and Stewart (2004): about −0.3 for long-haul international business travel, −0.7 for short-haul business, −1.0 for long-haul international leisure and −1.5 for short-haul leisure (Smyth and Pearce 2008). Short-haul travel is more elastic because car and rail are substitutes.

Market, route and carrier elasticities An elasticity is only defined once you say what is held fixed. IATA's study for its member airlines recommends −1.4 for a fare change on a single route or market, −0.8 when fares change across a whole country, as with a national tax, and −0.6 across a group of countries, with a multiplier of 1.1 for short-haul routes and 1.2 for trans-Atlantic markets (InterVISTAS Consulting 2007). The pattern is not a contradiction: a fare increase on one route diverts travellers to other routes and airports, and that diversion nets out as the change becomes more general.

The same logic separates a market's elasticity from one carrier's. Suppose carrier a, with market share s_a, raises its fare by one percent and nobody follows. The market shrinks by about $s_a\,|\eta_M|$ percent, because a's fare has a weight of about s_a in the market's average fare, and a loses share to its competitors. In a logit share model with fare coefficient $\beta_P < 0$ (Chapter 5), the share response is $\beta_P P_a(1-s_a)$, so

$$\eta_a \approx s_a\,\eta_M + (1-s_a)\,\beta_P P_a.$$

With a market elasticity of −1.0, a share of 40 percent and $\beta_P P_a = -3$, the carrier faces $\eta_a \approx -2.2$. A carrier contemplating its own fare should use its own elasticity; a planner forecasting how a new nonstop will grow the whole market should use the market's.

PITFALL

Importing one elasticity into every decision. The handbook warns against carrying one price elasticity, or one frequency-share curve, from market to market (Network Planning Handbook 2026). An elasticity belongs to a market, a segment, a level of aggregation and a price range, and estimates from fare data are biased toward zero unless the endogeneity of fares is handled (Chapter 7).

4.4 Time, frequency and generalized cost

Next to price, the total time a trip takes is the most important determinant of demand (Belobaba 2006d). Time has several components, and frequency enters through one of them.

DEFINITION 4.3 Total trip time and schedule delay

The **total trip time** from origin to destination is

$$T = t_{\text{fixed}} + t_{\text{flight}} + t_{\text{sd}},$$

where t_{fixed} covers access to and egress from the airports and processing at both ends, t_{flight} is block time plus any connection time, and t_{sd} is the **schedule delay** (or schedule displacement): the gap between when the traveller would like to leave and when a flight actually leaves.

Schedule delay is what frequency buys. The concept has a long history in transport economics; Douglas and Miller split it into the frequency delay imposed by the timetable and the stochastic delay suffered when the preferred flight is full (Douglas and Miller 1974). The second part is spill seen from the passenger's side (Chapter 9). Belobaba models the first as K/f, a constant divided by the daily frequency (Belobaba 2006d). The constant can be derived.

RESULT 4.2 Schedule delay and the frequency elasticity of demand

Suppose desired departure times are spread uniformly over an operating day of H hours and the f daily flights leave at the midpoints of f equal sub-periods. A traveller who takes the nearest flight has expected schedule delay $H/(4f)$. One who cannot leave earlier than desired has expected delay $H/(2f)$ if each flight leaves instead at the end of its sub-period, so that every desired time has a later flight the same day. Hence $t_{\text{sd}} = K/f$ with $K = H/4$ in the first case. If demand follows (4.2) with $T = t_0 + K/f$, where $t_0 = t_{\text{fixed}} + t_{\text{flight}}$, then

$$\frac{\partial \ln D}{\partial \ln f} = |b| \, \frac{K/f}{T}. \tag{4.4}$$

In words, the frequency elasticity of market demand equals the absolute time elasticity times the share of schedule delay in total trip time.

Proof. With headway $h = H/f$, each sub-period of length h contains one flight at its centre, and a desired time uniform on the sub-period is at a distance from the flight that is uniform on $[0, h/2]$, with mean $h/4 = H/(4f)$. If only later flights are acceptable and each flight leaves at the end of its sub-period, the wait to the next flight is uniform on $[0, h)$ with mean $h/2$; with departures at the midpoints, travellers who want to leave after the last flight would have no flight that day. For the elasticity, $\partial \ln D/\partial \ln f = b\,(\partial \ln T/\partial \ln f) = b\,(f/T)(-K/f^2) = -bK/(fT)$, which is positive because $b < 0$. ∎

An 18-hour operating day gives $K = 4.5$ hours. Equation (4.4) explains four observations from Belobaba's lecture at once. Nonstops beat connections because they cut t_{flight}. More frequency raises market demand because it cuts t_{sd}. Frequency matters more on short-haul routes, where schedule delay is a large share of the trip. And several connecting departures through a hub can serve the average traveller better than a single nonstop a day, because the saving in schedule delay can exceed the extra flight and connection time. The model also has a **saturation frequency** built in: the marginal effect of the f-th flight falls roughly as $1/f^2$, so beyond some point extra flights add almost nothing to the market and only redistribute it.

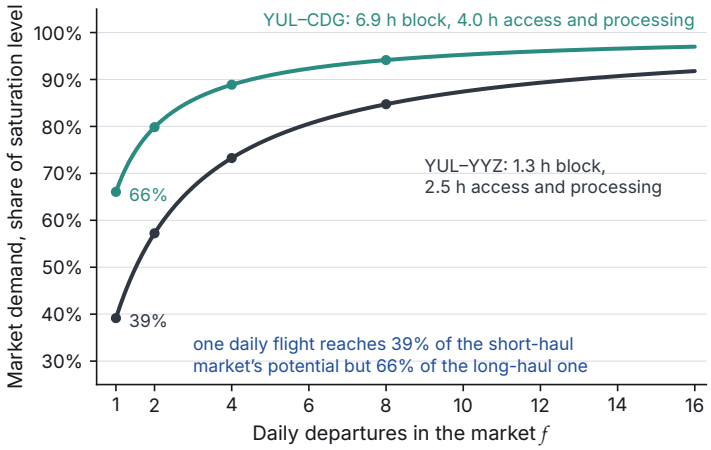


Figure 4.2 Market demand as a share of its saturation level (the demand with zero schedule delay) as daily frequency grows, from (4.2) with time elasticity $b = -1.2$ and $K = 4.5$ hours. The short-haul market needs many flights to approach its potential; the long-haul market is close to it with two or three.

WORKED EXAMPLE 4.2 **One more flight to Toronto or to Paris**

Take a time elasticity of $b = -1.2$, in the middle of the historical range of -0.8 to -1.6 that Belobaba reports, and $K = 4.5$ hours. For YUL–YYZ assume 2.5 hours of access and processing and 1.3 hours of block time; for YUL–CDG, 4.0 hours and 6.9 hours.

Toronto. With 8 daily flights in the market, $T = 2.5 + 1.3 + 4.5/8 = 4.36$ hours, of which schedule delay is 12.9 percent. The frequency elasticity is $1.2 \times 0.129 = 0.155$. A ninth flight cuts T to 4.30 hours and raises market demand by $(4.30/4.36)^{-1.2} - 1 = 1.7$ percent.

Paris. With one daily flight, $T = 4.0 + 6.9 + 4.5 = 15.4$ hours and schedule delay is 29 percent of it, so the elasticity is 0.351. A second daily flight cuts T to 13.15 hours and raises demand by 20.9 percent.

The comparison is not between short and long haul as such but between frequencies: the second Paris flight is worth more than the ninth Toronto flight, while the second Toronto flight would raise that market by 46 percent. Figure 4.2 shows the general pattern. With one daily flight the Toronto market reaches only 39 percent of the demand it would have with continuous ser-

vice, Paris 66 percent; at four flights the figures are 73 and 89 percent.

Two caveats. On transatlantic routes the acceptable departure window is narrow, because eastbound flights must leave in the evening to arrive in Europe in the morning, so H is shorter and frequency matters even less than the model suggests (Chapter 13). And below daily service, delay is measured in days: with n evenly spread weekly flights, a traveller who shifts to the nearest operating day loses on average $7/(4n)$ days, 10.5 hours for a four-times-weekly service. Many leisure travellers shift a whole trip by a day at little cost, which is why long-haul leisure routes can be launched below daily.

Generalized cost Price and time can be combined into a single measure. The **generalized cost** of a trip is

$$G = P + v\,T + \text{other disutilities},$$

where v is the traveller's value of time and the other terms price the inconvenience of fare restrictions, connections and unwanted departure times. The handbook lists the same ingredients as the basis of passenger choice: fare, elapsed time, departure-time inconvenience, connections, reliability, product and loyalty (Network Planning Handbook 2026). The MIT passenger origin-destination simulator (PODS) makes the idea operational: each simulated passenger draws a maximum willingness to pay and disutility costs, discards itineraries priced above that maximum, and books the path and fare with the lowest total of fare plus disutilities (Belobaba 2006g). Table 4.2 reproduces its illustrative mean parameters, historical values from 2006 that show the structure of the two segments rather than current magnitudes.

Table 4.2 Mean willingness to pay and disutility costs of fare restrictions for the two passenger types of the PODS simulator, with a four-fare example structure (adapted from Belobaba, 2006). Individual values are drawn around these means. PODS also adds a path-quality cost proportional to an index of 1 for a nonstop and 3 for a connection.

Mean parameter (USD, 2006)	Business	Leisure
Maximum willingness to pay	1,200	300
Saturday-night minimum stay	450	350
Non-refundable ticket	150	50
Change fee	150	50

Generalized cost explains why the two segments trade price against time differently.

RESULT 4.3 Price and time elasticities from generalized cost

If demand depends on price and time only through generalized cost, $D = M\,G^{\eta_G}$ with $G = P + vT$, then

$$\eta_P = \eta_G \frac{P}{G}, \qquad \eta_T = \eta_G \frac{vT}{G}, \qquad \eta_P + \eta_T = \eta_G .$$

Proof. $\partial \ln D / \partial \ln P = \eta_G \, (\partial \ln G / \partial \ln P) = \eta_G \, P/G$, and likewise for T with $\partial G / \partial T = v$. The shares P/G and vT/G sum to one. ∎

Travellers with a high value of time spend a small share of generalized cost on the fare, so their price elasticity is small and their time elasticity large. Suppose, for illustration, that business travellers value time at CAD 90 an hour and leisure travellers at CAD 25, and that both face $\eta_G = -1.6$ on YUL–CDG with one daily flight ($T = 15.4$ hours). For a business traveller on the CAD 1,149 fare, generalized cost is CAD 2,535 ($1{,}149 + 90 \times 15.4$) and the fare is 45 percent of it: $\eta_P = -0.73$ and $\eta_T = -0.87$. For a leisure traveller on the CAD 549 fare, generalized cost is CAD 934 ($549 + 25 \times 15.4$) and the fare is 59 percent of it: $\eta_P = -0.94$ and $\eta_T = -0.66$. Leisure trips are also more discretionary, so their η_G is larger in absolute value, which pushes their price elasticity further out. Belobaba's qualitative summary, business demand price inelastic but time elastic and leisure demand the reverse, is what Result 4.3 predicts (Belobaba 2006d).

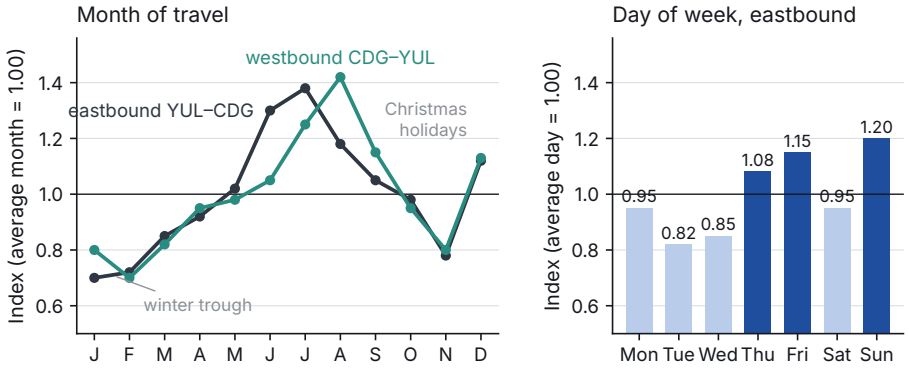


Figure 4.3 An illustrative demand profile for Laurentian's Montreal–Paris market. Left: monthly indices by direction. Eastbound demand peaks in July as Montreal residents leave on holiday; westbound demand peaks in August as they return and French visitors arrive; both directions share a winter trough and a Christmas bump. Right: the eastbound day-of-week index, with Thursday, Friday and Sunday above average. The numbers are invented for teaching.

4.5 Seasonality and the weekly cycle

Demand for "the Paris flight" is not one number. It moves with the month, the day of the week, the time of day and the direction of travel, and it jumps around holidays. Boeing's 1978 guide to load factor analysis insists that data be separated along exactly these lines before any statistics are computed: one direction at a time, one season at a time, similar days of the week together, holiday peaks removed (Boeing Commercial Airplane Company 1978). The reason is that predictable variation and random variation call for different responses. The planner answers predictable variation with the schedule, varying frequency and aircraft by season and day; revenue management absorbs part of the remaining random variation; spill models price what is left (Chapter 9).

A convenient model is multiplicative. Expected demand on day t is

$$\mathbb{E}[D_t] = \bar{D}\, S_{m(t)}\, W_{w(t)},$$

where $\bar{D}$ is the average daily demand, S_m the index of month m and W_w the index of weekday w, each normalized so that their averages are one. Real profiles add holiday effects and interactions, since the weekly pattern of a ski charter season differs from that of a business-heavy autumn.

Figure 4.3 shows a profile typical of a leisure-heavy transatlantic market; its numbers are invented, not measured. Three features matter for planning. The swing is large: if an average departure has demand of 250 passengers, a Sunday in July has $250 \times 1.38 \times 1.20 = 414$

and a Tuesday in January $250 \times 0.70 \times 0.82 = 144$, a ratio of almost three. No single aircraft fits both, which is why airlines change frequency and gauge between the IATA summer and winter seasons (Chapter 13). The directions are not symmetric: the eastbound peak comes about a month before the westbound peak, and an aircraft must fly both, so the binding direction changes during the summer. And averages mislead. An annual average PDEW, or an annual average load factor, mixes departures that spill heavily with departures that fly half empty, which is precisely the information that capacity decisions need.

Time of day adds a third layer. In many markets, departures early in the morning and late in the afternoon attract the most travellers, but rotations, turn times and time zones prevent every flight from leaving at a peak; transatlantic flights from North America leave in the evening so that they reach Europe in the morning (Belobaba 2006f). How travellers choose among departures at different times is the subject of the decision window models of Chapter 6.

4.6 Market size, market share and the S-curve

An airline's traffic in a market is the product of two quantities:

$$D_a = \mathrm{MS}_a \times D_M,$$

where D_M is total market demand and MS_a is airline a's market share. They respond to different levers. Market size responds to the general level of fares and to the best trip time available, the effects of Sections 4.3 and 4.4. Share responds to an airline's offer relative to its competitors: frequency, departure times, path quality, fares and product (Belobaba 2006d). Frequency acts on both: it adds to market size by cutting schedule delay, and it takes share from competitors.

> **DEFINITION 4.4 Frequency share and market share**
>
> In a market served by nonstop carriers $b = 1, \ldots, n$ with daily frequencies f_b, the **frequency share** of carrier a is $\mathrm{FS}_a = f_a / \sum_b f_b$, and its **market share** MS_a is its fraction of the market's passengers.

The rule of thumb is that, other things equal, market share is approximately equal to frequency share. The empirical pattern is an **S-curve**: carriers with a high frequency share win a disproportionately high market share, and carriers with a low frequency share a

disproportionately low one. Belobaba's explanation is that the carrier with more flights captures every traveller who wants to leave at a time when only it flies, and shares the travellers who want to leave when both fly (Belobaba 2006d). The lecture's model is

$$\mathrm{MS}_a = \frac{\mathrm{FS}_a^{\alpha}}{\sum_b \mathrm{FS}_b^{\alpha}},$$

with an exponent α greater than one, generally between 1.3 and 1.7 according to the lecture. The formula is a logit model in disguise, and seeing it that way explains what α means and how to estimate it.

RESULT 4.4 The S-curve as a logit model in log frequency

Let each traveller's utility for carrier a be $U_{na} = \alpha \ln f_a + \gamma' x_a + \varepsilon_{na}$, where x_a collects the carrier's other attributes and the ε_{na} are independent extreme-value errors. Then

$$\mathrm{MS}_a = \frac{f_a^{\alpha} e^{\gamma' x_a}}{\sum_b f_b^{\alpha} e^{\gamma' x_b}},$$

which reduces to the S-curve when the x_b are equal. Moreover:

1. log odds are linear in log frequency:

$$\ln \frac{\mathrm{MS}_a}{\mathrm{MS}_b} = \alpha \ln \frac{f_a}{f_b} + \gamma'(x_a - x_b); \tag{4.5}$$

2. in a duopoly the curve passes through $(\frac{1}{2}, \frac{1}{2})$ with slope α there, lies below the diagonal for $\mathrm{FS}_a < \frac{1}{2}$ and above it for $\mathrm{FS}_a > \frac{1}{2}$ when $\alpha > 1$, and the reverse when $\alpha < 1$;
3. if each of a carrier's f_a flights is a separate, equally attractive alternative in a plain logit, then $\alpha = 1$; if instead the flights of one carrier form a nest with log-sum parameter $\lambda \le 1$, then $\alpha = \lambda$.

Proof. The logit formula (derived in Chapter 5) gives $\mathrm{MS}_a = e^{V_a} / \sum_b e^{V_b}$ with $V_a = \alpha \ln f_a + \gamma' x_a$, and $e^{\alpha \ln f_a} = f_a^{\alpha}$. Dividing numerator and denominator by $(\sum_b f_b)^{\alpha}$ turns frequencies into frequency shares. Part 1 is the ratio of two shares. For part 2, write $x = \mathrm{FS}_a$ and $g(x) = x^{\alpha}/(x^{\alpha} + (1-x)^{\alpha})$, so that $\ln[g/(1-g)] = \alpha \ln[x/(1-x)]$: the log odds of market share are α times those of frequency share, which

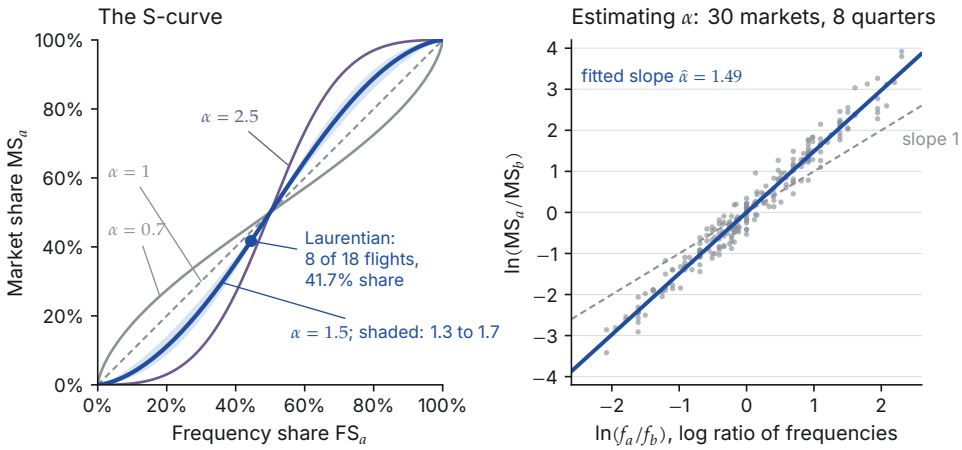


Figure 4.4 Left: the S-curve for a duopoly. The diagonal is $\alpha = 1$; the shaded band covers the lecture's range of 1.3 to 1.7, with 1.5 drawn in cobalt; $\alpha = 2.5$ shows a steep S and $\alpha = 0.7$ the inverse shape. Right: estimating α by regressing log share ratios on log frequency ratios, (4.5), in a simulated panel of 30 duopoly markets observed for 8 quarters with true $\alpha = 1.5$.

pushes g away from $\frac{1}{2}$ when $\alpha > 1$ and toward it when $\alpha < 1$. Differentiating, $g'(x) = \alpha g(1-g)/[x(1-x)]$, which equals α at $x = \frac{1}{2}$. For part 3, if all f_a flights of carrier a have utility V_a plus independent errors, the probability of choosing one of them is $f_a e^{V_a} / \sum_b f_b e^{V_b}$, so shares are proportional to frequency. In a nested logit with one nest per carrier, the nest's inclusive value is $I_a = \ln(f_a e^{V_a/\lambda})$ and the choice among carriers depends on $\lambda I_a = \lambda \ln f_a + V_a$ (Chapter 6), so frequency enters with coefficient λ. ■

Part 3 is the key to interpreting α. A carrier whose flights were merely more copies of the same product would gain share in proportion to frequency ($\alpha = 1$), or less than in proportion if its flights were close substitutes for one another ($\alpha < 1$). An exponent above one means that each flight becomes more attractive when the carrier has more of them: V_a itself rises with $\ln f_a$, with slope $\alpha - 1$. Several mechanisms produce that effect. A carrier with more flights covers more of the time windows in which travellers are willing to leave, as Boeing's decision window model formalizes: when only one carrier has a flight inside a traveller's window, it wins that traveller outright (Boeing Commercial Airplane Group 1993). Frequent travellers value the option to change to an earlier or later flight, corporate contracts and loyalty programs favour the dominant carrier at a city, and a carrier with many flights is more visible to buyers. None of these mechanisms is equally strong in every market, so α is a market-specific parameter, not a constant of nature.

Estimation Equation (4.5) turns the estimation of α into a linear regression. Collect, for a set of nonstop markets and periods, each carrier's frequency and its share of O&D passengers; pick a reference carrier in each market; regress the log share ratio on the log frequency ratio and on differences in fares, departure times and product. Market size cancels from the ratio, which is why the S-curve can be estimated without knowing it. The right panel of Figure 4.4 shows the method on simulated data: 240 market-quarters, with shares observed from a finite number of passengers and a preference shock of standard deviation 0.25 in the log odds. The fitted slope is $\hat{\alpha} = 1.49$ with a standard error of 0.02, against a true value of 1.5. Real estimates are far less precise, and they face a problem that the simulation leaves out.

PITFALL

Treating frequency as exogenous. Carriers add flights where they expect to be strong, so frequency ratios are correlated with the unobserved preference shocks in (4.5). If the log frequency ratio is $z + \kappa u$, where u is the share shock with variance σ_u^2 and z an independent component with variance σ_z^2, least squares converges to

$$\alpha + \frac{\kappa\,\sigma_u^2}{\sigma_z^2 + \kappa^2 \sigma_u^2}$$

instead of α. With $\alpha = 1.5$, $\sigma_u = 0.3$, $\sigma_z = 0.6$ and $\kappa = 1$, the estimate is 1.70: an S-curve steeper than the truth, which would make every added frequency look more valuable than it is. Instruments for frequency, such as aircraft availability or cost shifters, are the remedy (Chapter 7).

WORKED EXAMPLE 4.3 Frequency competition on YUL–YYZ

Suppose the Montreal–Toronto market has 2,400 PDEW, Laurentian operates 8 daily flights against a competitor's 10, and $\alpha = 1.5$. Laurentian's frequency share is $8/18 = 44.4$ percent and its market share is

$$\frac{0.444^{1.5}}{0.444^{1.5} + 0.556^{1.5}} = 41.7 \text{ percent},$$

1,001 PDEW, or 125 per daily flight.

Laurentian adds two flights and matches the competitor at 10. Its share rises to 50 percent. Total frequency rises from 18 to 20, which by Result 4.2 (with the Toronto parameters of Example 4.2) grows the market by 0.75 percent to 2,418 PDEW. Laurentian's traffic rises to 1,209 PDEW: the competitor falls from 1,399 to 1,209, so 190 of the 208 extra passengers come from the competitor and only 18 from market growth. Each added flight brings 104 PDEW, less than the 125 average of the existing ones, but well above the 71 that the proportional rule ($\alpha = 1$) would give.

Now the competitor responds with two flights of its own, to 12. Laurentian's frequency share falls to 45.5 percent and its market share to 43.2 percent, so with a market of 2,433 PDEW it carries 1,051. Its two extra flights now add 50 PDEW between them, a quarter of the first estimate, and both carriers fly more seats in a market that has barely grown. The planner's question is never only what a frequency is worth today, but what it is worth after the competitor's best response.

The example is a small game, and it has a clean equilibrium. Let two carriers divide market revenue R according to the S-curve, and let each daily frequency cost c. Carrier 1's profit is $Rf_1^\alpha/(f_1^\alpha + f_2^\alpha) - cf_1$.

RESULT 4.5 Symmetric frequency equilibrium

For $0 < \alpha < 2$, the symmetric Nash equilibrium of the frequency game has each carrier operating $f^* = \alpha R/(4c)$ flights, and the two carriers together spend $\alpha R/2$ on frequency.

Proof. The first-order condition is $R\alpha f_1^{\alpha-1} f_2^\alpha/(f_1^\alpha + f_2^\alpha)^2 = c$; at $f_1 = f_2 = f$ it reads $R\alpha/(4f) = c$. The share function has a negative second derivative at the symmetric point for every $\alpha > 0$, so the point is a local maximum of each carrier's profit. Against $f_2 = f^*$, carrier 1's marginal profit has the sign of $4x^{\alpha-1}/(1 + x^\alpha)^2 - 1$ with $x = f_1/f^*$, which rises and then falls in x (or only falls when $\alpha \le 1$), so any other stationary point is a local minimum at a lower frequency. The alternative of not flying earns zero, while the equilibrium earns $R/2 - cf^* = R(2 - \alpha)/4 > 0$ when $\alpha < 2$. Total frequency cost is $2cf^* = \alpha R/2$. ∎

With $\alpha = 1$, competition dissipates half of market revenue in frequency costs; with α between 1.3 and 1.7, between 65 and 85 per-

cent. This is the economic logic behind Belobaba's observation that competing carriers tend to match frequencies in nonstop markets to protect their shares (Belobaba 2006d), and behind the tendency of frequency competition to overshoot what travellers would choose to pay for.

4.7 Quality of service and the traffic a new nonstop creates

The S-curve handles competing nonstops. Most markets are also served by connections, and a planner needs a share model that compares a daily one-stop through Madrid with a four-times-weekly nonstop.

> **DEFINITION 4.5 Quality of service index**
>
> A **quality of service index (QSI)** assigns each itinerary i in a market a score q_i from its service level (nonstop, one-stop, online or interline connection) and adjustments for elapsed time, departure time, aircraft and carrier presence, multiplied by its number of departures in the period. Each itinerary's predicted share of the market is its score divided by the sum of all scores, $q_i / \sum_k q_k$, and a carrier's share is the sum over its itineraries.

A QSI model is a logit model with $V_i = \ln q_i$ whose coefficients are calibrated by judgement or from past launches rather than estimated by maximum likelihood. It inherits the logit's strengths, transparency and additivity, and its weakness, the proportional substitution pattern known as independence of irrelevant alternatives (Chapter 5). QSI models were the traditional tool for forecasting itinerary shares; Coldren and coauthors describe logit itinerary share models, estimated on booking data at a major US airline, as the improvement on them (Coldren et al. 2003). The idea survives in modern planning systems, which chain an O&D demand model with a share model, then spill and recapture, then costs and revenues (see the box at the end of the chapter).

How much is a connection worth relative to a nonstop? Boeing's decision window methodology offers one set of answers, as coefficients that depend on a path quality index (1 for a nonstop, 2 for a one-stop on the same aircraft, 3 for an online connection, 4 for an interline connection) and on the length of the trip (Boeing Commercial Airplane Group 1993). Table 4.3 reproduces them. The striking

pattern is that path quality matters less as the trip gets longer: at 1,000 miles an online connection is worth less than 4 percent of a nonstop, at 6,000 miles more than three quarters. Two extra hours matter much less on a ten-hour journey.

Table 4.3 Airline service-based coefficients by path quality index (PQI) and trip length in statute miles (adapted from Boeing, 1993). PQI 1 is a nonstop, 2 a one-stop on the same aircraft, 3 an online connection, 4 an interline connection; higher values add stops, connections or excess time. Historical values; they show how the penalty for connecting shrinks with distance.

PQI	1,000 mi	2,000 mi	3,000 mi	4,000 mi	5,000 mi	6,000 mi
1	100.00	100.00	100.00	100.00	100.00	100.00
2	19.45	33.86	49.82	65.28	78.28	87.73
3	3.78	11.46	24.82	42.62	61.28	76.96
4	0.74	3.88	12.36	27.82	47.97	67.52
5	0.14	1.31	6.16	18.16	37.55	59.23

Market size with an outside option A share model divides a fixed market. To predict how much a new nonstop enlarges the market, the model needs an outside option, the choice not to travel, or, as the handbook points out, it can only redistribute a market of fixed size among the itineraries it contains (Network Planning Handbook 2026). The simplest consistent device is a nested structure with air travel in one nest and the outside option in the other (Chapter 6). When air travellers are a small fraction of the population, market demand is then proportional to A^{λ}, where $A = \sum_i q_i$ is the total score of all itineraries and $\lambda \in (0, 1]$ is the nest's log-sum parameter. That gives a compact rule for the effect of a new service.

RESULT 4.6 Stimulation by a new itinerary

Let market demand be $D_M = MA^{\lambda}$ with $A = \sum_i q_i$, and let new itineraries with total score q_{new} be added, winning a share $s = q_{\text{new}}/(A_0 + q_{\text{new}})$ of the enlarged market. Then

$$\frac{D_1}{D_0} = (1 - s)^{-\lambda}. \tag{4.6}$$

The new itineraries carry $N = sD_1$ passengers, of which $D_1 - D_0$ are **stimulated** (new to the market) and $D_0 - (1 - s)D_1$ are **diverted** from the existing itineraries, each existing itinerary losing in proportion to its previous share.

Proof. $D_1/D_0 = (A_1/A_0)^{\lambda}$ with $A_1 = A_0 + q_{\text{new}}$, and $A_1/A_0 = 1/(1 - s)$ because $1 - s = A_0/A_1$. Existing itinerary i had share q_i/A_0 and now has $q_i/A_1 = (1 - s)\, q_i/A_0$, so the existing itineraries together carry $(1 - s)D_1$, and their loss $D_0 - (1 - s)D_1$ is split in proportion to q_i/A_0. Finally $(D_1 - D_0) + [D_0 - (1 - s)D_1] = sD_1$. ∎

The result says that a nonstop stimulates a market in proportion to how much better it is than what already exists, measured by the share it wins. A nonstop that takes 10 percent of a market served by excellent connections barely changes it; one that takes 60 percent of a market served by poor connections can enlarge it by a third or more. The parameter λ governs how strongly a better offer turns nontravellers into travellers, and it is the hardest number in the analysis to know.

WORKED EXAMPLE 4.4 Adding a nonstop to Seville

Laurentian's YUL–SVQ mirrors a service publicly announced for launch on 8 September 2027, three times a week and four times a week from October (Network Planning Handbook 2026); every number below is invented. Before the nonstop, the Montreal–Seville market is served only by connections. Assume it carries 560 eastbound passengers a week (80 PDEW) on three daily itineraries: Carrier A over Madrid and Carrier B over Lisbon, both online connections, and Laurentian over Paris with a partner's flight to Seville, an interline connection.

Scores. YUL–SVQ is 3,453 statute miles. Interpolating Table 4.3 gives about 33 for an online connection and 19 for an inter-

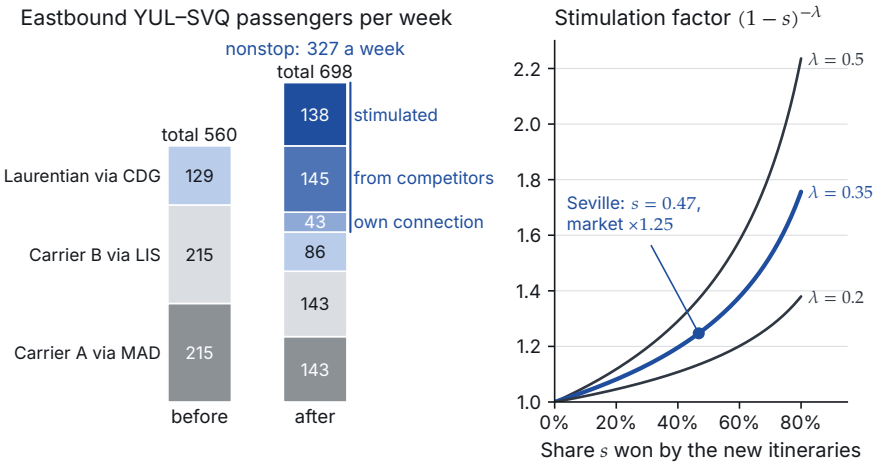


Figure 4.5 Left: the Montreal–Seville market before and after Laurentian's four-times-weekly nonstop, eastbound passengers per week. The nonstop's 327 weekly passengers come from three sources: travellers it creates, travellers it takes from competitors' connections, and travellers it takes from Laurentian's own connection over Paris. Right: the stimulation factor of (4.6) as a function of the share won by the new itineraries, for three values of λ.

line one at that distance; penalizing the long connections at Madrid, Lisbon and Paris, take 25, 25 and 15 points per departure, against 100 for the nonstop. Weekly scores are $7\times25 = 175$ for each competitor and $7 \times 15 = 105$ for Laurentian, so $A_0 = 455$ and the pre-launch shares are 38.5, 38.5 and 23.1 percent: 215, 215 and 129 passengers a week.

Share. The nonstop adds $4 \times 100 = 400$ points. Its share of the enlarged market is $s = 400/855 = 46.8$ percent.

Stimulation. With $\lambda = 0.35$, (4.6) gives a factor of $(1 - 0.468)^{-0.35} = 1.247$: the market grows to 698 passengers a week, or 99.8 PDEW. The nonstop carries $0.468 \times 698 = 327$ passengers a week, 82 per departure.

Sources. Of the nonstop's 327 passengers, 138 are stimulated (42 percent). The rest are diverted in proportion to previous shares: 145 from the two competitors and 43 from Laurentian's own connection over Paris. After the launch the competitors carry 143 each and Laurentian's connection 86. Laurentian's total traffic in the market rises from 129 to about 412 passengers a week, an increase of 283, not 327.

Sensitivity. The stimulation parameter is the weak link. With $\lambda = 0.2$ the market grows by 13.5 percent and the nonstop

carries 297 a week (74 per departure); with $\lambda = 0.5$ the market grows by 37.1 percent and the nonstop carries 359 (90 per departure). A rough cross-check with the time elasticity model: assigning trip times of 11 hours to the nonstop, 15 to the competitors' connections and 16 to Laurentian's, the share-weighted trip time falls from 15.2 to 13.3 hours, and with $b = -1.2$ that implies 18 percent more demand, between the low and central cases.

The planner learns four things. First, local demand of 82 passengers per departure is less than half of the 182 seats, so the route depends on connecting traffic over Montreal, which the running case's shoulder-season leg demand of 150 includes (Chapters 10 and 12). The 112 that Table 4.1 assigns to the YUL–SVQ market is larger because the four-leg bank shows only two feeders: it also stands for connections from spokes outside the model (Chapter 10). Second, the 43 passengers taken from Laurentian's own connection are not new revenue, but they free seats on the Paris flight, which Table 4.1 shows is already over-demanded: their value is the displacement cost of those seats, not zero (Chapter 12). Third, the proportional draw from each connection is an assumption of the logit form; in reality a nonstop may draw more heavily from the itinerary most similar to it (Chapter 6). Fourth, λ should come from comparable past launches rather than from a textbook, which is the subject of Chapter 8.

The handbook asks for the same discipline when a new route is presented: newly created demand, demand won from competitors and demand moved from the airline's own services belong on separate lines (Network Planning Handbook 2026). Figure 4.5 is that decomposition in picture form. The three components have very different values. Stimulated travellers are new revenue for the network, travellers won from competitors are new revenue that competitors will try to win back, and travellers moved from the airline's own itineraries are worth only the capacity they release elsewhere.

IN PRACTICE **How planners size markets and shares in 2026**

Market sizes now come from commercial datasets that combine several sources. OAG's Traffic Analyser estimates O&D passenger demand from global distribution system bookings, with hub-flow breakdowns into local, behind, beyond and bridge traffic, point of sale, average fares and market shares (OAG n.d.). Cirium's traffic and fares product estimates monthly O&D and segment passengers for more than 600 airlines from over 24 sources, including booking data, government statistics and the US ticket sample, fills gaps such as low-cost carriers' direct sales with regression and scaling, and is first released about 55 days after the end of a month and restated at 115 days (Cirium 2026a). The handbook advises treating such adjusted totals as estimates rather than a census and asking how direct bookings, groups and point-of-sale gaps are handled (Network Planning Handbook 2026). Coverage is a live issue: by May 2026 about one in five US travel agency transactions was settled through NDC channels (Travel Distribution News 2026), so booking data from traditional distribution capture a shrinking share of sales, one reason to ask any data vendor how it fills the gap. On the modelling side, Lufthansa Systems describes its NetLine/Plan system as evaluating O&D profitability with a chain of demand and market share, spill and recapture, and cost and revenue models (Lufthansa Systems 2026), the same sequence as this chapter and Chapter 9. Research on capacity allocation with endogenous demand stimulation was on the program of the AGIFORS scheduling and strategic planning meeting in 2026 (AGIFORS 2026), a sign that stimulation is still treated as an open modelling problem rather than a solved one.

4.8 Summary

- Demand is generated by O&D markets and counted in one-way trips (PDEW); capacity is supplied by legs, each serving many markets. Leg demand is the sum of the demands of the itineraries that use the leg, and a leg's load says little about any one market.
- Market demand depends on income and population, price, total trip time and the product offered. The multiplicative model $D = MP^aT^b$ has constant price and time elasticities and is estimated in logarithms, with care for the endogeneity of fares and frequencies.

- Revenue rises with a fare increase only where demand is inelastic. Linear and constant-elasticity demand agree locally and diverge for large changes. Business demand is less price elastic and more time elastic than leisure demand, and carrier-level elasticities exceed market elasticities.
- Total trip time includes schedule delay, K/f with $K = H/4$ for evenly spaced flights. The frequency elasticity of demand equals the time elasticity times the share of schedule delay in trip time, which is why frequency matters most where trips are short and frequency is low.
- Generalized cost combines fare and time: the price elasticity is the generalized-cost elasticity times the fare's share of generalized cost.
- Demand varies predictably by month, weekday, time of day and direction. Planners answer that variation with the schedule and never average it away.
- An airline's traffic is market size times market share. The S-curve $\mathrm{MS}_a = \mathrm{FS}_a^{\alpha} / \sum_b \mathrm{FS}_b^{\alpha}$ is a logit in log frequency; $\alpha > 1$ means that more flights make each flight more attractive, and frequency competition then dissipates a large share of market revenue.
- QSI models are logit share models with calibrated weights. With an outside option, a new itinerary that wins share s multiplies the market by $(1 - s)^{-\lambda}$; its traffic splits into stimulated travellers, travellers diverted from competitors, and travellers moved from the airline's own itineraries.

Table 4.4 Key formulas of the chapter.

Quantity	Formula
Leg demand	$D_l = \sum_j a_{lj} D_j$
Market demand	$D = M\,P^a\,T^b$
Price elasticity	$\eta = \partial \ln D / \partial \ln P$; linear: $-BP/(A - BP)$; constant: a
Revenue response	$\mathrm{d}\ln R / \mathrm{d}\ln P = 1 + \eta$
Total trip time	$T = t_{\text{fixed}} + t_{\text{flight}} + K/f$, $K = H/4$
Frequency elasticity	$\partial \ln D / \partial \ln f = \lvert b \rvert\ (K/f)/T$
Generalized cost	$\eta_P = \eta_G P/G$, $\eta_T = \eta_G v T/G$
S-curve	$\ln(\mathrm{MS}_a/\mathrm{MS}_b) = \alpha \ln(f_a/f_b)$
Frequency equilibrium	$f^* = \alpha R/(4c)$ per carrier
Stimulation	$D_1/D_0 = (1 - s)^{-\lambda}$

4.9 Further reading

Belobaba's airline economics review (Belobaba 2006d) is the source of the demand model, the trip time decomposition and the S-curve, and his PODS lecture (Belobaba 2006g) shows how the same ideas become a passenger choice simulator. Boeing's decision window methodology (Boeing Commercial Airplane Group 1993) is the clearest account of why frequency and timing win share. For a textbook treatment, read the chapter on airline economics and demand in The Global Airline Industry (Belobaba, Odoni, and Barnhart 2015, ch. 3), and Doganis (2019) for the commercial view. Douglas and Miller (1974) introduced schedule delay into the economics of air transport. The elasticity evidence is reviewed by IATA (Smyth and Pearce 2008; InterVISTAS Consulting 2007). Coldren and coauthors (2003) document the move from QSI to logit share models, and Garrow (2010) surveys discrete choice modelling of air travel demand.

CHAPTER

5

Choice Models I: Logit

IN THIS CHAPTER

- How a traveller's choice among itineraries becomes a probability, and which parts of utility the data can identify.
- The logit formula derived from extreme-value errors, and the S-shaped curve it traces.
- Independence of irrelevant alternatives: proportional substitution, and where it fails for airline itineraries.
- Own and cross elasticities, the scale of utility, and willingness to pay as a ratio of coefficients.
- Maximum likelihood estimation, worked through on simulated Laurentian YUL–CDG itinerary choices.
- The logsum: valuing a schedule change in dollars, and turning probabilities into market shares.

DRAWS ON Train, Discrete choice methods with simulation (2009), chapters 2 and 3; Belobaba, Introduction to PODS passenger choice model (MIT 16.75J, 2006); handbook

Almost every proposal a network planner makes changes the menu that travellers choose from. Moving Laurentian's evening departure to Paris from 18:30 to 19:30, adding a second daily frequency, shortening a connection at the Montreal hub, or watching a competitor launch a nonstop: each changes which itineraries exist and what they offer in fare, elapsed time, departure time and number of stops. The market-level tools of Chapter 4, market size, the frequency-share curve and the quality of service index, summarize the outcome. To predict how a particular change moves passengers between itineraries, between airlines and between flying and not flying, we need a model of the individual decision.

This chapter builds the workhorse model of that decision, multinomial logit. It is the language of itinerary choice in airline planning and research (Coldren et al. 2003; Garrow 2010), and, for a reader trained in industrial organization, it is also McFadden's conditional

logit, the foundation of the demand systems used in merger simulation (McFadden 1974). We derive it from first principles, study what it implies about substitution and price response, estimate it by maximum likelihood on simulated Laurentian data, and use it to put a dollar value on a schedule change.

Logit has one famous weakness, the independence of irrelevant alternatives, and it bites exactly where airline planners need answers: when a new itinerary is a close substitute for an existing one. Chapter 6 relaxes it with nested and mixed logit and turns to the airline-specific models (decision windows and the PODS passenger choice model) that sit inside planning and revenue management simulators. Throughout this chapter fares are treated as exogenous; Chapter 7 explains why they are not, and what to do about it.

5.1 Random utility

Consider a traveller, indexed n, who wants to fly from Montreal to Paris next Thursday evening. Her choice set C_n contains the itineraries she can buy, indexed $j = 1, \dots, J_n$, plus an **outside option** $j = 0$: not booking any of them, which covers travelling another day, flying another airline outside the set, or not travelling at all. She takes the alternative she likes best. The analyst sees some of what she weighs (the fare, the elapsed time, how far each departure is from the time she wanted, the number of connections, the carrier) and misses the rest: her loyalty status, the seat map, the meeting she cannot move.

DEFINITION 5.1 Random utility model

Traveller n chooses from C_n the alternative with the highest utility

$$U_{nj} = V_{nj} + \varepsilon_{nj}, \qquad V_{nj} = \beta' x_{nj},$$

where x_{nj} is a vector of observed attributes, V_{nj} is the **representative utility** and ε_{nj} collects everything the analyst does not observe, treated as random with joint density $f(\varepsilon_n)$. The **choice probability** of alternative i is

$$\begin{aligned} P_{ni} &= \Pr(U_{ni} > U_{nj} \text{ for all } j \neq i) \\ &= \int \mathbb{1}(\varepsilon_{nj} - \varepsilon_{ni} < V_{ni} - V_{nj} \;\; \forall j \neq i) f(\varepsilon_n)\, \mathrm{d}\varepsilon_n. \end{aligned}$$

The cleanest reading of P_{ni} is as a share. Among all travellers who face the same observed attributes as traveller n, the unobserved terms differ, and P_{ni} is the fraction of them who choose i (Train 2009, ch. 2). This is why a choice model fitted to individual bookings can be summed into itinerary shares, and why its errors are not noise in the engineering sense: they are the part of travellers' tastes that the model does not explain. The better the specification of V_{nj}, the less is left in ε_{nj}.

Two facts about utility decide what the data can and cannot tell us. Both are consequences of the fact that a choice reveals only which alternative ranks first.

Only differences in utility matter. Adding the same constant to every utility leaves every ranking, and therefore every probability, unchanged. Three practical consequences follow. First, with J alternatives at most $J - 1$ **alternative-specific constants** are identified; we normalize the utility of the outside option to $V_{n0} = 0$, so that a constant on the booking alternatives measures the attractiveness of flying relative to not flying. Second, a characteristic of the traveller, such as income or trip purpose, cannot enter utility on its own, because it would shift every alternative equally; it enters only through interactions with alternative attributes (a fare coefficient that differs for business travellers) or with alternative-specific coefficients. Third, only the $J - 1$ differences of the error terms affect the probabilities, so one dimension of the error distribution is not identified and must be fixed by assumption.

The scale of utility is arbitrary. Multiplying every utility by a positive number also preserves the ranking. The analyst fixes the scale by fixing the variance of the errors. If the true utility is $\tilde{\beta}' x_{nj} + \tilde{\varepsilon}_{nj}$ and

the errors have standard deviation proportional to σ, the coefficients that the data identify are $\beta = \tilde{\beta}/\sigma$: every coefficient measures the effect of an attribute relative to the importance of the unobserved factors. Two models estimated on different data, say business and leisure travellers, can therefore show different coefficient levels simply because one group's choices are noisier, and the levels cannot be compared directly. Ratios of coefficients can, because σ cancels; Section 5.5 builds on this.

A choice model identifies differences in utility, measured in units of unobserved noise. Constants are relative to a base alternative, coefficients are relative to the error scale, and only ratios of coefficients carry units such as dollars per hour.

5.2 From extreme-value errors to the logit formula

To turn the integral in Definition 5.1 into a number we must choose a distribution for the errors. The normal distribution is the obvious candidate, and it gives the probit model, but its choice probabilities have no closed form beyond two alternatives. The extreme-value distribution gives a closed form, and in data it is hard to distinguish from the normal (Train 2009, ch. 3). What matters is not its shape but the assumption that the errors are independent across alternatives: knowing that a traveller particularly likes the 18:30 tells us nothing about how she feels about the 20:45.

DEFINITION 5.2 Gumbel distribution

A random variable ε has the standard **Gumbel** (type I extreme value) distribution if

$$F(\varepsilon) = \exp(-e^{-\varepsilon}), \qquad f(\varepsilon) = e^{-\varepsilon}\exp(-e^{-\varepsilon}).$$

Its mode is 0, its mean is Euler's constant $\gamma \approx 0.5772$ and its variance is $\pi^2/6 \approx 1.645$. With location a and scale $\sigma > 0$ the distribution function is $\exp(-e^{-(\varepsilon-a)/\sigma})$.

Figure 5.1 shows the two densities that matter. A single Gumbel term is skewed to the right: the probability of exceeding 4 is 1.8 percent, against 0.4 percent for a normal with the same mean and vari-

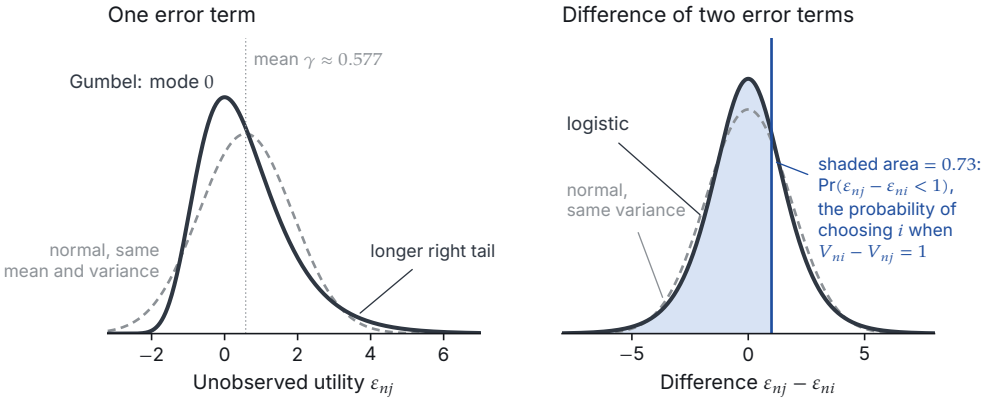


Figure 5.1 Left: the standard Gumbel density of one unobserved utility term, against a normal density with the same mean and variance. The Gumbel is skewed, with a longer right tail. Right: the difference of two independent Gumbel terms has the symmetric logistic density, close to a normal with the same variance but with slightly heavier tails. The shaded area is the probability that the difference falls below 1, which is the logit probability of choosing alternative i when its representative utility exceeds that of j by one unit.

ance. Choice probabilities, however, depend only on differences of errors, and the difference of two independent Gumbel terms is symmetric and nearly normal. Two properties of the Gumbel family do all the work in this chapter.

RESULT 5.1 Two properties of Gumbel errors

Let $\varepsilon_1, \dots, \varepsilon_J$ be independent standard Gumbel variables and $V_1, \dots, V_J$ constants.

1. The difference $\varepsilon_2 - \varepsilon_1$ has the **logistic** distribution: $\Pr(\varepsilon_2 - \varepsilon_1 < x) = 1/(1 + e^{-x})$.
2. The maximum $\max_j (V_j + \varepsilon_j)$ is Gumbel with unit scale and location $\ln \sum_j e^{V_j}$.

Proof. For the first property, condition on $\varepsilon_1 = s$. Then $\Pr(\varepsilon_2 < s + x) = \exp(-e^{-s}e^{-x})$, and integrating over the density of s,

$$\Pr(\varepsilon_2 - \varepsilon_1 < x) = \int_{-\infty}^{\infty} \exp(-e^{-s}(1 + e^{-x}))\, e^{-s}\, \mathrm{d}s.$$

With $t = e^{-s}$, so that $\mathrm{d}t = -e^{-s}\,\mathrm{d}s$ and t runs from ∞ to 0, the integral becomes $\int_0^{\infty} \exp(-t(1 + e^{-x}))\,\mathrm{d}t = 1/(1 + e^{-x})$.

For the second, the maximum is below u exactly when every term is, so by independence

$$\Pr\Big(\max_j (V_j + \varepsilon_j) \le u\Big) = \prod_j \exp(-e^{-(u - V_j)}) = \exp\Big(-e^{-u} \sum_j e^{V_j}\Big),$$

which is $\exp(-e^{-(u-I)})$ with $I = \ln \sum_j e^{V_j}$: a Gumbel distribution with location I. ∎

The second property is called **max-stability**: the best of several Gumbel alternatives is again Gumbel, with a location that is a smooth version of the largest V_j. It is the source of the logsum of Section 5.7. We can now derive the model.

RESULT 5.2 Logit choice probabilities

If the errors ε_{nj} are independent standard Gumbel variables, the probability that traveller n chooses alternative i is

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_{j \in C_n} e^{V_{nj}}}. \tag{5.1}$$

Proof. Drop the index n. Fix the error of the alternative of interest at $\varepsilon_i = s$. Alternative i then beats alternative j exactly when $\varepsilon_j < s + V_i - V_j$, an event of probability $F(s + V_i - V_j)$. The errors are independent, so the conditional probability that i beats every rival

is the product of these terms, and the choice probability averages it over the density of s:

$$P_i = \int_{-\infty}^{\infty} \prod_{j \neq i} \exp(-e^{-(s+V_i-V_j)})\, e^{-s} \exp(-e^{-s})\, \mathrm{d}s. \tag{5.2}$$

The last factor, $\exp(-e^{-s})$, is exactly the term the product would contain for $j = i$, because $V_i - V_i = 0$. Absorbing it, the integrand becomes $e^{-s}\exp(-e^{-s}S_i)$ with

$$S_i = \sum_{j \in C} e^{-(V_i - V_j)} = e^{-V_i} \sum_{j \in C} e^{V_j}.$$

Substitute $t = e^{-s}$ as in the previous proof:

$$P_i = \int_0^{\infty} e^{-tS_i}\, \mathrm{d}t = \frac{1}{S_i} = \frac{e^{V_i}}{\sum_j e^{V_j}}.$$

■

There is a second route to the same formula that is worth knowing, because it explains why logit behaves as it does. By max-stability, the best rival of alternative i has utility $\max_{j \neq i} U_j = I_{-i} + \tilde{\varepsilon}$, where $I_{-i} = \ln \sum_{j \neq i} e^{V_j}$ and $\tilde{\varepsilon}$ is a standard Gumbel term independent of ε_i. Choosing i means beating that composite rival, and by the logistic property

$$P_i = \Pr(\tilde{\varepsilon} - \varepsilon_i < V_i - I_{-i}) = \frac{1}{1 + e^{I_{-i} - V_i}} = \frac{e^{V_i}}{e^{V_i} + e^{I_{-i}}},$$

which is (5.1). Every multinomial logit is a binary logit between one alternative and a composite of all the others whose utility is the log of the sum of their exponentiated utilities.

The formula has the properties a share model needs: each probability lies strictly between 0 and 1 and they sum to 1 over the choice set. No alternative ever has probability exactly zero, so alternatives that no traveller would consider (a two-stop itinerary taking 30 hours) should be removed from the choice set rather than left to absorb a small share.

With two alternatives, (5.1) becomes $P_1 = 1/(1 + e^{-(V_1 - V_2)})$, the logistic function of the utility difference. Figure 5.2 draws it for a traveller choosing between Laurentian's nonstop and a competitor's nonstop at the same time of day, with $V_L - V_C = \alpha_L + \beta_f (f_L - f_C)$, a Laurentian preference of $\alpha_L = 0.25$ and a fare coefficient of $\beta_f = -0.004$ per dollar. At equal fares Laurentian wins 56.2 percent of such travellers. The traveller is indifferent when Laurentian charges

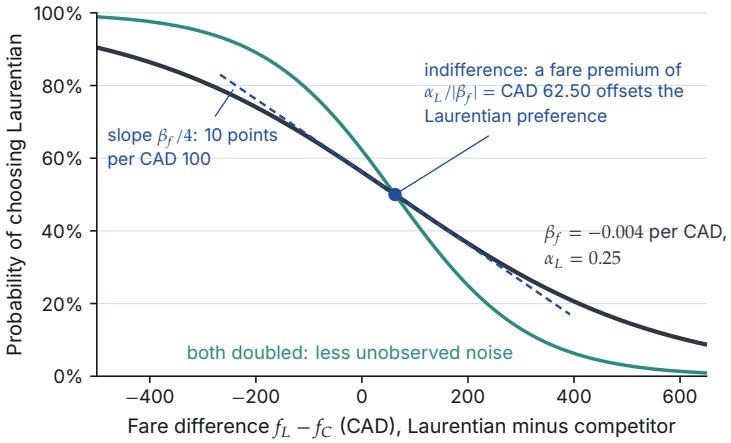


Figure 5.2 The binary logit curve: probability that a traveller chooses Laurentian's nonstop over a competitor's nonstop at the same time of day, as a function of the fare difference. The dashed tangent has slope $\beta_f/4$. The teal curve doubles both coefficients, which is what a smaller variance of unobserved utility does; the indifference point does not move, because it depends only on the ratio of the coefficients.

CAD 62.50 more, the premium $\alpha_L/|\beta_f|$ that exactly offsets the preference. The slope of the curve is $\mathrm{d}P/\,\mathrm{d}V = P(1-P)$, largest at $P = 1/2$, where it equals $1/4$: near the indifference point each CAD 100 of fare difference moves 10 points of share. Far from it the curve is flat. When one itinerary dominates, small changes in its fare barely matter, an observation we will formalize as elasticities.

The scale parameter. Suppose the errors are Gumbel with scale σ, so that their variance is $\sigma^2\pi^2/6$. Dividing all utilities by σ brings the errors back to the standard Gumbel distribution, and the probabilities become

$$P_{ni} = \frac{e^{V_{ni}/\sigma}}{\sum_j e^{V_{nj}/\sigma}}.$$

As $\sigma \to 0$ the probabilities approach 1 for the alternative with the largest V_{nj} and 0 for the others: choice becomes deterministic. As $\sigma \to \infty$ every alternative gets $1/J$. The steeper teal curve in Figure 5.2 is the same traveller with half the unobserved noise. Because only $\tilde{\beta}/\sigma$ is identified, a segment whose choices are more predictable shows larger coefficients across the board. When segments with different noise are pooled, the model needs a relative scale parameter for one of them, estimated jointly with the coefficients (Train 2009, ch. 3).

5.3 Substitution and the independence of irrelevant alternatives

The denominator of (5.1) is shared by every alternative. That single fact fixes the pattern of substitution that logit can represent.

RESULT 5.3 Independence of irrelevant alternatives

In the logit model, for any two alternatives i and k in the choice set,

$$\frac{P_{ni}}{P_{nk}} = e^{V_{ni} - V_{nk}},$$

whatever other alternatives are available and whatever their attributes. Equivalently, if an alternative m is added to the choice set, every existing probability is multiplied by the same factor, and the new alternative takes from each existing alternative j an amount proportional to its share:

$$P^0_{nj} - P^1_{nj} = P^0_{nj} P^1_{nm}.$$

Proof. In the ratio the denominators cancel. When m is added, the denominator grows from $S = \sum_j e^{V_{nj}}$ to $S + e^{V_{nm}}$, so $P^1_{nj} = P^0_{nj}\, S/(S + e^{V_{nm}})$, and $1 - S/(S + e^{V_{nm}}) = P^1_{nm}$. ∎

The property is called the **independence of irrelevant alternatives** (IIA). Its practical content is in the second statement: under logit, the share of a new itinerary's passengers drawn from each existing alternative equals that alternative's share before the change. That is reasonable when all alternatives are equally different from one another. It is wrong when some alternatives are close substitutes, and airline choice sets are full of close substitutes.

The starkest airline case is the **codeshare duplicate**. A flight operated by Laurentian and marketed under a partner's code appears twice in a booking display. If both listings enter a logit model as separate alternatives, the physical flight receives two shares of the denominator and its predicted share rises, although nothing about the product changed. This is the airline version of the classic red-bus, blue-bus paradox. Itinerary choice models must be run on operated alternatives, or on structures that recognize duplicates as the same product.

The economically important case is subtler, and it is the one planners face every season.

WORKED EXAMPLE 5.1 A second daily frequency to Paris

Take a representative traveller on YUL–CDG whose representative utilities are 0 for not booking, 0.40 for Laurentian's 18:30 nonstop, 0.25 for competitor B's 20:45 nonstop and −0.60 for carrier C's 17:15 one-stop itinerary via London Heathrow. The exponentials are 1, 1.492, 1.284 and 0.549, which sum to 4.325, so the shares are 23.1 percent (no booking), 34.5 percent (Laurentian), 29.7 percent (B) and 12.7 percent (C).

Laurentian adds a 21:15 nonstop with the same fare, aircraft and, for this traveller, the same convenience as the 18:30, so its representative utility is also 0.40. The denominator rises to 5.816 and every existing share is multiplied by $4.325/5.816 = 0.744$, with the results in Table 5.1.

Laurentian's total share rises from 34.5 to 51.3 percent, a gain of 16.8 points, and the last column reproduces the first exactly, as Result 5.3 says it must.

Now look at what the model has assumed. The new 21:15 departure leaves half an hour after the competitor's 20:45 and two and three-quarter hours after Laurentian's own flight. It is a close substitute for both nonstops and a poor substitute for the decision not to travel. Logit nevertheless says that almost a quarter of its passengers are new to the market, and that it takes from the connection, proportionally, as much as from the competitor's nonstop. Whether a second frequency mostly cannibalizes Laurentian's own flight or mostly takes passengers from the competitor is the question the planner is being paid to answer, and logit answers it by assumption. Example 6.1 reworks this example with nested logit, where the answer ranges from a gain of 8.2 points to 21.1 points depending on which itineraries are allowed to compete most closely.

Table 5.1 Multinomial logit shares on YUL–CDG before and after Laurentian adds a 21:15 nonstop, for the representative traveller of Example 5.1. The last column shows where the new flight's passengers come from; it equals the first column.

Alternative	Before	After	Change (points)	Share of new flight's traffic
No booking	23.1%	17.2%	−5.9	23.1%
Laurentian 18:30	34.5%	25.6%	−8.9	34.5%
Competitor B 20:45	29.7%	22.1%	−7.6	29.7%
Carrier C 17:15 via LHR	12.7%	9.4%	−3.3	12.7%
Laurentian 21:15 (new)		25.6%	+25.6	

PITFALL

Counting near-duplicates as distinct alternatives. Codeshare listings of one flight, the same itinerary sold through two channels, or connections that differ only by a feeder flight ten minutes apart all inflate the predicted share of whatever they duplicate. Collapse them to operated alternatives before estimating or forecasting, or use a nested structure that treats them as close substitutes.

IIA is not only a weakness. Because the odds between two alternatives do not depend on the others, a logit model can be estimated consistently on a random subset of each traveller's alternatives, which matters in markets with hundreds of connecting itineraries (Train 2009, ch. 3). The same logic yields a test: if IIA holds, coefficients estimated after dropping an alternative from the choice set should not differ systematically from those estimated on the full set (J. Hausman and McFadden 1984). Chapter 6 gives a more constructive test, which nests logit inside a model that does not impose IIA.

5.4 Elasticities

Planners and revenue managers talk about price response in elasticities, because elasticities are free of units. For logit they have simple closed forms, and the forms restate IIA in a second language.

RESULT 5.4 Derivatives and elasticities of logit probabilities

Let x_{njk} be attribute k of alternative j, entering utility linearly with coefficient β_k. Then

$$\frac{\partial P_{ni}}{\partial x_{nik}} = \beta_k P_{ni}(1 - P_{ni}), \qquad \frac{\partial P_{ni}}{\partial x_{njk}} = -\beta_k P_{ni} P_{nj} \quad (j \neq i),$$

and the own and cross elasticities are

$$\eta_{ii} = \beta_k \, x_{nik} \, (1 - P_{ni}), \qquad \eta_{ij} = -\beta_k \, x_{njk} \, P_{nj} \quad (j \neq i). \tag{5.3}$$

Proof. Write $P_{ni} = e^{V_{ni}}/S_n$ with $S_n = \sum_j e^{V_{nj}}$, and note that $\partial S_n / \partial x_{njk} = \beta_k e^{V_{nj}}$. For the own attribute, the quotient rule gives $\beta_k e^{V_{ni}}/S_n - e^{V_{ni}} \beta_k e^{V_{ni}}/S_n^2 = \beta_k P_{ni}(1 - P_{ni})$. For an attribute of $j \neq i$, only the denominator moves: $-e^{V_{ni}} \beta_k e^{V_{nj}}/S_n^2 = -\beta_k P_{ni} P_{nj}$. Multiplying each derivative by x/P_{ni} gives the elasticities. ∎

Three readings of (5.3) are useful. First, the own elasticity is proportional to $1 - P_{ni}$: an itinerary that already wins most travellers has few left to win, so it is less elastic. With $\beta_f = -0.004$ per dollar, a fare of CAD 760 and a share of 34.5 percent, Laurentian's own-fare elasticity is $-0.004 \times 760 \times 0.655 = -1.99$. Second, the cross elasticity with respect to an attribute of j does not depend on i: a 1 percent cut in Laurentian's fare reduces the probability of every other alternative by the same $0.004 \times 760 \times 0.345 = 1.05$ percent, whether it is the competitor's nonstop at a similar time or the decision not to travel. This is proportional substitution again. Third, the derivatives with respect to any one attribute sum to zero over alternatives, because the probabilities always sum to one: a share gained by one itinerary is taken from the others.

Itinerary elasticity and market elasticity. The outside option controls how much of a price change is stimulation rather than diversion. Raise every air fare by the same small proportion δ. Since $\partial P_{n0}/\partial f_{nj} = -\beta_f P_{n0} P_{nj}$, the probability of booking any itinerary, $1 - P_{n0}$, changes by $\beta_f P_{n0} \sum_{j \geq 1} f_{nj} P_{nj}$ per unit of δ, and the **market elasticity** is

$$\eta_{\text{air}} = \beta_f \bar{f}_n \, P_{n0}, \qquad \bar{f}_n = \frac{\sum_{j \geq 1} f_{nj} P_{nj}}{1 - P_{n0}}, \tag{5.4}$$

where $\bar{f}_n$ is the expected fare paid by those who book. For the representative traveller of Example 5.1, with fares of CAD 760, 720 and 540, the expected fare is CAD 708 and $\eta_{\text{air}} = -0.004 \times 708 \times 0.231 \approx -0.65$. The same traveller's itinerary-level elasticities are three times larger, because most of an itinerary's response to its own fare is diversion to other itineraries, which cancels when all fares move together. Equation (5.4) also shows that a model without an outside option ($P_{n0} = 0$) implies a market that never grows or shrinks: it only reallocates a fixed number of travellers, as the handbook warns (Network Planning Handbook 2026). The outside share is the least well measured number in airline data, since booking records contain no one who did not book; Chapter 4 discusses market size and Chapter 7 returns to the assumption in demand estimation.

5.5 Willingness to pay and the value of time

Coefficients are measured in units of unobserved noise, which no one can interpret. Their ratios are measured in dollars, which everyone can.

DEFINITION 5.3 Marginal utility of money and willingness to pay

When the fare enters utility linearly with coefficient $\beta_f < 0$, $\alpha = -\beta_f$ is the **marginal utility of money**. The money value of attribute k is

$$v_k = \frac{\beta_k}{\alpha} = -\frac{\beta_k}{\beta_f},$$

the fare change that leaves utility unchanged when x_k rises by one unit: from $\mathrm{d}U = \beta_f \,\mathrm{d}f + \beta_k \,\mathrm{d}x_k = 0$, $\mathrm{d}f/\mathrm{d}x_k = -\beta_k/\beta_f$. The **value of time** is $\text{VOT} = -v_T = \beta_T/\beta_f$, in dollars per hour, where $\beta_T < 0$ is the coefficient of elapsed time.

Dividing representative utility by α turns it into money. For an itinerary with fare f_{nj}, elapsed time T_{nj}, departure displacement d_{nj} (the gap between its departure and the time the traveller wanted) and c_j connections,

$$\frac{V_{nj}}{\alpha} = \text{constants} - (f_{nj} + \text{VOT}\, T_{nj} + v_d\, d_{nj} + v_c\, c_j),$$

where v_d and v_c, defined like VOT as the ratio of the attribute's coefficient to the fare coefficient, are the dollar costs of an hour of displacement and of a connection. The bracket is the **generalized cost**

of the itinerary, the quantity the handbook asks planners to think in (Network Planning Handbook 2026), and logit is a model in which travellers choose the itinerary with the lowest generalized cost plus a random term. Because the scale σ cancels in every ratio, values of time and willingness to pay can be compared across segments and data sets even when coefficient levels cannot. Departure displacement is the individual-level form of the schedule delay of Chapter 4; Koppelman, Coldren, and Parker (2008) estimate its effect on airline itinerary demand.

A value of time is a ratio of two estimates, so its uncertainty combines both. By the delta method, with $r = \beta_T / \beta_f$,

$$\operatorname{Var}(\hat{r}) \approx \frac{1}{\beta_f^2}\Big[\operatorname{Var}(\hat{\beta}_T) - 2r\operatorname{Cov}(\hat{\beta}_T, \hat{\beta}_f) + r^2 \operatorname{Var}(\hat{\beta}_f)\Big]. \tag{5.5}$$

The approximation is good when the fare coefficient is estimated precisely. When it is not, the ratio's sampling distribution is skewed and can even put weight on negative values, as the Monte Carlo experiment of the next section shows.

PITFALL

Taking a value of time from a model with an endogenous fare. Airlines charge more when demand is high and on flights travellers like for reasons the analyst does not see, so observed fares are positively correlated with the unobserved part of utility. A naive fare coefficient is then biased toward zero, and every value derived from it, values of time, of nonstop service and of brand, is inflated. Lurkin et al. (2017) find this pattern for fare coefficients and values of time in airline itinerary choice models estimated on US markets. Chapter 7 shows how instruments and control functions correct it.

5.6 Estimation by maximum likelihood

Observe N independent choices and let $y_{ni} = 1$ if traveller n chose alternative i and 0 otherwise. The probability of the observed sample is $\prod_n \prod_i P_{ni}^{y_{ni}}$, and the **log-likelihood** is

$$\begin{aligned} \mathrm{LL}(\beta) &= \sum_{n=1}^{N} \sum_{i \in C_n} y_{ni} \ln P_{ni}(\beta), \\ \ln P_{ni} &= \beta' x_{ni} - \ln \sum_{j \in C_n} e^{\beta' x_{nj}}. \end{aligned} \tag{5.6}$$

The maximum likelihood estimator $\hat{\beta}$ maximizes it. For logit the maximization is unusually well behaved.

RESULT 5.5 Score, Hessian and concavity of the logit log-likelihood

Let $\bar{x}_n(\beta) = \sum_j P_{nj} x_{nj}$ be the probability-weighted mean of traveller n's attribute vectors. The gradient and Hessian of (5.6) are

$$\begin{aligned} g(\beta) &= \sum_n \sum_i (y_{ni} - P_{ni})\, x_{ni}, \\ H(\beta) &= -\sum_n \sum_j P_{nj} (x_{nj} - \bar{x}_n)(x_{nj} - \bar{x}_n)'. \end{aligned}$$

$H(\beta)$ is negative semidefinite for every β, so the log-likelihood is globally concave. If no linear combination of attributes is constant across every traveller's alternatives, H is negative definite and the maximum, when it exists, is unique.

Proof. Differentiating the log of the sum, $\nabla \ln P_{ni} = x_{ni} - \bar{x}_n$. Weighting by y_{ni} and using $\sum_i y_{ni} = 1 = \sum_i P_{ni}$ gives $g = \sum_n (\sum_i y_{ni} x_{ni} - \bar{x}_n) = \sum_n \sum_i (y_{ni} - P_{ni}) x_{ni}$. For the Hessian, differentiate $\bar{x}_n$: since $\nabla P_{nj} = P_{nj}(x_{nj} - \bar{x}_n)$,

$$\frac{\partial \bar{x}_n}{\partial \beta'} = \sum_j P_{nj}\, x_{nj}(x_{nj} - \bar{x}_n)' = \sum_j P_{nj}(x_{nj} - \bar{x}_n)(x_{nj} - \bar{x}_n)',$$

where the second equality subtracts $\bar{x}_n \sum_j P_{nj}(x_{nj} - \bar{x}_n)' = 0$. Hence $H = -\sum_n \partial \bar{x}_n / \partial \beta'$, a negative sum of probability-weighted covariance matrices. For any vector c, $c'Hc = -\sum_n \sum_j P_{nj} (c'(x_{nj} - \bar{x}_n))^2 \le 0$, with equality only if $c'x_{nj}$ is the same for all alternatives of every traveller. ∎

The first-order condition $g(\hat{\beta}) = 0$ has a useful reading: the residuals $y_{ni} - P_{ni}$ are uncorrelated with the attributes, the same moment condition that defines least squares. Applied to an alternative-specific constant, whose attribute is a dummy for that alternative, it says that the predicted share of the alternative equals its observed share in the sample. A logit model with a full set of constants therefore always reproduces the sample shares; that is a property of the estimator, not evidence that the model is right.

Under correct specification and exogenous attributes, $\hat{\beta}$ is consistent, asymptotically normal and efficient, with asymptotic covariance matrix $(-H(\hat{\beta}))^{-1}$, whose diagonal gives the standard errors. Nested hypotheses are tested with the **likelihood ratio** statistic $2(\mathrm{LL}_u - \mathrm{LL}_r)$, asymptotically χ^2 with as many degrees of freedom as restrictions. The usual summary of fit is McFadden's $\rho^2 = 1 - \mathrm{LL}(\hat{\beta})/\mathrm{LL}(0)$, which is not an R^2: its values are much lower, and it is useful only for comparing models on the same data.

Concavity makes Newton's method reliable: iterate $\beta \leftarrow \beta - H(\beta)^{-1} g(\beta)$. Started from $\beta = 0$, as in the code below, it converges in a few iterations; from a poor start a full step can overshoot, and halving the step until the log-likelihood rises fixes that. The whole estimator fits in a dozen lines of Python.

```
import numpy as np

def fit_logit(X, y, iters=25):
    """Conditional logit, Newton-Raphson.
    X: (N, J, K) attributes; the outside
       option is a row of zeros.
    y: (N,) chosen alternative."""
    N, J, K = X.shape
    b = np.zeros(K)
    for _ in range(iters):
```

```
        V = X @ b
        V = V - V.max(axis=1, keepdims=True)
        P = np.exp(V)
        P /= P.sum(axis=1, keepdims=True)
        xbar = np.einsum("nj,njk->nk", P, X)
        xy = X[np.arange(N), y]
        g = (xy - xbar).sum(axis=0)
        D = X - xbar[:, None, :]
        H = -np.einsum("nj,njk,njl->kl",
                       P, D, D)
        b = b - np.linalg.solve(H, g)
    se = np.sqrt(np.diag(np.linalg.inv(-H)))
    return b, se
```

WORKED EXAMPLE 5.2 Estimating itinerary choice on YUL–CDG

We simulate 4,000 shopping sessions for one-way YUL–CDG travel on summer evenings. The data are illustrative, generated from a known model so that the estimates can be checked against the truth. Each session offers four alternatives: Laurentian's 18:30 nonstop, competitor B's 20:45 nonstop, carrier C's 17:15 one-stop via LHR, and no booking. Fares vary across sessions with the booking date and the inventory state, averaging CAD 782, 740 and 556. Elapsed times vary by day of operation: 6.5 to 7.3 hours for Laurentian's nonstop, 6.7 to 7.5 hours for B's, 8.8 to 13.8 hours for the connection. Each traveller has a preferred departure time (average 19:30, standard deviation two hours), which we assume the booking engine records from the search request; the departure displacement d_{nj} is the absolute gap between it and each departure. Representative utility is

$$V_{nj} = \alpha_{\text{air}} + \alpha_L \mathbb{1}_L(j) + \beta_c c_j + \beta_f f_{nj}/100 + \beta_T T_{nj} + \beta_d d_{nj},$$

with $V_{n0} = 0$, where $\mathbb{1}_L(j)$ marks Laurentian and c_j marks the connection. Because the connection is the only itinerary of carrier C, its coefficient also absorbs anything else that distinguishes C, an identification limit worth remembering whenever a variable is perfectly aligned with one alternative. Choices are simulated by adding independent Gumbel draws to the true utilities and picking the largest. In the sample, 25.5 percent of sessions end without a booking, 35.2 percent with Laurentian, 30.1 percent with B and 9.2 percent with the connection.

Newton's method from $\beta = 0$ raises the log-likelihood from $\mathrm{LL}(0) = 4{,}000 \ln(1/4) = -5{,}545.2$ to $-4{,}860.5$ after one step, $-4{,}828.1$ after two and $-4{,}827.3$ after three, where it stays; $\rho^2 = 0.129$. Table 5.2 reports the estimates.

Every estimate lies within one standard error of its true value. Dropping the Laurentian constant lowers the log-likelihood to $-4{,}848.0$; the likelihood ratio statistic $2 \times 20.74 = 41.5$ far exceeds the 5 percent critical value of 3.84, so the data clearly detect a preference for Laurentian.

Money values. The estimated marginal utility of money is 0.0040 per dollar, the fare coefficient divided by 100. The value of time is $0.228/0.400 \times 100 =$ CAD 57.0 per hour, with a delta-method standard error of 9.6 and a 95 percent confidence interval from CAD 38 to CAD 76 (the true value is CAD 62.50). An hour of departure displacement costs CAD 50.3 (standard error 4.7), the connection penalty is CAD 226 (standard error 38) on top of the connection's longer elapsed time, and the Laurentian preference is worth CAD 67.5 (standard error 10.4). A typical connection is 4.4 hours slower than a nonstop; leaving at the same time and without a brand difference, it must therefore be about $4.4 \times 57 + 226 \approx$ CAD 476 cheaper to be equally attractive, and CAD 67.5 more than that to match Laurentian's own nonstop.

Elasticities. Table 5.3 reports aggregate fare elasticities computed by sample enumeration: for each session the individual elasticity of (5.3), averaged with weights equal to the session's probability of choosing the alternative whose bookings are being measured.

Own-fare elasticities are close to -1.8, and in each column the cross elasticities are nearly, but not exactly, equal. IIA holds for each traveller, but the aggregate mixes travellers with different preferred departure times, for whom the alternatives have different probabilities; heterogeneity in observed attributes already bends the aggregate substitution pattern slightly. Evaluating the formula once at the average traveller instead of enumerating gives a Laurentian own-fare elasticity of -2.01 rather than -1.81, a first sign of the aggregation error discussed in Section 5.8.

For the planner, the table translates directly into revenue arithmetic. A 10 percent increase in Laurentian's fare, about CAD

78, loses roughly 18 percent of its bookings in this market. The lost travellers spread over the competitor, the connection and no booking roughly in proportion to their shares, which is exactly the pattern the next chapter will question.

Table 5.2 Maximum likelihood estimates of the YUL–CDG itinerary choice model on 4,000 simulated shopping sessions (illustrative data). Log-likelihood $-4{,}827.3$.

Parameter	True value	Estimate	Standard error	t
Air constant	5.30	5.129	0.302	17.0
Laurentian constant	0.25	0.270	0.042	6.4
Connection	-0.90	-0.904	0.157	-5.8
Fare (per CAD 100)	-0.40	-0.400	0.018	-21.9
Elapsed time (per hour)	-0.25	-0.228	0.037	-6.2
Departure displacement (per hour)	-0.20	-0.201	0.017	-12.0

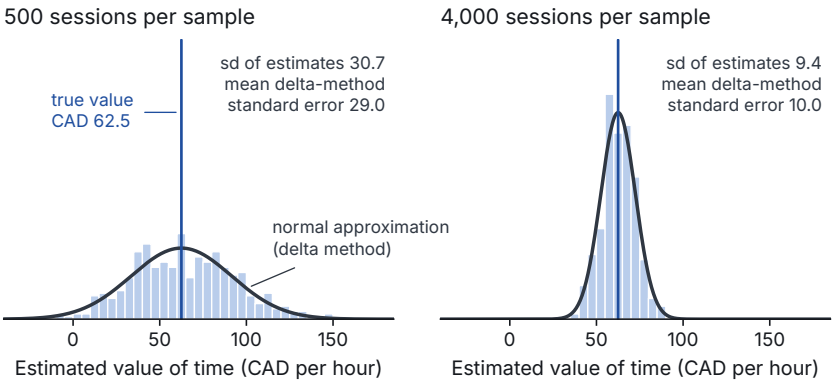


Figure 5.3 Sampling distribution of the estimated value of time across 400 simulated YUL–CDG samples of 500 and of 4,000 shopping sessions. The curves are the normal approximations implied by the delta-method standard errors; the vertical line is the true value, CAD 62.50 per hour.

Table 5.3 Aggregate fare elasticities implied by the estimated model. Each column gives the response to a 1 percent change in one carrier's fare. The market elasticity of total bookings to a uniform change in all fares is -0.70.

Bookings of	Laurentian fare	Competitor fare	Connection fare
No booking	1.00	0.84	0.21
Laurentian	−1.81	0.77	0.19
Competitor B	0.95	−1.84	0.18
Connection C	1.03	0.77	−1.87

How precise are such estimates in samples of realistic size? Figure 5.3 repeats the whole exercise 400 times for samples of 500 and 4,000 sessions. With 4,000 sessions the estimated values of time centre on the truth (mean CAD 62.6), their standard deviation across samples (9.4) matches the average delta-method standard error (10.0), and nominal 95 percent intervals cover the truth in 95.8 percent of samples. With 500 sessions the spread triples, to a standard deviation of 30.7, the distribution is visibly skewed to the right, and 5 of the 400 samples even produce a negative value of time. Values of time are among the least precisely estimated quantities in any itinerary choice model; report them with intervals.

5.7 The logsum and consumer surplus

A better departure time benefits every traveller who might use it, including those who keep choosing the same itinerary. To value a schedule change from the traveller's side we need the expected util-

ity of the best available alternative, expressed in money. Logit delivers it in closed form.

RESULT 5.6 The logsum: expected maximum utility and consumer surplus

Under independent standard Gumbel errors,

$$\mathbb{E}\Big[\max_{j\in C_n} U_{nj}\Big] = \ln \sum_{j\in C_n} e^{V_{nj}} + \gamma, \qquad \frac{\partial}{\partial V_{ni}} \ln \sum_{j} e^{V_{nj}} = P_{ni}. \tag{5.7}$$

If utility is linear in money with constant marginal utility $\alpha = -\beta_f$, the change in traveller n's expected consumer surplus when the choice set or its attributes change from situation 0 to situation 1 is

$$\Delta\mathbb{E}[\mathrm{CS}_n] = \frac{1}{\alpha}\Big[\ln \sum_{j\in C_n^1} e^{V_{nj}^1} - \ln \sum_{j\in C_n^0} e^{V_{nj}^0}\Big]. \tag{5.8}$$

Proof. By the max-stability property of Result 5.1, $\max_j U_{nj}$ is Gumbel with unit scale and location $I_n = \ln \sum_j e^{V_{nj}}$, so its mean is $I_n + \gamma$. Consumer surplus is the money equivalent of the utility obtained, $\mathrm{CS}_n = \max_j U_{nj}/\alpha$, plus a constant that reflects the arbitrary level of utility; the constant and γ cancel in the difference between two situations. The derivative of I_n with respect to V_{ni} is $e^{V_{ni}} / \sum_j e^{V_{nj}} = P_{ni}$. ∎

The quantity $I_n = \ln \sum_j e^{V_{nj}}$ is the **logsum**, or inclusive value, of the choice set. Three of its properties guide its use. It is a smooth maximum: it always exceeds $\max_j V_{nj}$, because the traveller also benefits from the chance that another alternative turns out better for her on the day, and it approaches the maximum when one alternative dominates. It rises with every alternative added, with diminishing returns. And its gradient is the vector of choice probabilities, the discrete-choice counterpart of Roy's identity. A small fare cut Δf on alternative i therefore raises expected surplus by $P_{ni}\Delta f$: the probability of buying times the saving, as in any textbook welfare calculation. For larger changes, averaging the probabilities before and after gives the rule of a half, $\Delta\mathbb{E}[\mathrm{CS}_n] \approx \frac{1}{2}\sum_j (P_{nj}^0 + P_{nj}^1)\,\Delta V_{nj}/\alpha$. The welfare interpretation of the logsum is due to Small and Rosen (1981), and it requires the marginal utility of money not to vary with the change

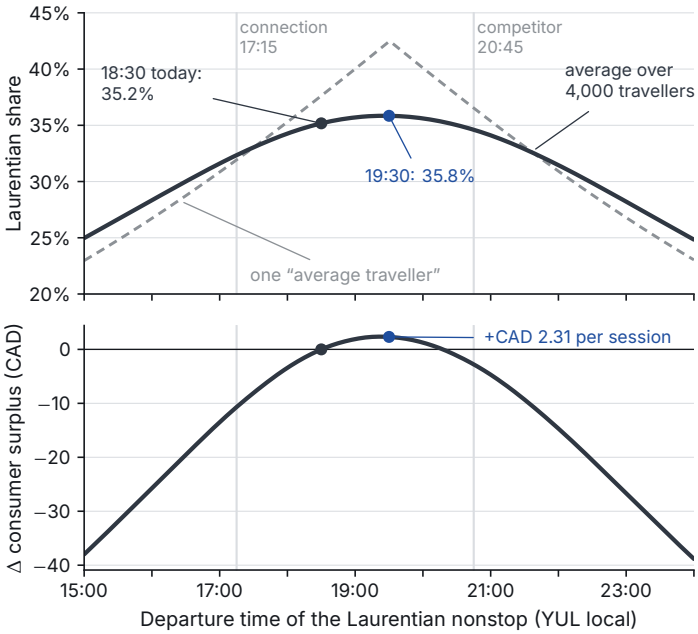


Figure 5.4 Retiming Laurentian's YUL–CDG nonstop, with the competitor at 20:45 and the connection at 17:15 unchanged. Top: Laurentian's share computed for the 4,000 sessions of Example 5.2 and averaged (dark), and computed for a single traveller with average preferences and attributes (grey). Bottom: change in expected consumer surplus per session relative to the current 18:30 departure, from (5.8).

considered, which is reasonable for fare differences that are small relative to income.

The logsum also measures the attractiveness of a market. With the outside option normalized to zero, the probability of booking any itinerary is $1 - P_{n0} = e^{I_n^{\text{air}}}/(1 + e^{I_n^{\text{air}}})$, where I_n^{air} is the logsum over the air alternatives only. The decision to fly at all is a binary logit in the logsum. This is the bridge from itinerary choice to market size and stimulation in Chapter 4, and it is the structure that nested logit generalizes in Chapter 6.

WORKED EXAMPLE 5.3 What is a better departure time worth?

Use the estimated model of Example 5.2 and its 4,000 sessions to evaluate moving Laurentian's nonstop from 18:30 to 19:30, with fares, elapsed times and competitors unchanged. Only the departure displacement of the Laurentian alternative changes, by ± 1 hour or less depending on each traveller's preferred time, and Figure 5.4 traces the result for every departure time between 15:00 and midnight.

Shares. Averaged over sessions, Laurentian's share rises from 35.2 to 35.8 percent, competitor B's falls from 30.1 to 29.6 percent, the connection's edges up from 9.2 to 9.3 percent (travellers who wanted to leave early lose the 18:30 and some take the 17:15) and no booking falls from 25.5 to 25.3 percent. To the nearest quarter hour, nineteen-thirty is the share-maximizing time for this demand.

Consumer surplus. Applying (5.8) session by session with $\hat{\alpha} = 0.0040$ per dollar, expected surplus rises by CAD 2.31 per session on average; the rule of a half gives the same CAD 2.31. The average hides a distribution: travellers who wanted an early departure lose (the 5th percentile is a loss of CAD 25) and those who wanted a late one gain (the 95th percentile is a gain of CAD 25), and 59 percent of sessions are closer to the new time than to the old one.

Scaling up. Suppose the market generates 600 such sessions a day, consistent with the roughly 215 local passengers a day that the running case assigns to Laurentian's YUL–CDG flight at a 35 percent share. The retiming is then worth CAD 1,386 a day to travellers, about CAD 0.51 million a year. Laurentian gains 4.0 bookings and CAD 2,940 of revenue a day, about CAD 1.07 million a year, but only 1.1 of the 4.0 additional bookings are new to the market; the rest are taken from competitors.

Reading the result. Laurentian's revenue gain is twice the travellers' surplus gain, because most of it is a transfer from rivals rather than new value. The airline cares about its own revenue, and that is what the network plan should maximize, subject to cost. The logsum answers a different question, how much value the change creates for travellers, which matters when a planner argues for a slot or a route in a public process, when pricing and revenue management ask how much of the improvement

could be captured, and when a competition authority evaluates the effect of a schedule change on consumers. It also shows that the surplus-maximizing time (19:15, again to the nearest quarter hour) is a little earlier than the share-maximizing one: the travellers who gain most are those poorly served by the competitor's 20:45.

5.8 From probabilities to market shares

A choice model predicts probabilities for individual travellers. Planning needs passengers per day by itinerary, which requires adding up correctly.

Sample enumeration. Apply the model to a representative sample of travellers (here, shopping sessions) and sum: the expected number choosing i is $\hat{N}_i = \sum_n w_n P_{ni}$, where w_n is the number of travellers in the population that session n represents (Train 2009, ch. 2). **Segmentation** does the same with a small number of homogeneous groups: $\hat{N}_i = \sum_s N_s P_{si}$.

The average-traveller error. It is tempting to compute one probability for a traveller with average attributes. Because P_{ni} is a nonlinear, S-shaped function of utility, the probability at the average is not the average probability. The grey curve in Figure 5.4 shows how large the error can be. A single traveller with the average preferred departure time of 19:30 and average fares would give Laurentian 37.6 percent at 18:30 and 42.4 percent at 19:30, a gain of 4.8 points; the correct enumeration over 4,000 travellers gives a gain of 0.67 points. The representative traveller overstates the effect of the retiming sevenfold, because it puts every traveller's preferred time at the same point, whereas real preferences are spread over the evening. The same error inflates elasticities, as Example 5.2 showed. Enumerate over a sample, or over segments, that reproduces the spread of preferences and attributes.

Forecasting proceeds the same way: change the attributes in the sample to those of the future schedule and fares, reweight the sample to the future population, and recompute.

Recalibrating constants. Itinerary choice models are often estimated on data pooled across many markets or from an earlier period, then applied to a target market whose current shares are known from booking data. The constants absorb the average effect of everything the model leaves out, which differs between markets,

so they are recalibrated: starting from the estimated constants $\delta_j^{(0)}$, iterate

$$\delta_j^{(t+1)} = \delta_j^{(t)} + \ln(S_j / \hat{S}_j^{(t)}) - \ln(S_0 / \hat{S}_0^{(t)}),$$

where S_j is the observed share and $\hat{S}_j^{(t)}$ the share predicted with the current constants, until the two agree (Train 2009, ch. 2). The last term applies the same correction to the outside option and then shifts every constant so that the outside option's stays at zero. For a single representative traveller one step is exact; with heterogeneous travellers a few steps suffice.

WORKED EXAMPLE 5.4 Recalibrating to a new season

Suppose next season's booking data for YUL–CDG show shares of 24 percent no booking, 38 percent Laurentian, 28 percent B and 10 percent connection, against the model's 25.5, 35.2, 30.1 and 9.2 percent. Writing the three booking constants of the estimated model as $\delta = (5.399, 5.129, 4.224)$ for Laurentian, B and the connection (the air constant plus the Laurentian or connection term), one update moves them to $(5.537, 5.116, 4.368)$ and the predicted shares to 24.1, 37.8, 28.2 and 10.0 percent. A second update gives $(5.548, 5.115, 4.377)$ and shares that match the targets to the first decimal. In total the Laurentian constant rises by 0.150, the competitor's falls by 0.014 and the connection's rises by 0.153; the coefficients on fare, time and displacement are untouched, so the model's responses to schedule and fare changes are those estimated from the full data, anchored to the target market's current position.

Calibrated constants carry a warning. They absorb everything the model misses, including errors in the assumed market size and any distortion caused by IIA. A market that needs large adjustments is telling you that the model's variables do not explain it. And holding the calibrated constants fixed while the schedule changes assumes that the unobserved factors stay put, which is acceptable for a retiming and risky for the entry of a new carrier or a new hub.

IN PRACTICE **Estimating itinerary choice in 2026**

Three developments shape how itinerary choice models are estimated today. First, data: booking records such as MIDT and DDS show which itinerary was bought but not which alternatives were displayed or who searched without buying, whereas shopping data link searches to bookings and so supply both the choice set and an outside option. Working with search and booking data from European markets, Lhéritier et al. (2019) report that a machine-learning classifier predicted itinerary choices better than multinomial logit and latent class logit. Second, vendor planning systems continue to rely on the logit family: Lufthansa Systems' 2025 product information for NetLine/Plan lists multinomial and nested logit market models, with training options on PaxIS, DDS or MIDT data (Lufthansa Systems 2025c). Third, the econometric caution of the pitfall above has reached airline research: Lurkin et al. (2017) show that correcting for fare endogeneity changes fare coefficients and lowers the implied values of time. For the analyst the consequences are practical: know which data a model was trained on, whether its choice sets contain an outside option, and whether its fare coefficient was identified from exogenous variation.

5.9 Summary

- A random utility model explains choices as utility maximization with an unobserved component. Only utility differences are identified, constants are relative to a base alternative (here, not booking), and coefficients are scaled by the noise of unobserved utility.
- Independent Gumbel errors give the logit formula (5.1), derived either by integration or from the logistic and max-stability properties of the Gumbel distribution.
- The logit curve is S-shaped; its slope $P(1-P)$ is largest at even odds, so changes matter most where alternatives are close.
- IIA: the odds between two alternatives ignore all others, and a new alternative draws from existing ones in proportion to their shares. That fails for close substitutes such as codeshare duplicates and nearby frequencies, which are central to airline planning.

- Own elasticities are $\beta x(1-P)$ and cross elasticities $-\beta x_j P_j$. Market elasticities are much smaller than itinerary elasticities and depend on the outside share.
- Ratios of coefficients are money values: the value of time is β_T/β_f, and generalized cost is utility divided by the marginal utility of money. Endogenous fares inflate them.
- The logit log-likelihood is globally concave; Newton's method converges quickly, and with constants the model reproduces sample shares by construction.
- The logsum is the expected maximum utility. Divided by the marginal utility of money, its change values a schedule change for travellers; revenue gains for one airline are partly transfers from others.
- Aggregate by sample enumeration, never at the average traveller, and recalibrate constants to the target market's observed shares.

Table 5.4 Key formulas of the chapter.

Quantity	Formula
Logit probability	$P_{ni} = e^{V_{ni}} / \sum_j e^{V_{nj}}$
Odds (IIA)	$P_{ni}/P_{nk} = e^{V_{ni}-V_{nk}}$
Own, cross elasticity	$\beta_k x_{nik}(1-P_{ni})$, $-\beta_k x_{njk} P_{nj}$
Market elasticity	$\eta_{\text{air}} = \beta_f \bar{f}_n P_{n0}$
Value of attribute k	$v_k = -\beta_k/\beta_f$; VOT $= \beta_T/\beta_f$
Score	$g = \sum_n \sum_i (y_{ni} - P_{ni})\, x_{ni}$
Hessian	$H = -\sum_n \sum_j P_{nj}(x_{nj} - \bar{x}_n)(x_{nj} - \bar{x}_n)'$
Expected maximum utility	$\ln \sum_j e^{V_{nj}} + \gamma$
Change in consumer surplus	$\frac{1}{\alpha}[\ln \sum_j e^{V^1_{nj}} - \ln \sum_j e^{V^0_{nj}}]$
Constant recalibration	$\delta_j \leftarrow \delta_j + \ln(S_j/\hat{S}_j) - \ln(S_0/\hat{S}_0)$

5.10 Further reading

Belobaba's PODS lecture (Belobaba 2006g) frames passenger choice as a trade-off between fare, restrictions, path quality and schedule, the attributes this chapter has put into utility, and the handbook's demand section (Network Planning Handbook 2026) states the planner's version of the outside-option and IIA warnings. Train (2009, chs. 2 and 3) is the best single treatment of random utility, identification, logit, consumer surplus and estimation, and the model itself is McFadden's (1974). Ben-Akiva and Lerman (1985) remain the classic transport-oriented text. Small and Rosen (1981) establish the welfare interpretation of the logsum, and Hausman and McFad-

den (1984) give the standard test of IIA. For airline applications, Coldren et al. (2003) describe logit itinerary share models developed at a major US airline, Koppelman, Coldren and Parker (2008) model schedule delay in itinerary choice, Garrow (2010) surveys discrete choice modelling of air travel demand, and Lurkin et al. (2017) show what fare endogeneity does to itinerary choice estimates.

CHAPTER

6

Choice Models II: Nested, Mixed and Itinerary Choice

IN THIS CHAPTER

- The generalized extreme value family: one generating function, many patterns of substitution.
- Nested logit derived from its generating function, the dissimilarity parameter and its bounds, and the split into marginal and conditional logits.
- Mixed logit: random coefficients and error components, simulated probabilities, and how business and leisure heterogeneity breaks IIA.
- The Boeing decision window model: schedule tolerance, coverage, path quality and airline preference.
- The PODS passenger choice model: willingness-to-pay distributions and the disutility costs of restrictions and path quality.
- What network planning tools do with itinerary choice, and the questions an analyst should ask about them.

DRAWS ON Train, Discrete choice methods with simulation (2009), chapters 4 and 6; Boeing, Decision window path preference methodology (1993, MIT 16.75J lecture notes); Belobaba, Introduction to PODS passenger choice model and ePODS educational game (MIT 16.75J, 2006); handbook

Chapter 5 ended on a question that multinomial logit cannot answer honestly. When Laurentian adds a second evening frequency to Paris, whose passengers does it take: its own, the competitor's, or new travellers'? Under the independence of irrelevant alternatives the answer is fixed in advance by the base shares. For a network planner that is not a detail. A frequency that mostly cannibalizes the existing flight is worth a fraction of one that wins passengers from a competitor, and the recapture rates that drive spill estimates

(Chapter 9) and itinerary-based fleet assignment (Chapter 14) are statements about exactly this substitution.

This chapter develops the models that let the data, or an explicit description of behaviour, decide. Two econometric families relax IIA in complementary ways. Generalized extreme value models, of which nested logit is the workhorse, group alternatives that share unobserved attributes. Mixed logit lets tastes differ across travellers, so that business and leisure travellers substitute differently. We then turn to two models built inside the airline industry that encode schedules and willingness to pay more literally: Boeing's decision window model and the passenger choice model of the Passenger Origin-Destination Simulator (PODS), which Belobaba's 2006 lecture calls the widely recognized state of the art in simulating revenue management in competitive markets (Belobaba 2006h). The chapter closes with how planning tools use these models and what an analyst should ask about them.

The running example continues from Chapter 5: Laurentian's 18:30 nonstop from Montreal to Paris, competitor B's 20:45 nonstop, carrier C's 17:15 one-stop via London Heathrow, and the option of not booking. All Laurentian numbers are illustrative.

6.1 Correlated alternatives and the GEV family

Logit's substitution pattern comes from its assumption that unobserved utility is independent across alternatives. Itineraries violate it in recognizable ways. Two departures of the same airline share the frequent flyer program, the lounge and the cabin product. Two departures in the late evening share their fit with a traveller's working day. Two connections over the same hub share its reliability. A traveller with a strong unobserved liking for one member of such a group probably likes the others too, and when one of them improves, it draws disproportionately from its group. The generalized extreme value (GEV) family keeps the convenience of closed-form probabilities while allowing this correlation (McFadden 1978).

DEFINITION 6.1 GEV generating function

Let $Y_j > 0, j = 1, \dots, J$. A function $G(Y_1, \dots, Y_J)$ is a **GEV generating function** if (i) $G \geq 0$; (ii) G is homogeneous of degree one, $G(\rho Y) = \rho\, G(Y)$ for $\rho > 0$; (iii) $G \to \infty$ when any $Y_j \to \infty$; and (iv) its partial derivatives alternate in sign: $G_i = \partial G/\partial Y_i \geq 0$, $G_{ij} = \partial G_i/\partial Y_j \leq 0$ for $i \neq j$, $G_{ijk} \geq 0$ for distinct i, j, k, and so on. Then

$$F(\varepsilon_1, \dots, \varepsilon_J) = \exp(-G(e^{-\varepsilon_1}, \dots, e^{-\varepsilon_J}))$$

is a joint distribution function for the unobserved utilities, each of whose marginals is Gumbel.

RESULT 6.1 GEV choice probabilities and expected maximum utility

If $U_j = V_j + \varepsilon_j$ with ε distributed as F, and $Y_j = e^{V_j}$, then

$$P_i = \frac{Y_i\, G_i(Y)}{G(Y)}, \qquad \mathbb{E}\Big[\max_j U_j\Big] = \ln G(Y) + \gamma. \tag{6.1}$$

Proof. Alternative i is chosen with its error in $[s, s + \mathrm{d}s]$ when ε_i falls there and every other ε_j lies below $s + V_i - V_j$. That probability is the partial derivative $\partial F/\partial \varepsilon_i$ evaluated at $\varepsilon_j = s + V_i - V_j$ for all j, times $\mathrm{d}s$. Differentiating F gives $\partial F/\partial \varepsilon_i = F\, G_i(e^{-\varepsilon})\, e^{-\varepsilon_i}$. At the evaluation point, $e^{-\varepsilon_j} = \tau Y_j$ with $\tau = e^{-s-V_i}$, the same factor for every j. Homogeneity of degree one makes $G(\tau Y) = \tau G(Y)$, and its derivative G_i is homogeneous of degree zero, so $G_i(\tau Y) = G_i(Y)$. The integrand is therefore $\exp(-\tau G)\, G_i\, \tau Y_i$. Changing variable from s to τ, with $\mathrm{d}s = -\,\mathrm{d}\tau/\tau$,

$$P_i = \int_0^\infty e^{-\tau G}\, G_i\, Y_i\, \mathrm{d}\tau = \frac{Y_i G_i}{G}.$$

For the maximum, $\Pr(\max_j U_j \leq u) = F(u - V_1, \dots, u - V_J) = \exp(-G(e^{-u}Y)) = \exp(-e^{-u}G(Y))$, a Gumbel distribution with location $\ln G(Y)$, whose mean is $\ln G(Y) + \gamma$. ∎

Logit is the simplest member: with $G = \sum_j Y_j$, $G_i = 1$ and (6.1) returns $e^{V_i}/\sum_j e^{V_j}$ and the logsum of Chapter 5. Every GEV model has a logsum, $\ln G$, which values changes in the choice set for consumer surplus exactly as in Result 5.6. Different generating functions give

different substitution patterns; the one that matters most in practice groups alternatives into nests.

6.2 Nested logit

Partition the alternatives into nests $B_1, \dots, B_K$ and give nest k a parameter λ_k. The **nested logit** model uses

$$G(Y) = \sum_{k=1}^{K} \Big(\sum_{j \in B_k} Y_j^{1/\lambda_k} \Big)^{\lambda_k}.$$

Conditions (i) to (iii) of Definition 6.1 are immediate. For (iv), the first derivative for $i \in B_k$ is positive, and the cross derivative for $m \in B_k$, $m \neq i$,

$$G_{im} = \frac{\lambda_k - 1}{\lambda_k} (Y_i Y_m)^{1/\lambda_k - 1} \Big(\sum_{j \in B_k} Y_j^{1/\lambda_k} \Big)^{\lambda_k - 2},$$

is negative when $\lambda_k < 1$ (it is zero for alternatives in different nests); the higher derivatives alternate as required when $0 < \lambda_k \leq 1$.

RESULT 6.2 Nested logit probabilities

For alternative i in nest B_k, let $I_k = \ln \sum_{j \in B_k} e^{V_j/\lambda_k}$ be the **inclusive value** of the nest. Then

$$P_i = \frac{e^{V_i/\lambda_k} \big(\sum_{j \in B_k} e^{V_j/\lambda_k} \big)^{\lambda_k - 1}}{\sum_{l=1}^{K} \big(\sum_{j \in B_l} e^{V_j/\lambda_l} \big)^{\lambda_l}}, \tag{6.2}$$

which factors into a marginal probability of the nest and a conditional probability within it, $P_i = P_k \, P_{i|k}$, with

$$P_k = \frac{e^{\lambda_k I_k}}{\sum_l e^{\lambda_l I_l}}, \qquad P_{i|k} = \frac{e^{V_i/\lambda_k}}{\sum_{j \in B_k} e^{V_j/\lambda_k}}. \tag{6.3}$$

Proof. Write $D_k = \sum_{j \in B_k} Y_j^{1/\lambda_k} = e^{I_k}$. For $i \in B_k$, $G_i = \lambda_k D_k^{\lambda_k - 1} \cdot \frac{1}{\lambda_k} Y_i^{1/\lambda_k - 1} = Y_i^{1/\lambda_k - 1} D_k^{\lambda_k - 1}$, so $Y_i G_i = Y_i^{1/\lambda_k} D_k^{\lambda_k - 1}$. Dividing by $G = \sum_l D_l^{\lambda_l}$ and substituting $Y_j = e^{V_j}$ gives (6.2). For the decomposition, write the numerator as $(Y_i^{1/\lambda_k}/D_k) \cdot D_k^{\lambda_k}$: the first factor is $P_{i|k}$ and the second is $e^{\lambda_k I_k}$. ∎

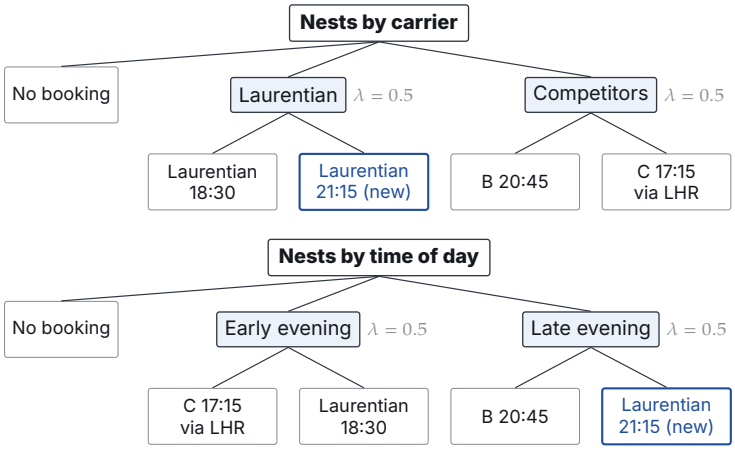


Figure 6.1 Two nesting structures for the YUL–CDG choice set after Laurentian adds a 21:15 nonstop. Top: nests by carrier, so that the two Laurentian departures are close substitutes. Bottom: nests by time of day, so that the new flight competes most closely with the competitor's 20:45. The outside option forms its own nest. The dissimilarity parameters are illustrative.

The dissimilarity parameter. The parameter λ_k, also called the log-sum coefficient, measures how different the alternatives in nest k are from one another. For two alternatives in the same nest, the joint distribution of their errors is $\exp\left(-\left(e^{-\varepsilon_i/\lambda_k}+e^{-\varepsilon_j/\lambda_k}\right)^{\lambda_k}\right)$, whose correlation is $1-\lambda_k^2$; errors in different nests are independent. With $\lambda_k = 1$ the alternatives are independent and nested logit collapses to multinomial logit. As $\lambda_k \to 0$ they become perfectly correlated, and the nest behaves like a single alternative whose utility is that of its best member. Values in $(0, 1]$ are consistent with utility maximization for all data; values above 1 are consistent only over a limited range of the explanatory variables, and non-positive values never are (Train 2009, ch. 4). Because logit is the special case $\lambda_k = 1$ for all k, IIA can be tested by a likelihood ratio or t test of that restriction.

Two logits, upper and lower. The decomposition (6.3) gives the model its intuitive reading. The lower model is a logit among the alternatives of a nest, with utilities divided by λ_k: within a nest of close substitutes, small differences in V produce large differences in probability. The upper model is a logit among nests, in which each nest's attractiveness is $\lambda_k I_k$, its expected maximum utility. A nest gains share when any of its members improves, but much of the improvement is absorbed inside the nest. Within a nest IIA holds; across nests the odds depend on the members of the two nests concerned and on no others.

The model can be estimated by maximum likelihood with (6.2) as the choice probability. The log-likelihood is no longer globally concave, so starting values matter. Estimating the lower models first and then the upper model with the estimated inclusive values is consistent but inefficient and understates standard errors; joint estimation is preferable (Train 2009, ch. 4).

PITFALL

Two different models are sold under the name nested logit. In the one derived here, utilities inside a nest are divided by λ_k. Some software estimates a variant that omits the division; when coefficients such as the fare coefficient are shared across nests, that variant is not consistent with utility maximization and its substitution patterns differ (Train 2009, ch. 4). When a planning tool or a consultant reports nested logit results, ask which form was used.

The pattern of substitution is clearest in elasticities.

RESULT 6.3 Nested logit elasticities

Let attribute x_m of alternative m enter utility with coefficient β, and let $m \in B_k$. The elasticity of P_i with respect to x_m is

$$\eta_{im} = \begin{cases} \beta x_m [(1 - P_m) + (\frac{1}{\lambda_k} - 1)(1 - P_{m|k})] & i = m, \\ -\beta x_m [P_m + (\frac{1}{\lambda_k} - 1) P_{m|k}] & i \neq m,\ i \in B_k, \\ -\beta x_m P_m & i \notin B_k. \end{cases}$$

Proof. From (6.2), for $i \in B_l$, $\ln P_i = V_i/\lambda_l + (\lambda_l - 1) I_l - \ln \sum_h e^{\lambda_h I_h}$. For $m \in B_k$, $\partial I_k / \partial V_m = P_{m|k}/\lambda_k$, and the derivative of the last term is $P_k P_{m|k} = P_m$. Hence $\partial \ln P_i / \partial V_m$ equals $1/\lambda_k + (\lambda_k - 1) P_{m|k}/\lambda_k - P_m$ when $i = m$, $(\lambda_k - 1) P_{m|k}/\lambda_k - P_m$ when $i \neq m$ is in the same nest, and $-P_m$ otherwise. Rearranging and multiplying by $\partial V_m / \partial \ln x_m = \beta x_m$ gives the three cases. ∎

Relative to logit, each own elasticity gains the term $(1/\lambda_k - 1)(1 - P_{m|k})$ and each cross elasticity within the nest gains $(1/\lambda_k - 1) P_{m|k}$; with $\lambda_k = 1$ both vanish. Alternatives outside the nest see exactly the logit cross elasticity. An improvement in one member of a nest of close substitutes draws mainly from its nest-mates.

Airline research uses exactly these structures. Building on the multinomial itinerary share models of Coldren et al. (2003), Coldren and Koppelman (2005) developed GEV models of itinerary shares, including nested structures that group itineraries by time of day, by carrier and by level of service, so that the data rather than the analyst decide which itineraries compete most closely.

WORKED EXAMPLE 6.1 The second frequency to Paris, revisited

Return to Example 5.1. Base shares are 23.1 percent no booking, 34.5 percent Laurentian 18:30, 29.7 percent competitor B 20:45 and 12.7 percent carrier C 17:15 via LHR. We compare four models, each calibrated to reproduce these base shares exactly, then add Laurentian's 21:15 nonstop with the same representative utility as its 18:30. All dissimilarity parameters are set to $\lambda = 0.5$ for illustration.

Calibration. With the outside option alone in its nest and $V_0 = 0$, (6.3) can be inverted: for j in nest B_k with total share S_k,

$$V_j = \ln(S_k / S_0) + \lambda_k \ln(S_j / S_k).$$

For the nests by time of day of Figure 6.1, the early-evening nest holds Laurentian and the connection with $S_k = 0.472$, which gives $V = 0.557$ for Laurentian and 0.057 for the connection, while the competitor alone in the late-evening nest keeps $V = \ln(0.297/0.231) = 0.25$.

Results. Table 6.1 and Figure 6.2 show what happens when the 21:15 is added.

The four models agree perfectly on today and disagree sharply on tomorrow. Laurentian's gain ranges from 8.2 to 21.1 points. Nested by carrier, the two Laurentian flights are close substitutes and 62 percent of the new flight's passengers come from the 18:30: the frequency mostly redistributes Laurentian's own customers. Nested by time of day, the 21:15 competes most closely with the competitor's 20:45 and takes 56 percent of its passengers from B. Nesting all air alternatives against not booking halves the stimulation that logit predicts, from 23 to 13 percent of the new flight's traffic; this is the structure behind the stimulation rule of Chapter 4, which assumes that travellers who fly are a small fraction of the potential market. Here they are 77 percent, so stimulation is small under every model. IIA no longer holds across nests: the ratio of Laurentian 18:30 to B moves from 1.16 to 0.82 under carrier nesting and to 1.96 under time-of-day nesting, while logit keeps it at 1.16.

The elasticities show the same structure. With carrier nesting, after the 21:15 is added the 18:30 has a share of 21.3 percent and half of Laurentian's nest. At a fare of CAD 760 and $\beta_f = -0.004$ per dollar, Result 6.3 gives an own-fare elasticity of $-0.004 \times 760 \times [0.787 + 1 \times 0.5] = -3.91$, against -2.26 under logit; the 21:15's elasticity with respect to the 18:30's fare is $+2.17$, while the competitor, the connection and the outside option see only $+0.65$. A fare cut on one Laurentian departure would mostly move passengers between Laurentian's two flights.

The planner's lesson is that the nesting structure is not a technicality. It is an assumption about which itineraries compete, it decides whether a frequency is worth adding, and it should be estimated, by comparing the likelihood of competing structures and testing $\lambda_k = 1$, rather than chosen.

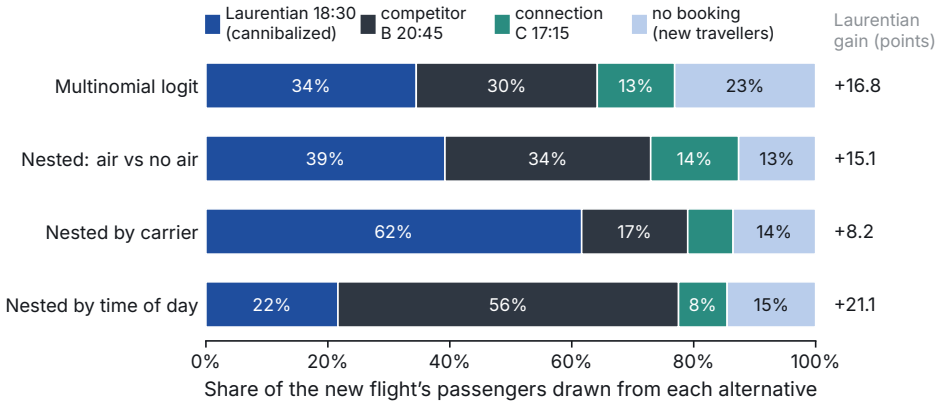


Figure 6.2 Origin of the new 21:15 flight's passengers under four choice models calibrated to the same base shares on YUL–CDG. The figures on the right are Laurentian's total gain in share points.

Table 6.1 Laurentian's total share after adding a 21:15 nonstop to YUL–CDG, and the origin of the new flight's passengers, under four models calibrated to the same base shares: multinomial logit, all air alternatives nested against not booking, nests by carrier and nests by time of day ($\lambda = 0.5$ in every nest).

	Logit	Air nest	Carrier nests	Time nests
Laurentian share after	51.3%	49.6%	42.7%	55.6%
Gain (points)	16.8	15.1	8.2	21.1
New flight's passengers from:				
Laurentian 18:30	34.5%	39.2%	61.6%	21.7%
Competitor B 20:45	29.7%	33.7%	17.4%	55.9%
Connection C 17:15	12.7%	14.4%	7.4%	8.0%
No booking (new travellers)	23.1%	12.6%	13.5%	14.5%

6.3 Mixed logit

Nested logit groups alternatives; it says nothing about differences between travellers. Yet the most important heterogeneity in airline markets is between travellers: a consultant flying on her client's account and a family going on holiday weigh fare, time and flexibility very differently. A model with a single fare coefficient averages them, and the average traveller does not exist.

DEFINITION 6.2 Mixed logit

A **mixed logit** model has choice probabilities

$$P_{ni} = \int L_{ni}(\beta) f(\beta \mid \theta)\, \mathrm{d}\beta, \qquad L_{ni}(\beta) = \frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}},$$

a weighted average of logit probabilities over a **mixing distribution** $f(\beta \mid \theta)$ whose parameters θ (for example the means and variances of the coefficients) are estimated. When the mixing distribution is discrete, with segments s of weight π_s and coefficients β_s, $P_{ni} = \sum_s \pi_s L_{ni}(\beta_s)$: the **latent class** model.

Two readings of the same formula guide specification (Train 2009, ch. 6). In the **random coefficients** reading, $U_{nj} = \beta_n' x_{nj} + \varepsilon_{nj}$ with β_n varying across travellers according to f; conditional on her own β_n the traveller is a logit chooser, and the analyst, who does not know β_n, integrates over it. In the **error components** reading, $U_{nj} = \alpha' x_{nj} + \mu_n' z_{nj} + \varepsilon_{nj}$, where the random terms μ_n with covariance W create covariance $z_{ni}' W z_{nj}$ between the unobserved utilities of alternatives i and j. A random term shared by all alternatives of a nest, with variance σ_k^2, produces within-nest correlation $\sigma_k^2/(\sigma_k^2+\pi^2/6)$, a mixed logit analogue of nested logit. The two readings are the same model: decompose β_n into its mean α and a deviation μ_n. Their generality is complete in a precise sense: McFadden and Train (2000) show that any random utility model can be approximated arbitrarily well by a mixed logit.

RESULT 6.4 Substitution in mixed logit

In mixed logit the ratio $P_{ni}/P_{nk} = \int L_{ni} f \, \mathrm{d}\beta / \int L_{nk} f \, \mathrm{d}\beta$ depends on every alternative, because the logit denominators sit inside the integrals. For an attribute x_{njm} of alternative $j \neq i$ with random coefficient β_m,

$$\eta_{ij} = -\frac{x_{njm}}{P_{ni}} \int \beta_m \, L_{ni}(\beta) \, L_{nj}(\beta) \, f(\beta \mid \theta) \, \mathrm{d}\beta. \tag{6.4}$$

The cross elasticity is largest for the alternatives whose logit probabilities move together with L_{nj} across the distribution of tastes.

Proof. Differentiate under the integral sign, using $\partial L_{ni}/\partial x_{njm} = -\beta_m L_{ni} L_{nj}$ from Result 5.4, and multiply by x_{njm}/P_{ni}. ∎

The integral in (6.4) is the whole story of how heterogeneity breaks IIA. An improvement in alternative j appeals to the travellers who like j; it draws from the alternatives those same travellers would otherwise choose. If cheap itineraries are chosen by price-sensitive travellers, a new cheap itinerary draws mainly from the other cheap itineraries.

Simulation. For continuous mixing distributions the integral has no closed form, but it is an expectation and can be simulated: draw $\beta^1, \ldots, \beta^R$ from $f(\beta \mid \theta)$ and average,

$$\check{P}_{ni} = \frac{1}{R} \sum_{r=1}^{R} L_{ni}(\beta^r).$$

The simulator is unbiased, strictly positive, smooth in θ and sums to one over alternatives, which makes it suitable for estimation by **maximum simulated likelihood**: maximize $\sum_n \sum_i y_{ni} \ln \check{P}_{ni}$. Because the logarithm is concave, $\ln \check{P}_{ni}$ is biased downward for finite R; the estimator is consistent when R rises with the sample size, and asymptotically equivalent to maximum likelihood when R rises faster than $\sqrt{N}$ (Train 2009, ch. 10). Quasi-random sequences such as Halton draws cover the distribution more evenly than independent draws and reduce simulation error dramatically for the same R.

WORKED EXAMPLE 6.2 Business and leisure travellers to Paris

Suppose YUL–CDG travellers come in two segments. Business travellers, 30 percent of sessions, have a fare coefficient of −0.0010 per dollar; leisure travellers, 70 percent, have −0.0070. Each segment has its own constants (business 3.40, 2.80 and 1.40 for Laurentian, B and C; leisure 6.50, 6.10 and 4.90), and the average fares are CAD 900 for Laurentian, 800 for B and 600 for the connection. Table 6.2 gives the segment shares. Business travellers are 52.5 percent of Laurentian's passengers but only 10.9 percent of the connection's.

An analyst who does not know about the segments fits a single logit to 200,000 simulated sessions from this population, with fares varying by about 15 percent around their averages. The fitted fare coefficient is −0.00453 per dollar, between the two segments' values, and the fitted constants reproduce the sample shares. On today's market the two models agree. They part company as soon as something changes.

Fare response. At CAD 900, Laurentian's own-fare elasticity is −2.59 in the two-segment model and −2.79 in the single logit. The two-segment value is the share-weighted average of −0.42 for business and −4.99 for leisure travellers, with weights 0.525 and 0.475. The left panel of Figure 6.3 shows the difference away from today's fare: at CAD 1,500, the two-segment model keeps 11.9 percent of the market for Laurentian, the single logit only 3.0 percent, because business travellers keep flying. A single logit undervalues the high end of the fare structure, the part that fences and revenue management exist to protect (Chapters 17 and 21).

Entry. A low-fare carrier opens a nonstop at CAD 560 with a product that appeals to leisure travellers but not to business travellers (constants 5.16 and 1.30). In the two-segment model it wins 28.4 percent of the market: 37.0 percent of leisure travellers and 8.4 percent of business travellers. Give the single logit a constant for the entrant that produces the same 28.4 percent. The right panel of Figure 6.3 compares where the entrant's share comes from. The single logit takes 28.4 percent of every alternative's share, by Result 5.3. The two-segment model takes 34.1 percent of the no-booking share (price-sensitive leisure travellers who now fly), 33.9 percent of the connection's, 27.5 percent of B's and only 22.0 percent of Laurentian's. The sin-

gle logit overstates Laurentian's loss by almost 30 percent and understates the connection's. The ratio of Laurentian's share to B's moves from 1.04 to 1.12 in the two-segment model and does not move at all in the single logit.

Simulation. If, within the leisure segment, the fare coefficient itself varies (minus a lognormal variable with mean 0.007 per dollar and coefficient of variation 0.5), the probability that a leisure traveller chooses Laurentian is 22.8 percent, computed exactly by quadrature, rather than the 20.8 percent of a logit at the mean coefficient: heterogeneity raises the probability of the expensive option, because some leisure travellers are much less price-sensitive than average. Figure 6.4 shows the cost of simulating it. With independent draws the root mean squared error is 0.042 at $R = 10$ and 0.0042 at $R = 1{,}000$, a tenfold improvement for a hundredfold increase in draws. With Halton draws it is 0.0028 already at $R = 100$ and 0.00026 at $R = 1{,}000$, sixteen times smaller than with independent draws.

The planner's lesson: the same base shares can hide very different travellers, and a model that averages them will mispredict exactly the events planners care about, a competitor's entry, a fare change, a new product at either end of the market.

Table 6.2 Shares by segment in the two-segment YUL–CDG model of Example 6.2 at average fares of CAD 900 (Laurentian), 800 (B) and 600 (connection). Illustrative values.

Segment	No booking	Laurentian	Competitor B	Connection C
Business (30%)	4.4%	53.4%	32.4%	9.8%
Leisure (70%)	17.0%	20.8%	28.0%	34.2%
All travellers	13.2%	30.6%	29.3%	26.9%

6.4 The Boeing decision window model

Logit-family models treat time of day as one more attribute in utility. In the early 1990s Boeing took a more literal view: a traveller has a window of time within which a trip is acceptable, and only itineraries inside it are considered at all. The method is documented in a 1993 presentation included in the MIT airline management lecture notes (Boeing Commercial Airplane Group 1993), and Belobaba's history of PODS traces the simulator's passenger choice model to it (Belobaba 2006h). Boeing framed market share eval-

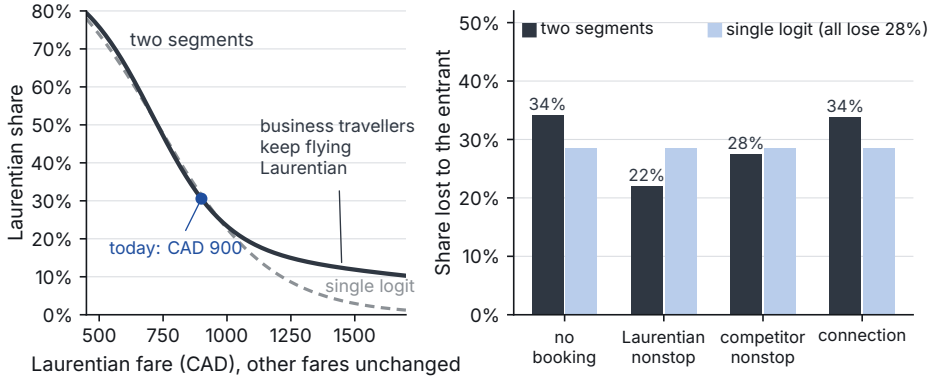


Figure 6.3 Business and leisure heterogeneity on YUL–CDG. Left: Laurentian's share as its fare changes, in the two-segment model and in a single logit fitted to the same travellers. Right: share of each alternative's base share lost to a low-fare entrant that wins 28.4 percent of the market in both models.

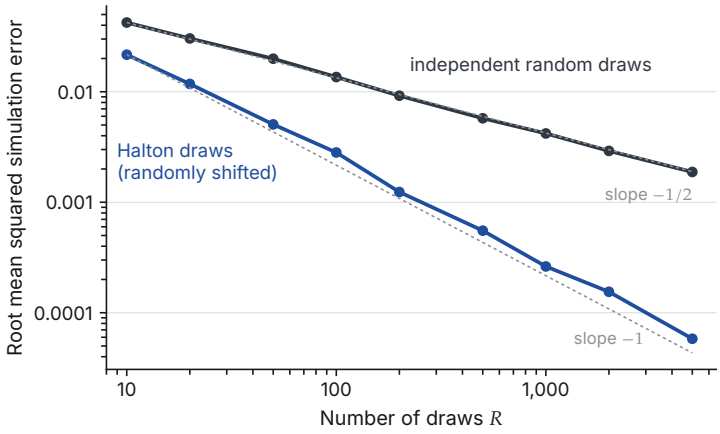


Figure 6.4 Root mean squared error of the simulated probability that a leisure traveller chooses Laurentian, when her fare coefficient is minus a lognormal variable with mean 0.007 per dollar and coefficient of variation 0.5, over 300 replications at each number of draws. Independent draws converge at the rate $R^{-1/2}$; randomly shifted Halton draws converge close to the rate R^{-1}.

uation as three tasks, demand estimation, path generation and allocation, with the decision window model doing the allocation: the same chain that today's planning systems run (Section 6.6). All of its parameter values are historical.

DEFINITION 6.3 Decision window and path quality index

A traveller's **decision window** runs from the earliest departure time she will accept to the latest arrival time, on her preferred day. Its width is the market's **Delta-T**, the local arrival time minus the local departure time of the best service in the market, plus her **schedule tolerance**. A path is considered only if it fits entirely within the window. The **path quality index** (PQI) of a path is 1 for a nonstop plus one point for each stop, each connection and each interline connection: a one-stop through flight scores 2, an online connection 3, an interline connection 4 and a path with two interline connections 7.

Boeing measured window width with Delta-T rather than distance because its data showed travellers responding to the clock times they experience, which include time-zone changes: a trip from London to Seattle has a Delta-T of about two hours. The model has three steps (Boeing Commercial Airplane Group 1993).

Step 1: place and size each window. Day-of-week demand picks the day, time-of-day demand picks the position, and the tolerance, drawn from a distribution estimated from surveys, sets the width beyond Delta-T. A market is a population of windows, some narrow and some wide.

Step 2: find the paths in each window. For each window the model records which paths fit, its **coverage state**. In Boeing's illustration with two YY flights and one ZZ flight, 15 percent of travellers had only the first YY flight in their window, 30 percent only the second, 35 percent the second YY flight and the ZZ flight, 5 percent both YY flights, 10 percent all three and 5 percent none; two of the eight combinations could not occur given the timings. A traveller with no path in the window must re-plan, either with another flight the same day or with the same time on another day. A small share of travellers, rising with distance to about a fifth at 6,000 miles, also discriminate by aircraft and drop paths not flown by the best aircraft category in their window. Only the best path of each airline, by path quality, is carried forward.

Step 3: choose among the paths in the window. Each airline has an **integrated image** built from ratings, on a scale from 1 (extremely poor) to 7 (excellent) with a baseline airline at 4, on four factors whose weights depend on stage length: availability and reliability, marketing programs, service quality and passenger environment, weighted 31, 14, 31 and 24 percent for trips under six hours and 11, 18, 36 and 35 percent for longer ones. The weighted rating is turned into a coefficient of preference, scaled so that an airline's coefficient divided by the sum of its own and the baseline's equals the probability that the airline is rated above the baseline. The **market share equation** then gives each airline with a path in the window a share equal to its coefficient divided by the sum of the coefficients. Path quality enters through decision orientation. Schedule-oriented travellers first restrict attention to the best path quality available and then choose an airline; airline-oriented travellers choose their airline first and then its best path. The schedule-oriented share falls with distance, from about 85 percent at the shortest ranges to under 10 percent at 6,000 miles. Boeing combined the two orientations into service-based coefficients by path quality and range, which Chapter 4 reproduces: a nonstop scores 100 at every range, and an online connection scores 3.78 at 1,000 miles, 24.82 at 3,000 miles and 76.96 at 6,000 miles. Within an airline, paths of equal quality split its share: with two equally convenient YY nonstops and one ZZ nonstop in a window and equal airline preference, each airline gets half, and each YY flight a quarter.

The coefficients encode a familiar fact: the penalty for a stop or connection falls with distance. At 1,000 miles a one-stop path is worth a fifth of a nonstop; at 6,000 miles, almost nine-tenths.

WORKED EXAMPLE 6.3 Decision windows on YUL–CDG

Boeing's examples. With short-range weights, airline ZZ, rated 5, 3, 4 and 6 on the four factors, has a combined rating of $0.31 \times 5 + 0.14 \times 3 + 0.31 \times 4 + 0.24 \times 6 = 4.65$, and airline CC, rated 4, 4, 4 and 5, has 4.24. Boeing's coefficients of preference are 100 for the baseline YY, 309.52 for ZZ and 150.60 for CC, so a traveller with all three in her window chooses them with probabilities 0.1785, 0.5526 and 0.2689 (Boeing Commercial Airplane Group 1993). Two observations connect this to Chapter 5. First, the market share equation is a logit with utilities $V_a = \ln \text{Coef}_a$: ZZ's utility exceeds YY's by $\ln 3.0952 = 1.13$, and CC's exceeds YY's by $\ln 1.5060 = 0.41$. Sec-

ond, the coefficients are consistent with comparing normally distributed ratings: $309.52/409.52 = 0.756 = \Phi(0.65/0.938)$ and $150.60/250.60 = 0.601 = \Phi(0.24/0.938)$, the same spread of 0.938 rating points for both airlines. In Boeing's 1,000-mile example, two airlines of equal image offer a nonstop and a one-stop path; the 67.4 percent of travellers who are schedule-oriented all take the nonstop, and the airline-oriented 32.6 percent split evenly, so the nonstop wins $67.4 + 16.3 = 83.7$ percent, which is $100/(100 + 19.45)$: Boeing's service-based coefficient for a one-stop path at 1,000 miles is 19.45.

Laurentian's market. Apply the method to YUL–CDG (3,432 statute miles), with equal airline images. The Delta-T of Laurentian's 18:30 nonstop, arriving at 07:24 local time, is 12.9 hours, which sets the market Delta-T; B's 20:45 arrives at 09:51 (13.1 hours) and C's 17:15 via LHR at 09:45 (16.5 hours), so the connection fits only windows with at least 3.6 hours of tolerance. Interpolating Boeing's coefficients linearly between 3,000 and 4,000 miles, the online connection's coefficient is 32.5 against 100 for a nonstop. For illustration, let earliest departure times be normal with mean 17:30 and standard deviation 1.5 hours, and tolerances follow a gamma distribution with mean 4.5 hours. Simulating 200,000 travellers gives the coverage states: 26.7 percent have only Laurentian in their window, 21.0 percent only B, 20.7 percent both nonstops, 16.9 percent all three paths, 0.6 percent Laurentian and the connection, and 14.1 percent nothing, so they must re-plan. Laurentian wins 44.8 percent of all travellers, B 38.6 percent and C 2.5 percent.

The second frequency. Add Laurentian's 21:15 (Delta-T 12.9 hours). Laurentian's share rises to 55.7 percent, a gain of 10.9 points: 10.0 points from B and 0.9 points from travellers who previously had no path, nothing from the connection. The new flight carries 18.9 percent of travellers, but 8.0 points of that come from the 18:30, because the 34.2 percent of travellers who have both Laurentian flights in their window simply split their Laurentian choice between them. The decision window model behaves like a nested logit with three special features: alternatives outside the window are excluded, equal paths of one airline are perfect substitutes, and frequency raises an airline's share only by covering windows that were not covered before.

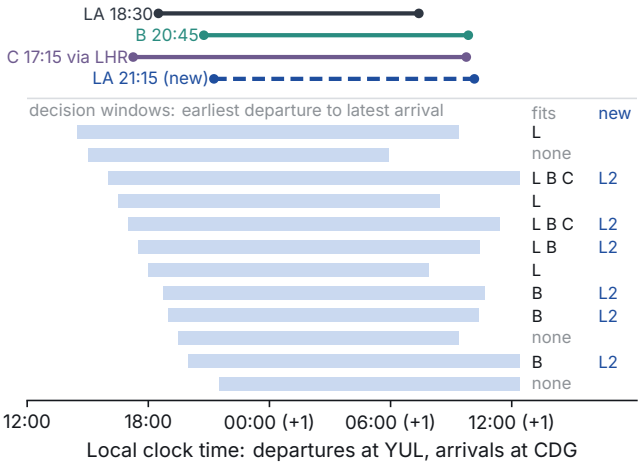


Figure 6.5 Decision windows on YUL–CDG. Paths run from local departure at YUL to local arrival at CDG; each bar is one illustrative traveller's window, from earliest departure to latest arrival. A path is considered only if it lies entirely inside the window. The right-hand columns list the paths that fit today (L: Laurentian 18:30, B: competitor 20:45, C: connection via LHR) and whether Laurentian's proposed 21:15 (L2) would fit.

Figure 6.5 makes Boeing's observations visible (Boeing Commercial Airplane Group 1993). More frequencies at different times raise the chance that a window contains at least one of an airline's paths, but frequencies saturate: a second flight inside windows already covered adds nothing. Short paths fit more windows than long ones. Timing matters: a departure at a popular time sits in many windows. These are the mechanics behind the frequency-share relationship of Chapter 4 and the timetable design of Chapter 13.

6.5 The PODS passenger choice model

The Passenger Origin-Destination Simulator was developed by Boeing in the early 1990s to simulate how passengers choose among airlines and paths, using the decision window model, and was developed jointly with MIT from 1994, adding choice among fare products and realistic revenue management systems (Belobaba 2006h). It simulates competing airlines through many repetitions of a booking process: before departure, the booking horizon is divided into 16 time frames; at the start of each, every airline's revenue management system sets availability using forecasts built only from previously simulated history; passengers arrive at random and choose among what is open; the historical database is updated (Belobaba 2006g).

Because it simulates the interaction of competing revenue management systems with passenger choice (Belobaba 2006h), PODS can measure effects that leg-level models cannot, such as buy-down (Chapter 21). Here we look at its passenger model.

PODS generates total daily demand for each origin-destination market and two passenger types, business and leisure, with booking curves over the 16 time frames. In the lecture's illustration, leisure passengers have made about three-quarters of their bookings three weeks before departure, business passengers about a third. The types differ in three ways: time-of-day demand and schedule tolerance, which set their decision windows; maximum out-of-pocket willingness to pay; and the disutility, in dollars, that they attach to fare restrictions, to path quality and to re-planning. Willingness to pay and disutilities are drawn for each simulated passenger from normal distributions whose means and k-factors are inputs (Belobaba 2006g).

The choice rule is lexicographic, then compensatory. A passenger first discards every path and fare combination whose lowest available fare exceeds her maximum willingness to pay. Among the rest, she picks the one with the lowest **total cost**: fare plus restriction disutilities, plus a path quality cost proportional to the path quality index (1 for a nonstop, 3 for a connection in PODS), plus a re-planning cost if the path lies outside her decision window, plus a cost for an unfavoured airline. Path quality, re-planning and airline costs are set as functions of the market's base fares.

Willingness to pay. The lecture's example formulation makes the probability of paying at least f a shifted exponential,

$$\Pr(\text{WTP} \geq f) = \min\Big\{1,\ \exp\Big(-\ln 2\,\frac{f-b}{(m-1)\,b}\Big)\Big\}, \qquad (6.5)$$

where b is a base fare (the lowest, Q, fare for leisure passengers and 2.5 times the Q fare for business passengers) and m is a multiplier, 1.2 for leisure and 3 for business (Belobaba 2006g). Everyone will pay the base fare. Setting the exponent to $-\ln 2$ shows that m fixes the median: half of all passengers will pay mb or more, that is 1.2 Q fares for leisure passengers and 7.5 Q fares for business passengers. In a structure with fares at 1, 1.5, 2 and 4 times the Q fare, (6.5) implies that 17.7 percent of leisure passengers would pay 1.5 times the Q fare and 3.1 percent twice it, while 81.2 percent of business passengers would pay the full fare at four times the Q fare. The left panel of Figure 6.6 draws both curves. The formula is one of two illustrations in the lecture, and the two do not match: with a Q fare of 200 it implies mean willingness to pay of about 1,943 for busi-

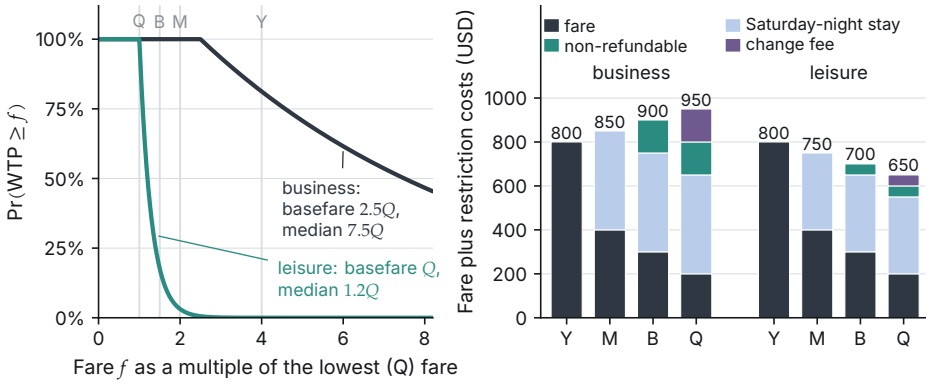


Figure 6.6 The PODS passenger choice model with historical lecture settings. Left: probability that a passenger's maximum willingness to pay is at least f, from (6.5), with fares measured in multiples of the Q fare. Right: mean total cost of each fare class, fare plus restriction disutilities, for the average business and leisure passenger of Table 6.3.

ness and 258 for leisure passengers, whereas the normal draws of Table 6.3 have means of 1,200 and 300.

Restriction disutilities. The restriction costs were tuned so that, on average, business passengers rank the fare classes from the least to the most restricted and leisure passengers the reverse (Belobaba 2006g). The lecture's example fare structure and mean parameters are reproduced in Table 6.3.

Table 6.3 The PODS lecture example (historical simulation settings, USD; adapted from Belobaba, 2006). Mean maximum willingness to pay is 1,200 for business and 300 for leisure passengers. Mean disutilities are 450 and 350 for the Saturday-night stay, 150 and 50 for non-refundability and 150 and 50 for the change fee. Total cost is fare plus the disutilities of the restrictions that apply.

Class	Fare	Restrictions	Business total cost	Leisure total cost
Y	800	none	800	800
M	400	7-day advance purchase, Saturday-night stay	850	750
B	300	14-day advance purchase, Saturday-night stay, non-refundable	900	700
Q	200	21-day advance purchase, Saturday-night stay, non-refundable, change fee	950	650

WORKED EXAMPLE 6.4 Fences at work in PODS

Take the mean passengers of Table 6.3. The average business passenger can afford every class, and her total costs rise as the fare falls, from 800 for Y to 950 for Q: she buys Y even when Q is open. The Saturday-night stay, non-refundability and change fee do their job: for her they cost more than the fare differences save. Advance purchase works differently: it makes the cheaper classes unavailable to anyone booking after the deadline, and business passengers book late. The average leisure passenger has a willingness to pay of 300, so only Q and B are affordable; Q costs her 650 in total against 700 for B, so she buys Q. If revenue management closes Q, she sells up to B. If it closes B as well, the cheapest open fare, M at 400, exceeds her willingness to pay: she looks for another path or airline, or does not travel.

Individual draws create the realistic mixture. A business passenger whose Saturday-night cost is 250 rather than 450 sees total costs of 800 for Y, 650 for M, 700 for B and 750 for Q, and buys M: a buy-down. To see the aggregate, give every willingness to pay and disutility a k-factor of 0.3 (an assumption made here for illustration, not a lecture value). With all classes open, 55 percent of business passengers buy Y, 36 percent M, 6 percent B and 2 percent Q; 87 percent of leisure passengers buy Q and 13 percent cannot afford any fare. Closing Q leaves the business mix almost unchanged but splits leisure passengers evenly between selling up to B and leaving; closing Q and B as well sells M to only 13 percent of them.

Structurally, the PODS passenger model is a random-coefficients choice model in which, given her draws, a passenger chooses deterministically, with two additions that logit lacks: a non-compensatory screen (no fare above willingness to pay, whatever its other merits) and a consideration set (the decision window, with a cost for leaving it). Choices conditional on availability are exactly what a revenue management simulator needs, because sell-up, buy-down and recapture then emerge from passenger behaviour rather than being assumed. The settings are historical: they describe the restricted fare structures of the early 2000s, which branded fares have since replaced in many markets (Chapter 17). The logic of maximum willingness to pay plus the dollar cost of product features carries over to branded fares directly. The ePODS teaching game built on

the same model fixed the fare ratios at 4, 2 and 1.5 times a Q fare equal to USD 50 plus 4 cents per mile (Belobaba 2006h).

6.6 Itinerary choice in network planning tools

A network planning system turns a schedule into passengers and revenue in five steps. It sizes each origin-destination market by day and segment, from booking data and traffic statistics, possibly with stimulation for new service (Chapter 4). A connection builder generates the itineraries that can be sold, using minimum and maximum connecting times and limits on circuity. A **share model** allocates each market's travellers to its itineraries. The resulting itinerary demands are loaded onto flight legs, where capacity spills some passengers and recapture reassigns some of them (Chapter 9). Revenues are prorated to legs and compared with costs. The share model is where the choice models of this and the previous chapter live.

The oldest share models use the **quality of service index** of Chapter 4: each itinerary gets a score built by multiplying factors for its level of service (nonstop, one-stop, connection), departure time, aircraft type and carrier, and its share is its score divided by the sum of the scores in the market. The connection to logit is exact.

RESULT 6.5 QSI shares are logit shares, and so are QSI recapture rates

If $\mathrm{QSI}_i = \prod_k a_k^{x_{ik}}$ for factors $a_k > 0$ and attributes x_{ik}, then the QSI share $s_i = \mathrm{QSI}_i / \sum_j \mathrm{QSI}_j$ is a multinomial logit probability with $V_i = \sum_k x_{ik} \ln a_k$. The recapture rate used in passenger mix and itinerary-based fleet assignment models (Barnhart 2003),

$$b_{pr} = \frac{\mathrm{QSI}_r}{\mathrm{QSI}_r + \mathrm{msc}},$$

where msc is the total QSI share of competitors' itineraries, is the logit probability of choosing r when the spilled passenger's choice set is reduced to r and the competitors' itineraries, with no outside option.

Proof. $\ln \mathrm{QSI}_i = \sum_k x_{ik} \ln a_k = V_i$, so $\mathrm{QSI}_i / \sum_j \mathrm{QSI}_j = e^{V_i} / \sum_j e^{V_j}$. Normalizing scores to sum to one does not change the ratios, so

the logit probability of r in the set $\{r\} \cup$ competitors is $\mathrm{QSI}_r / (\mathrm{QSI}_r + \sum_{\text{comp}} \mathrm{QSI}) = \mathrm{QSI}_r / (\mathrm{QSI}_r + \text{msc})$. ■

A QSI model therefore inherits IIA in full, usually with factors set by judgement rather than estimated, so they come without standard errors. Its recapture rates inherit IIA too. In the YUL–CDG choice set with both Laurentian departures, the inside scores normalized to one are 0.310 for each Laurentian flight, 0.267 for B and 0.114 for C. The QSI rule recaptures $0.310/(0.310 + 0.381) = 44.9$ percent of passengers spilled from the 18:30 onto the 21:15. The models of Example 6.1, which include an outside option, divert 34.5 percent of the 18:30's passengers to the 21:15 under logit, 61.6 percent under carrier nesting and 21.9 percent under time-of-day nesting when the 18:30 is unavailable. The recapture rates that spill models and fleet assignment take as inputs (Chapters 9 and 14) are outputs of a choice model, and they are only as good as its substitution structure.

Modern systems replace or supplement QSI with estimated logit and nested logit models (Coldren et al. 2003; Lufthansa Systems 2025c), with time-of-day preferences that can differ with the market's Delta-T and with the direction of travel (Boeing Commercial Airplane Group 1993; Koppelman, Coldren, and Parker 2008), and with constants calibrated to observed shares as in Section 5.8.

IN PRACTICE **Itinerary choice inside planning systems in 2026**

Vendor documentation shows the logit family at the core of current network planning software. Lufthansa Systems' 2025 product information for NetLine/Plan describes a market model based on nested logit that predicts passenger choices from itinerary attributes such as itinerary type, connection quality and airline preference; it lists multinomial and nested logit market models, training options on PaxIS, DDS or MIDT data, and spill and recapture alongside connection building and cost and revenue models in a fully O&D-based evaluation (Lufthansa Systems 2025c). These are marketing descriptions without technical detail, and the handbook cautions that public product information does not establish any airline's configuration, calibration or workflow (Network Planning Handbook 2026). Research points in two directions: richer GEV structures for itinerary shares (Coldren and Koppelman 2005), and machine-learning predictors trained on search and booking data (Lhéritier et al. 2019). A planning analyst's job is less to build these models than to know what the one in use assumes.

What should an analyst ask? The questions follow from the two chapters.

- **The choice set.** Which itineraries does the connection builder generate (maximum connections, connecting times, circuity, elapsed time)? Are codeshare duplicates collapsed to operated flights? Is there an outside option, how is total market size set, and can it grow with better service?
- **The model.** QSI, multinomial logit or nested logit? If nested, by carrier, time of day, level of service or several, and with which dissimilarity parameters? Are business and leisure, or points of sale, modelled separately? How are time-of-day preferences represented, and do they differ by distance, direction and day of week?
- **The estimation.** On which data and which years were the coefficients estimated? Do fares enter the model, and if so, how was their endogeneity handled (Chapter 7)? What values of time and connection penalties do the coefficients imply, and are they plausible?
- **The calibration.** Which constants are calibrated to which observed shares, how often, and are they held fixed when the schedule changes? How large are the adjustments in the markets that matter?

- **The validation.** How well does the model back-cast past schedule changes: retimings, added frequencies, competitor entries and exits?
- **The consistency.** Are the recapture rates used for spill derived from the same choice model, with the same outside option? Are competitors' schedules and fares held fixed in a scenario, and is that reasonable for the decision?

Every network planning forecast contains a theory of substitution. Find out what it is before you trust the answer to a frequency, timing or entry question.

6.7 Summary

- GEV models are generated by a function G; choice probabilities are $Y_i G_i / G$ and the logsum is $\ln G$. Logit is the case $G = \sum_j Y_j$.
- Nested logit groups close substitutes. Within a nest, utilities are divided by λ_k; the nest's inclusive value enters the upper logit with coefficient λ_k. The within-nest error correlation is $1 - \lambda_k^2$, and $0 < \lambda_k \leq 1$ keeps the model consistent with utility maximization.
- The nesting structure decides substitution. Calibrated to the same base shares, a second Laurentian frequency to Paris gains 8.2 points if the nests are carriers and 21.1 points if they are times of day.
- Mixed logit averages logit over a distribution of tastes. It captures business and leisure heterogeneity, breaks IIA through the covariance of individual probabilities, and is estimated by simulated likelihood; Halton draws cut simulation error sharply.
- A single logit fitted to a mixed population undervalues demand at high fares and misattributes the losses from a low-fare entrant.
- Boeing's decision window model restricts each traveller to the paths that fit her window, then chooses among airlines with a logit-like share equation in which path quality matters less at longer range. Frequency raises share by covering more windows.
- PODS passengers screen fares by willingness to pay and minimize fare plus restriction, path quality and re-planning costs. Its 2006 settings make the average business passenger buy Y and the average leisure passenger buy Q.
- QSI shares and QSI recapture rates are logit probabilities and inherit IIA. Recapture rates are outputs of a choice model, not independent facts.

Table 6.4 Key formulas of the chapter.

Quantity	Formula
GEV probability, logsum	$P_i = Y_i G_i / G,\ \ln G + \gamma$
Nested logit generator	$G = \sum_k \left(\sum_{j \in B_k} Y_j^{1/\lambda_k}\right)^{\lambda_k}$
Nested logit decomposition	$P_i = P_k P_{i\mid k},\ P_k \propto e^{\lambda_k I_k},\ P_{i\mid k} \propto e^{V_i/\lambda_k}$
Within-nest correlation	$1 - \lambda_k^2$
Mixed logit	$P_{ni} = \int L_{ni}(\beta) f(\beta \mid \theta)\, \mathrm{d}\beta$
Simulated probability	$\check{P}_{ni} = R^{-1} \sum_r L_{ni}(\beta^r)$
PODS willingness to pay	$\Pr(\text{WTP} \geq f) = \min\{1, e^{-\ln 2\,(f-b)/((m-1)b)}\}$
QSI recapture	$b_{pr} = \text{QSI}_r / (\text{QSI}_r + \text{msc})$

6.8 Further reading

The Boeing decision window presentation (Boeing Commercial Airplane Group 1993) and Belobaba's PODS and ePODS lectures (Belobaba 2006g, 2006h) are short and repay careful reading for their parameters and their logic; the handbook places the PODS lecture in a planner's preparation (Network Planning Handbook 2026). Train (2009, chs. 4 and 6) covers GEV and mixed logit, including overlapping nests and the simulation methods of his later chapters, and McFadden (1978) introduced the GEV family. McFadden and Train (2000) prove that mixed logit can approximate any random utility model. For airline applications, Coldren and Koppelman (2005) develop GEV itinerary share models, Garrow (2010) surveys the field, and Barnhart's schedule planning notes (Barnhart 2003) show how QSI-based recapture enters passenger mix and fleet assignment models. Talluri and van Ryzin (2004a) show how a discrete choice model of the kind developed here becomes the demand model of revenue management, the subject of Chapter 21.

CHAPTER

7

Prices Are Not Random: Endogeneity

IN THIS CHAPTER

- Why the fares in airline data move with demand, so that a regression of traffic on fares measures the pricing rule as much as the demand curve.
- The direction of the bias, derived for a log-linear demand regression and for a logit share model.
- A simulated route priced by revenue management, where the naive elasticity has the wrong sign and an instrument recovers the truth.
- Instruments that make sense for airlines, from fuel cost per seat to rivals' products, and the ways each one fails.
- Berry's inversion, the BLP contraction and the control function of Petrin and Train.
- Low-cost entry and route launches as natural experiments, read with difference-in-differences.
- What to ask before using an elasticity from a vendor model or a past study.

DRAWS ON Train, Discrete choice methods with simulation, ch. 13 (2009); handbook; Belobaba, Airline economics review (MIT 16.75J, 2006); Berry (1994); Berry, Levinsohn and Pakes (1995); Petrin and Train (2010); Berry, Carnall and Spiller (1996); Air Canada, Q2 2026 MD&A; Competition Bureau (2025)

Sooner or later someone will hand you an elasticity. A revenue analyst may tell you that leisure demand between Montreal and Vancouver has a fare elasticity of -0.4; a planning tool carries a fare coefficient inside its itinerary-choice model; a consultant's case for a new transatlantic route assumes that fares 10 percent lower stimulate 15 percent more traffic. Each number drives a decision you will own: how much traffic a new nonstop creates, how far fares must fall to fill an upgauged aircraft, how much revenue a competitor's

entry will take. And each was estimated, if it was estimated at all, from data in which the airline chose the fares.

That is the problem of this chapter. Airline fares are set by people and systems that watch demand. Revenue management closes cheap booking classes when a flight books ahead of forecast, so busy departures sell at high fares and quiet ones at low fares. Better products are priced higher. Planners add frequency where they expect growth. In the data, high fares and high traffic go together for reasons that have nothing to do with how travellers respond to price, and a regression of traffic on fares mixes the demand curve with the airline's own behaviour. The handbook is blunt about it: fares and frequency are endogenous, a historical regression can forecast well while giving misleading causal elasticities, and the first question to ask of any coefficient is how it was identified (Network Planning Handbook 2026).

This chapter turns that question into tools. It builds on the demand functions of Chapter 4 and on the logit and mixed logit models of Chapters 5 and 6, which assumed that every explanatory variable is independent of the unobserved part of utility. Here that assumption is dropped. The same issue returns in forecasting (Chapter 8), in the estimation of willingness to pay inside revenue management systems (Chapter 21), and in every route case that rests on a stimulation assumption (Chapter 12).

7.1 Why airline fares are endogenous

Start from the constant-elasticity demand function of Chapter 4, written for one market or one flight observed over many periods:

$$\ln D_t = a + \eta \ln f_t + u_t, \qquad \mathbb{E}[u_t] = 0. \tag{7.1}$$

Here D_t is the number of passengers in period t (a departure, a week or a quarter), which equals demand as long as the flights do not fill (Chapter 9 treats the full ones), f_t the average fare, η the fare elasticity and u_t the **demand shock**: everything that moves demand and is not in the equation, from a convention in Vancouver to a rival's seat sale or a slowing economy. Ordinary least squares (OLS) applied to (7.1) recovers η only if the fare is unrelated to the shock.

DEFINITION 7.1 Exogenous and endogenous regressors

In a linear model $y_i = x_i'\beta + u_i$, a regressor is **exogenous** if it is uncorrelated with the error, $\mathbb{E}[x_i u_i] = 0$, and **endogenous** if it is not. OLS is consistent for β only when every regressor is exogenous. With an endogenous regressor, more data give a more precise estimate of the wrong number.

Airline fares fail the test for three reasons, and each has a planning counterpart.

Revenue management responds to demand This is the airline-specific reason, and the strongest. A booking-limit system (Chapter 18) compares each departure's bookings with its forecast. When a departure books ahead of forecast the cheap classes close early and the average fare paid rises; when it books behind, classes stay open and a sale may follow. Continuous pricing does the same thing directly. Either way the departures with a positive u_t are sold at high fares and the quiet ones at low fares, so $\mathrm{Cov}(\ln f_t, u_t) > 0$. Williams (2022) estimates a model of this behaviour on US airline data, with fares that respond to demand shocks and to the seats remaining as departure approaches. Within a single departure there is a second layer: travellers who book late pay more and are on average less sensitive to price, so data by booking date mix the response to price with a change in who is buying.

Unobserved quality Products differ in attributes the analyst does not measure: the convenience of the departure time, reliability, the cabin, the loyalty program, lounges, the brand. A better product costs more to provide and faces stronger demand, so its fare is higher for two reasons and its traffic is higher too (Train 2009, ch. 13). A cross-section of flights or carriers then credits the extra traffic of the better product to its higher fare, as if travellers liked paying more.

Fares, frequency and capacity are chosen together Planners add frequency and larger aircraft where they expect demand to grow, and rivals enter the same markets. More frequency raises demand through schedule convenience (the S-curve of Chapter 4); more seats push revenue management to open cheaper classes. Fares, frequency, capacity and market structure are all decided with demand in view. Adding frequency to the regression as a control does not help, because frequency is endogenous for the same reason.

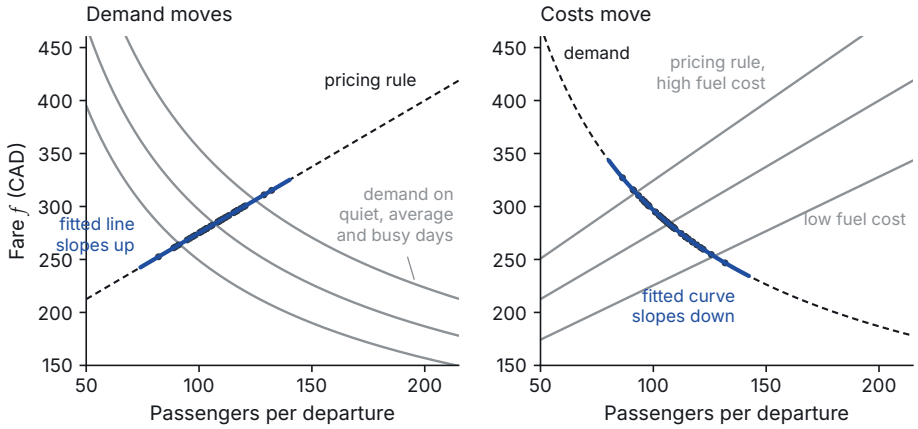


Figure 7.1 The identification problem. Left: demand shifts from day to day while the pricing rule stays fixed, so the observed combinations of fare and load (dots) lie on the pricing rule and a fitted line slopes upward. Right: cost changes shift the pricing rule while demand stays fixed, so the observed points trace the demand curve.

The difficulty is old. Working (1927) asked what statistical demand curves show, and answered with the picture redrawn in Figure 7.1. Each observation is the intersection of a demand curve and a pricing rule. If demand shifts while the pricing rule stays put, the intersections trace the pricing rule and a fitted line slopes upward. If the pricing rule shifts, because costs change, while demand stays put, the intersections trace the demand curve. Airline data contain both kinds of movement, and the share of each decides what a regression measures.

In airline data the fare is a decision taken after looking at demand. A regression of traffic on fares measures a mixture of the demand curve and the pricing rule.

IN PRACTICE Fares that watch demand

Modern pricing systems make the problem sharper. IATA defines dynamic pricing as setting prices from the shopping context, including the days to departure, competitor activity and the seats remaining, and continuous pricing as allowing an unlimited number of price points so that the price can track demand at each moment (International Air Transport Association 2021). Air Canada relies

on PROS's real-time dynamic pricing product, according to the vendor's case study (PROS 2025), and PROS describes its pricing as moving along a continuous curve between filed fares as demand and willingness to pay change (PhocusWire n.d.). In such data the fare is, by design, a function of the demand forecast. An elasticity estimated from it without an instrument or an experiment describes the pricing system as much as the travellers.

7.2 The direction of the bias

In a log-linear demand regression To see how large the problem can be, summarize the airline's pricing by a reduced-form rule,

$$\ln f_t = \gamma_0 + \gamma_1 w_t + \kappa\, u_t + v_t, \tag{7.2}$$

where w_t is a cost shifter such as the fuel price, $\kappa \geq 0$ measures how strongly revenue management passes demand shocks into fares (a demand day 10 percent stronger than usual raises the fare by 10κ percent), and v_t is pricing noise unrelated to demand. Take w_t, u_t and v_t to be mutually independent with variances σ_w^2, σ_u^2 and σ_v^2. The variance of the log fare then splits into a cost part, a demand part and noise, and the share of the demand part,

$$R = \frac{\kappa^2 \sigma_u^2}{\operatorname{Var}(\ln f_t)} = \frac{\kappa^2 \sigma_u^2}{\gamma_1^2 \sigma_w^2 + \kappa^2 \sigma_u^2 + \sigma_v^2},$$

governs the bias.

RESULT 7.1 The naive elasticity

Under (7.1) and (7.2) with $\kappa > 0$, the OLS slope of $\ln D_t$ on $\ln f_t$ converges to

$$\operatorname{plim} \hat{\eta}_{\mathrm{OLS}} = \eta + \frac{R}{\kappa} = (1 - R)\,\eta + R\left(\eta + \frac{1}{\kappa}\right). \tag{7.3}$$

The naive estimate is a weighted average of the demand elasticity η and of the slope $\eta + 1/\kappa$ of the path that demand shocks trace through the pricing rule, with weights equal to the shares of fare variance coming from costs and noise $(1 - R)$ and from demand (R). The bias is positive, so the estimate is less negative than η, and it is itself positive when $R > \kappa\,|\eta|$.

Proof. The OLS slope converges to $\operatorname{Cov}(\ln D_t, \ln f_t)/\operatorname{Var}(\ln f_t)$. Substituting (7.1), the numerator is $\eta \operatorname{Var}(\ln f_t) + \operatorname{Cov}(u_t, \ln f_t)$, and by independence $\operatorname{Cov}(u_t, \ln f_t) = \kappa\sigma_u^2$. The limit is therefore $\eta + \kappa\sigma_u^2/\operatorname{Var}(\ln f_t) = \eta + R/\kappa$, which rearranges to the weighted average. To see what $\eta + 1/\kappa$ measures, hold w_t and v_t fixed and let the shock change by $\mathrm{d}u$: the fare moves by $\mathrm{d}\ln f = \kappa\,\mathrm{d}u$ and traffic by $\mathrm{d}\ln D = (1+\eta\kappa)\,\mathrm{d}u$, a slope of $(1+\eta\kappa)/\kappa = \eta + 1/\kappa$. Finally, $\eta + R/\kappa > 0$ exactly when $R > -\eta\kappa$. ∎

The second form of (7.3) is the one to remember. Fare variation driven by costs identifies the demand curve; fare variation driven by demand identifies the pricing rule. If revenue management raises fares by less than enough to absorb a demand shock ($1 + \eta\kappa > 0$), a busy day still carries more passengers and the pricing path slopes upward. If the aircraft is full and revenue management raises the fare just enough to hold traffic at capacity ($\kappa = 1/|\eta|$), traffic does not move at all along the path and its slope is zero. A route on which most fare variation comes from revenue management therefore produces naive elasticities near zero or positive. With $\eta = -1.5$, $\kappa = 0.5$ and 80 percent of the fare variance due to demand, the naive estimate is $-1.5 + 0.8/0.5 = +0.1$.

In a logit share model The same logic applies to the share models of Chapter 5, with one extra step. Consider a market (an O&D pair in a quarter) with products $j = 1, \dots, J$, such as carrier and routing combinations, and an outside option $j = 0$: not flying, or travelling some other way. Product j gives every traveller the mean utility

$$\delta_j = x_j'\beta - \alpha f_j + \xi_j, \tag{7.4}$$

where x_j holds observed attributes (nonstop, frequency, carrier), $\alpha > 0$ is the fare coefficient and ξ_j is **unobserved quality**, the part of the product's appeal that all travellers share but the analyst does not measure. With logit errors the market shares are $s_j = e^{\delta_j}/(1+\sum_k e^{\delta_k})$ and the outside share is $s_0 = 1/(1+\sum_k e^{\delta_k})$, so $\ln s_j - \ln s_0 = \delta_j$: a linear regression of the log share ratio on attributes and fare, whose error is ξ_j (Section 7.6 develops this). Endogeneity enters through the fare.

Suppose each product is priced by a carrier that maximizes $(f_j - c_j)Ms_j$, where c_j is the marginal cost per passenger and M the market size. Since $\partial s_j/\partial f_j = -\alpha s_j(1-s_j)$, the first-order condition $s_j + (f_j - c_j)\,\partial s_j/\partial f_j = 0$ gives the logit markup rule

$$f_j = c_j + \frac{1}{\alpha\,(1-s_j)}. \tag{7.5}$$

A product with more unobserved quality has a larger share and therefore a larger markup.

RESULT 7.2 Bias in a logit share regression

(a) Under the markup rule (7.5), holding rivals' fares fixed, $\mathrm{d}f_j/\,\mathrm{d}\xi_j = s_j/\alpha > 0$. If instead capacity binds and revenue management sets the fare that just fills the aircraft, $\mathrm{d}f_j/\,\mathrm{d}\xi_j = 1/\alpha$.

(b) In the OLS regression of $\ln s_j - \ln s_0$ on x_j and f_j,

$$\operatorname{plim}\hat{\alpha}_{\mathrm{OLS}} = \alpha - \frac{\operatorname{Cov}(\tilde{f}_j, \xi_j)}{\operatorname{Var}(\tilde{f}_j)},$$

where $\tilde{f}_j$ is the fare net of its linear projection on x_j. When fares rise with unobserved quality the covariance is positive: the fare coefficient is biased toward zero, own-fare elasticities $-\alpha f_j(1 - s_j)$ are understated in magnitude, and every money value β_k/α derived from the model (the value of a nonstop, of frequency, of time) is overstated when β_k itself is estimated consistently.

Proof. For (a), write the markup condition as $g(f_j, \xi_j) = f_j - c_j - 1/(\alpha(1 - s_j)) = 0$, where s_j depends on δ_j and $\partial s_j/\partial\delta_j = s_j(1 - s_j)$. Then $\partial g/\partial f_j = 1 + s_j/(1-s_j) = 1/(1-s_j)$ and $\partial g/\partial\xi_j = -s_j/(\alpha(1-s_j))$, so the implicit function theorem gives $\mathrm{d}f_j/\,\mathrm{d}\xi_j = s_j/\alpha$. If capacity binds, the share is pinned at C_j/M, so δ_j cannot move and $\alpha\,\mathrm{d}f_j = \mathrm{d}\xi_j$. For (b), by the Frisch-Waugh-Lovell theorem the OLS fare coefficient equals the slope of the log share ratio on $\tilde{f}_j$, which converges to

$$\frac{\operatorname{Cov}(\tilde{f}_j, -\alpha f_j + \xi_j)}{\operatorname{Var}(\tilde{f}_j)} = -\alpha + \frac{\operatorname{Cov}(\tilde{f}_j, \xi_j)}{\operatorname{Var}(\tilde{f}_j)},$$

using $\operatorname{Cov}(\tilde{f}_j, x_j) = 0$ and $\operatorname{Cov}(\tilde{f}_j, f_j) = \operatorname{Var}(\tilde{f}_j)$. Because x_j is exogenous, $\operatorname{Cov}(\tilde{f}_j, \xi_j) = \operatorname{Cov}(f_j, \xi_j)$, so the bias has the sign of the covariance between fares and unobserved quality. ∎

The magnitudes are not small. With Laurentian's 26 percent share of the Montreal–Vancouver market in Example 7.3 and $\alpha = 0.006$ per dollar, one unit of unobserved quality adds s_j/α, or CAD 44, to the equilibrium fare. In the case study of new-vehicle demand that closes Train's chapter, the price coefficient in the regression of the

make-and-model constants on vehicle attributes is −0.0434 when estimated by OLS against −0.0733 with instruments, about 40 percent smaller in magnitude, exactly the direction the result predicts (Train 2009, sec. 13.6).

> **PITFALL**
>
> Controlling for the outcome. Adding the load factor, bookings to date or the number of days a flight was closed to a demand regression, in order to "control for demand", conditions on a variable that the fare itself affects. It does not remove the bias and usually makes the fare coefficient harder to interpret. Controls must be determined before the fare is set.

7.3 A route priced by revenue management

The formulas say what can happen. A simulation shows what does happen when a realistic pricing process generates the data. Consider Laurentian's daily evening departure from Montreal to Vancouver, flown by a 169-seat 737-8, over two years: 104 weeks and 728 departures. All numbers are illustrative.

Demand. At fare f a departure attracts $D_t(f) = 138\, e^{u_t} (f/249)^{-1.5}$ passengers, so the true elasticity is −1.5 and an average departure priced at CAD 249 would carry 138 passengers. The shock u_t combines a day-of-week pattern (Fridays and Sundays strong, Tuesdays and Saturdays weak) with day-to-day noise of standard deviation 0.22.

Fares. Pricing files a ladder of eight fares, each 12 percent above the one below. The lowest step, the floor, is indexed to the weekly jet fuel price, which wanders between CAD 0.52 and 1.36 per litre over the two years: a 10 percent rise in fuel raises every step of the ladder by about 3 percent.

Revenue management. For each departure the RM system forecasts u_t from bookings, with a small error (standard deviation 0.06), and opens the lowest step whose forecast demand fits into 88 percent of the seats. One fare per departure is a simplification: read f_t as the average fare paid.

The analyst sees fares, passengers and the fuel price. Figure 7.2 shows what those data look like.

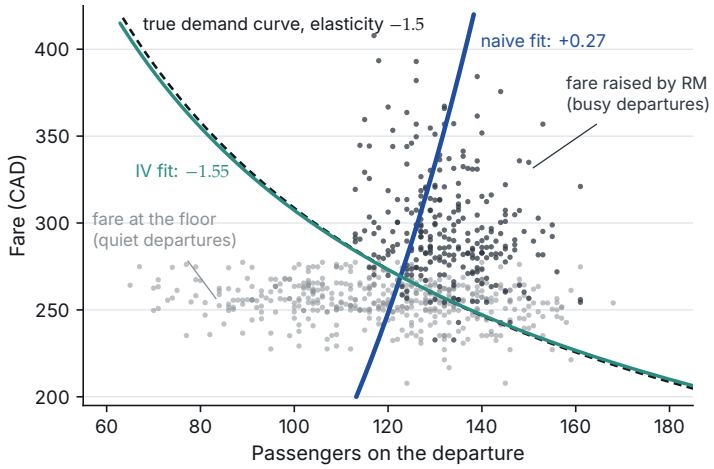


Figure 7.2 Two years of Laurentian's simulated daily YUL–YVR departure (728 departures of a 169-seat 737-8). Grey dots are departures sold at the fuel-indexed floor, dark dots departures where revenue management raised the fare. The dashed curve is the true demand curve for an average day. The naive log-log fit slopes the wrong way; the IV fit, which uses the fuel price as instrument, recovers the demand curve.

WORKED EXAMPLE 7.1 The elasticity of a revenue-managed departure

The data. On 63.5 percent of departures demand is weak enough that the floor is the fare; on the others revenue management raises it. The busy departures carry more passengers (132.8 on average against 119.2) at higher fares (CAD 297 against 255), because they are busy: their average demand shock is +0.22, against −0.13 on floor days.

The naive regression. Regressing $\ln D_t$ on $\ln f_t$ gives a slope of +0.27 (standard error 0.05, clustered by week). Taken at face value, a 10 percent fare increase would raise traffic by 2.7 percent.

The instrument. The weekly fuel price moves the floor and has nothing to do with demand for this flight. The first stage, the regression of $\ln f_t$ on the log fuel price, has slope 0.133 (standard error 0.019, F-statistic 49.7). The reduced form, the regression

of $\ln D_t$ on the log fuel price, has slope -0.206 (0.027). Their ratio is the IV estimate,

$$\hat{\eta}_{\text{IV}} = \frac{-0.206}{0.133} = -1.55,$$

with standard error 0.32 and a 95 percent confidence interval from -2.17 to -0.92 that covers the true -1.5.

Why fuel moves fares only on quiet days. Pricing passes 30 percent of a fuel change into the ladder, yet the first-stage slope is only 0.133. On busy departures revenue management offsets a higher ladder by opening a lower step: when seats are scarce, the fare is set by demand and capacity, not by cost. The instrument works through the quiet departures.

The airline's own fare history, read naively, says that travellers do not mind paying more, and every capacity or pricing scenario built on that number would be wrong. The instrument recovers the truth at a price in precision: the IV standard error is six times the OLS one.

One sample could be a fluke. Figure 7.3 repeats the exercise on 1,000 independent two-year samples. The naive slope averages $+0.23$, and 90 percent of samples fall between $+0.09$ and $+0.38$: precise, and wrong in sign. The IV estimate has a median of -1.50, with 90 percent of samples between -2.27 and -1.03. The fuel price explains under 6 percent of the variance of log fares in the sample of Example 7.1; the rest is revenue management at work, and that share, not the sample size, is what makes the naive estimate wrong. With weaker fuel indexation (a pass-through of 0.10 instead of 0.30) the median first-stage F falls to 4.8 and the IV estimates spread from -5.2 to -0.3 between the 5th and 95th percentiles: an instrument that barely moves the fare cannot pin down the response to it.

7.4 Instrumental variables and two-stage least squares

The fix in Example 7.1 used only the part of the fare variation that came from fuel. That is the general idea of instrumental variables (IV): find a variable that moves the fare for reasons unrelated to demand, and use only the variation it creates.

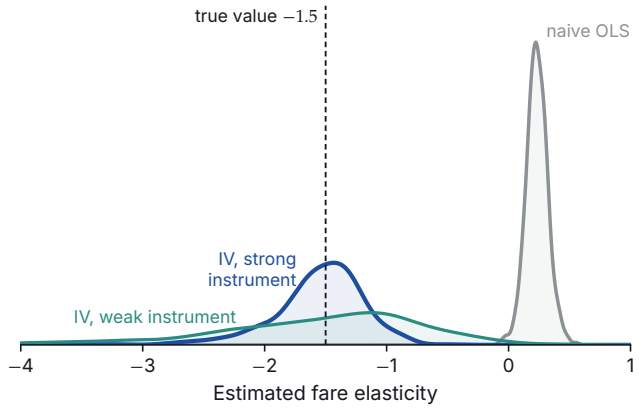


Figure 7.3 Sampling distributions over 1,000 simulated two-year samples of the route in Figure 7.2. The naive OLS slope is precise and wrong. The IV estimate with the fuel price as instrument is centred on the true elasticity, with a wider spread. With a weak instrument (median first-stage F of 4.8) the IV estimates scatter widely.

DEFINITION 7.2 Instrument

A variable z is a valid **instrument** for the fare in the demand equation (7.1) if it is

- **relevant**: $\mathrm{Cov}(z, \ln f) \neq 0$ after controlling for the exogenous regressors; and
- **excluded**: $\mathrm{Cov}(z, u) = 0$, which requires both that z is unrelated to the demand shocks and that it affects traffic only through the fare.

Relevance can be checked in the data. Exclusion cannot: it is an argument about how the market works.

Figure 7.4 draws the two conditions as a causal diagram. With one instrument the estimator follows in one line. Taking covariances of (7.1) with z gives $\mathrm{Cov}(z, \ln D) = \eta\,\mathrm{Cov}(z, \ln f) + \mathrm{Cov}(z, u)$, and exclusion sets the last term to zero:

$$\eta = \frac{\mathrm{Cov}(z, \ln D)}{\mathrm{Cov}(z, \ln f)} = \frac{\text{reduced-form slope}}{\text{first-stage slope}}. \tag{7.6}$$

The sample analogue is the IV estimator: the effect of the instrument on traffic, scaled by its effect on the fare. Relevance keeps the denominator away from zero. With several regressors, and possibly

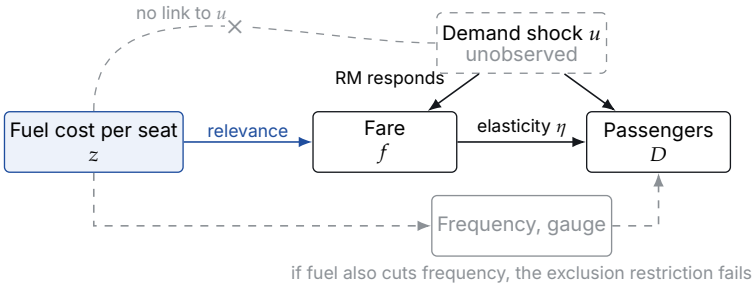


Figure 7.4 What an instrument may and may not do. Fuel cost per seat moves the fare and is unrelated to the demand shock that revenue management reacts to. If fuel prices also lead planners to cut frequency or downgauge, fuel reaches passengers through a second path and the exclusion restriction fails.

more instruments than endogenous variables, the estimator generalizes to two-stage least squares (2SLS).

RESULT 7.3 Two-stage least squares

Let $y = X\beta + u$ with n observations and k regressors, some of them endogenous, and let Z be an $n \times L$ matrix of instruments with $L \geq k$ (exogenous regressors serve as their own instruments), $\mathbb{E}[Z_i u_i] = 0$ and $\mathbb{E}[Z_i X_i']$ of rank k. The 2SLS estimator

$$\hat{\beta}_{2\mathrm{SLS}} = (X' P_Z X)^{-1} X' P_Z y, \qquad P_Z = Z(Z'Z)^{-1}Z', \tag{7.7}$$

is consistent and, under homoskedastic errors, approximately normal in large samples:

$$\hat{\beta}_{2\mathrm{SLS}} \approx \mathcal{N}(\beta,\ \sigma_u^2 (X' P_Z X)^{-1}).$$

With one regressor and one instrument it reduces to the ratio $\widehat{\mathrm{Cov}}(z,y)/\widehat{\mathrm{Cov}}(z,x)$ of (7.6).

Proof. The name describes the computation. The first stage regresses each column of X on Z and keeps the fitted values $\hat{X} = P_Z X$, the part of each regressor that the instruments explain. The second stage regresses y on $\hat{X}$: $\hat{\beta} = (\hat{X}'\hat{X})^{-1}\hat{X}'y$, which equals $(X' P_Z X)^{-1} X' P_Z y$ because P_Z is symmetric and idempotent. Equivalently, $\hat{\beta}$ minimizes $(y - X\beta)' P_Z (y - X\beta)$, the generalized method of moments criterion for the conditions $\mathbb{E}[Z_i u_i] = 0$. Substituting $y = X\beta + u$,

$$\hat{\beta} - \beta = \left(\frac{X'Z}{n} \left(\frac{Z'Z}{n}\right)^{-1} \frac{Z'X}{n} \right)^{-1} \frac{X'Z}{n} \left(\frac{Z'Z}{n}\right)^{-1} \frac{Z'u}{n}.$$

By the law of large numbers the sample moments converge to population moments, $Z'u/n \to 0$ by exogeneity, and the rank condition makes the inverse exist, so $\hat{\beta} \to \beta$. Multiplying by $\sqrt{n}$ and applying the central limit theorem to $Z'u/\sqrt{n}$, whose variance is $\sigma_u^2 \mathbb{E}[Z_i Z_i']$ under homoskedasticity, gives the stated covariance. With one regressor and one instrument (both demeaned), P_Z projects on a single vector and the formula collapses to $z'y/z'x$. ∎

Four practical points decide whether an IV estimate can be trusted.

Standard errors Run the two stages by hand and the second-stage standard errors will be wrong: the residuals must be computed with the actual regressors, $\hat{u} = y - X\hat{\beta}$, not with their first-stage predic-

tions. Use an IV routine, and cluster the standard errors by the unit within which shocks persist (the week, the route, the market).

Weak instruments When the first stage is weak, the 2SLS estimator is noisy, its distribution has fat tails, and in finite samples it is biased toward OLS. The rule of thumb asks for a first-stage F-statistic above 10 (Staiger and Stock 1997), and Angrist and Pischke (2009) give the approximation that the bias of 2SLS is roughly the OLS bias divided by $F + 1$. The weak case in Figure 7.3 shows what an F near 5 does.

Overidentification With more instruments than endogenous regressors, an overidentification test (Sargan's, or Hansen's J) checks whether the instruments agree with one another (Wooldridge 2010). It has no power against a bias that all of them share, which is the common case when every instrument is a cost shifter.

Whose elasticity If travellers differ in their response to price, IV estimates an average of the elasticities of the units whose fares the instrument moves, weighted by how much it moves them (Angrist and Pischke 2009). In Example 7.1 fuel moved fares only on quiet departures, so the estimate describes quiet departures. With a constant elasticity this does not matter; with a mix of business and leisure travellers it can.

7.5 Instruments for airline demand

No instrument is valid in general: validity is a property of an instrument in a particular data set and specification. Table 7.1 lists the families used in airline demand studies, why each moves fares and how each can fail.

Table 7.1 Instrument families for airline fares, why each is relevant and the most likely way it violates the exclusion restriction.

Instrument	Why it moves the fare	How exclusion can fail
Fuel price times distance (fuel cost per seat)	Fuel is a large cost, and more so on long sectors	Fuel cycles track the economy; fuel cuts frequency and gauge; on full flights fares follow demand, not cost
Airport charges, taxes, exchange rates	Shift the cost or the all-in price	Charges respond to traffic; currencies also move foreign demand
Rival products: number of rivals, rival nonstops, rival frequency	Closer substitutes squeeze markups	Rivals enter and add frequency where demand grows
Hausman: the carrier's fares in other markets	Shared cost shocks move fares everywhere	Shared demand shocks; markets that share a leg share its bid price
Capacity changed for operational reasons	More seats lower the fare that fills the aircraft	Planners also change gauge because demand is strong
Lagged fares	Fares are persistent	Demand shocks are persistent too

Cost shifters. Fuel is the standard airline instrument because its cost per seat scales with distance. With the illustrative burn of 0.028 litres per seat-kilometre and fuel at CAD 0.95 per litre, a seat costs about CAD 148 in fuel on YUL–SVQ and CAD 6 on YUL–YQB, so a fuel spike changes long-haul costs far more than short-haul ones. With market and period fixed effects, the instrument becomes the interaction of the fuel price with distance, and a national demand cycle that happens to coincide with oil prices is absorbed by the period effects. Two airline-specific failures deserve attention. Fuel shocks change capacity: when fuel spikes, planners cut marginal frequencies and downgauge, and frequency affects demand directly (Figure 7.4). And when capacity binds, cost barely moves the fare, as Example 7.1 showed, so the instrument's strength comes from the flights that are not full. The 2026 fuel shock shows both channels at once. In the second quarter of 2026 Air Canada paid 132.7 cents per litre of fuel, 51 percent more than a year earlier, and it reports responding by adjusting fares and fees and by making selective capacity adjustments (Air Canada 2026e): the fare response an instrument needs, and the capacity response that threatens its exclusion.

Rival products. In differentiated markets a product's markup depends on how close its substitutes are, so the characteristics of other

products shift fares. Berry, Levinsohn and Pakes built instruments from sums of the attributes of the same firm's other products and of rival products (S. T. Berry, Levinsohn, and Pakes 1995; Train 2009, ch. 13). For airlines the natural candidates are the number of rival carriers and rival nonstops, rival frequencies and whether a rival has a hub at an endpoint. They are valid only if the set of products is fixed relative to the demand shocks, which is plausible over a quarter and not over several years: carriers enter markets that are growing.

Hausman instruments. The same carrier's fares in other markets share its cost shocks but, the argument goes, not the demand shocks of this market (J. A. Hausman 1997; Train 2009, ch. 13). Airline networks undermine the argument. Markets that share a leg share its bid price (Chapter 20), so a demand shock in Montreal–Paris raises the fares of Quebec City passengers who connect to the same flight. Hausman instruments built from markets that share legs are contaminated by construction; markets far apart in the network fare better, but national shocks such as holidays, recessions and system-wide sales remain.

Capacity changes. A gauge swap made for operational reasons, an aircraft out for heavy maintenance or a delayed delivery, shifts the fare that fills a capacity-constrained flight and is plausibly unrelated to its demand, provided the analyst can separate such swaps from the ones planners make because demand is strong. A planner is well placed to know which is which.

WORKED EXAMPLE 7.2 A route-level fare elasticity from a market panel

Laurentian has 24 quarters of data on 36 O&D markets: its 15 nonstop routes from Montreal and 21 connecting markets from Quebec City, Ottawa and Boston to seven European cities over YUL. For each market m and quarter t we observe passengers, the average fare and the fuel cost per seat $z_{mt} = 0.028 \times \text{km}_m \times \text{fuel}_t$, where km_m is the flown distance and the quarterly jet fuel price ranges from CAD 0.72 to 1.33 per litre. The simulated truth has an elasticity of -1.2, persistent demand shocks, revenue management that raises the non-fuel part of the fare by 0.6 percent for each 1 percent demand shock, and 80 percent of

fuel cost per passenger passed into fares. The demand equation has market and quarter fixed effects:

$$\ln D_{mt} = a_m + b_t + \eta \ln f_{mt} + u_{mt}.$$

Three estimates. Table 7.2 reports them with standard errors clustered by market. Pooled OLS without fixed effects gives -1.56, a plausible-looking number obtained by comparing Seville with Quebec City: long-haul markets have higher fares and fewer travellers for many reasons besides the fare, and nothing guarantees that the comparison isolates the fare's effect. Two-way fixed effects OLS removes those differences and gives $+0.28$, because within a market the quarters with high fares are the quarters with high demand. 2SLS with fuel cost per seat as the instrument gives -1.13.

The pieces. The first stage says that CAD 100 more fuel cost per seat raises the log average fare by 0.138 (standard error 0.028, F-statistic 24.8), about 15 percent. The reduced form says it lowers log traffic by 0.156 (0.027), about 14 percent. Their ratio is $-0.156/0.138 = -1.13$, with a 95 percent confidence interval from -1.69 to -0.56 around the true -1.2.

The fixed effects did not solve the problem. They removed the comparisons across markets and left exactly the variation that revenue management creates. The instrument is what identifies the elasticity, and its interval is wide: the honest summary is "between -0.6 and -1.7", not "-1.13".

Table 7.2 Fare elasticity from the simulated panel of 36 Laurentian O&D markets over 24 quarters. The fixed-effects estimates include market and quarter effects; standard errors are clustered by market.

Estimator	Elasticity	Std. error	Identified by
Pooled OLS	−1.56	0.17	Differences across markets
Fixed effects OLS	+0.28	0.08	Changes within markets
Fixed effects 2SLS	−1.13	0.29	Fuel cost per seat (F = 24.8)
True value	−1.20		

7.6 Logit demand with endogenous fares: Berry and BLP

Choice models are nonlinear, so IV cannot be applied to them directly. S. T. Berry (1994) showed how to move the endogeneity out of the nonlinear choice model into a linear equation where IV works. The data are aggregate: for each market m (an O&D pair in a quarter) and product j (a carrier and routing, possibly a cabin) we observe passengers D_{jm} and fares. Market shares are $s_{jm} = D_{jm}/M_m$, where the **market size** M_m counts potential travellers and the outside share is $s_{0m} = 1 - \sum_j s_{jm}$. Market size is an assumption, often proportional to the endpoints' populations or to total O&D traffic, and it matters: it sets the outside share.

RESULT 7.4 Berry's inversion

With logit errors and an outside option, mean utilities are recovered from observed shares:

$$\ln s_{jm} - \ln s_{0m} = \delta_{jm} = x'_{jm}\beta - \alpha f_{jm} + \xi_{jm}. \tag{7.8}$$

In a nested logit with log-sum parameter λ for the nests of inside products and the outside option alone in its nest,

$$\ln s_{jm} - \ln s_{0m} = x'_{jm}\beta - \alpha f_{jm} + (1-\lambda)\ln s_{j|g,m} + \xi_{jm},$$

where $s_{j|g,m}$ is the share of product j within its nest g. Both equations are linear in the parameters with error ξ_{jm} and are estimated by 2SLS, with instruments for the fare and, in the nested case, for the within-nest share.

Proof. For logit, $s_{jm}/s_{0m} = e^{\delta_{jm}}$; take logarithms. For the nested logit of Chapter 6, let $I_g = \ln \sum_{k \in g} e^{\delta_k/\lambda}$ be the inclusive value of nest

g. Then $s_j = e^{\delta_j/\lambda + (\lambda-1)I_g}\, s_0$ and $s_{j|g} = e^{\delta_j/\lambda - I_g}$, so $\ln s_j - \ln s_0 = \delta_j/\lambda + (\lambda-1)I_g$. Substituting $I_g = \delta_j/\lambda - \ln s_{j|g}$ gives $\delta_j + (1-\lambda)\ln s_{j|g}$. ■

WORKED EXAMPLE 7.3 **Inverting shares on Montreal–Vancouver**

Laurentian's simulated data cover 60 O&D markets over 8 quarters, with two to four products per market (Laurentian or a rival, nonstop or one-stop): 1,336 product-quarters. Travellers value a nonstop ($\beta = 1.0$), frequency (0.45 per unit of log weekly frequency) and Laurentian's brand at its Montreal hub (0.3), and dislike fares ($\alpha = 0.006$ per dollar). Unobserved quality has a persistent part and a quarterly part. Carriers set fares by the markup rule (7.5), with costs that rise with distance, fuel and, slightly, quality.

Inversion. In the first quarter on YUL–YVR, Laurentian's nonstop (28 weekly flights, CAD 440) has 26.2 percent of the market, the rival nonstop (21 weekly, CAD 395) 16.0 percent and a rival one-stop over Toronto (CAD 384) 6.3 percent; 51.6 percent of potential travellers choose none of them. By (7.8) the mean utilities are

$$\delta_{\mathrm{L}} = \ln(0.2617/0.5157) = -0.678,$$
$$\delta_{\mathrm{R}} = \ln(0.1601/0.5157) = -1.170,$$
$$\delta_{1\mathrm{S}} = \ln(0.0625/0.5157) = -2.110.$$

Estimation. Pooling all markets, regress the log share ratio on the nonstop dummy, log frequency, the Laurentian dummy and the fare, with market and quarter fixed effects that absorb market size and each market's overall pull. The instrument for the fare is the fuel price times each product's flown distance, a cost shifter that differs between nonstops and the longer one-stop routings. Table 7.3 compares the estimates. OLS finds a fare coefficient of 0.0031 per dollar, about half the truth; 2SLS finds 0.0059 (first-stage F = 186).

Consequences. With the 2SLS coefficient, Laurentian's own-fare elasticity on YUL–YVR is $-0.00593 \times 440 \times 0.738 = -1.93$; with OLS it is -1.00. A 10 percent fare increase loses 19 percent of Laurentian's passengers, not 10 percent. The value of a nonstop, $\hat{\beta}_{\mathrm{ns}}/\hat{\alpha}$, is CAD 168 with 2SLS (delta-method standard error 37; truth 167) and CAD 337 with OLS: the naive model

would tell a planner that travellers pay twice as much for a nonstop as they do.

The bias does not stay in the fare coefficient. It flows into every elasticity and every dollar value of a schedule attribute that the model produces.

Table 7.3 Logit demand estimated from the inverted shares of 1,336 simulated product-quarters, with market and quarter fixed effects. Standard errors, clustered by market, in parentheses.

Parameter	True	OLS	2SLS
Nonstop	1.00	1.04 (0.05)	0.99 (0.06)
Log weekly frequency	0.45	0.45 (0.08)	0.48 (0.09)
Laurentian	0.30	0.27 (0.05)	0.28 (0.06)
Fare, per CAD 1,000	−6.0	−3.1 (0.5)	−5.9 (1.2)
Value of a nonstop (CAD)	167	337 (63)	168 (37)
Own elasticity, Laurentian YUL–YVR	−1.95	−1.00	−1.93

Plain logit also dictates where lost passengers go. If Laurentian raises its YUL–YVR fare, the travellers it loses spread over the alternatives in proportion to their shares: 21.7 percent to the rival nonstop, 8.5 percent to the one-stop and 69.8 percent out of the market. That is the independence of irrelevant alternatives of Chapter 5, and the main reason to let tastes vary.

Random coefficients. S. T. Berry, Levinsohn, and Pakes (1995) let tastes vary across travellers. With a random fare coefficient, for example, traveller n has $\alpha_n = \alpha + \sigma_\alpha \nu_n$ with ν_n standard normal, and utility $U_{nj} = \delta_j - \sigma_\alpha \nu_n f_j + \varepsilon_{nj}$ with δ_j as in (7.4). Market shares become the mixed logit of Chapter 6,

$$\hat{s}_j(\delta;\theta) = \int \frac{e^{\delta_j - \sigma_\alpha \nu f_j}}{1 + \sum_k e^{\delta_k - \sigma_\alpha \nu f_k}} \, \varphi(\nu)\, \mathrm{d}\nu, \tag{7.9}$$

computed by simulation over a fixed set of draws of ν, where θ collects the parameters of taste dispersion (σ_α here; more generally one per random coefficient). The log share ratio no longer equals δ_j, but S. T. Berry (1994) showed that for any θ exactly one vector δ makes predicted shares equal observed shares, and S. T. Berry, Levinsohn, and Pakes (1995) gave the iteration that finds it.

RESULT 7.5 The BLP contraction

For given θ and observed shares s, define the map

$$T(\delta) = \delta + \ln s - \ln \hat{s}(\delta; \theta).$$

Its Jacobian has nonnegative entries and row sums

$$\sum_k \frac{\partial T_j}{\partial \delta_k} = 1 - \frac{\mathbb{E}_n[P_{nj} P_{n0}]}{\hat{s}_j} < 1,$$

where P_{nj} is traveller n's logit probability of choosing j and P_{n0} of choosing the outside option. On any bounded convex set of δ, T is therefore a contraction in the maximum norm, with modulus $\rho < 1$ equal to the largest row sum on that set. It has a unique fixed point $\delta^\star(\theta)$, at which $\hat{s}(\delta^\star; \theta) = s$, and the iteration $\delta^{t+1} = T(\delta^t)$ converges to it from any starting point, with $\max_j \left|\delta_j^t - \delta_j^\star\right| \le \rho^t \max_j \left|\delta_j^0 - \delta_j^\star\right|$.

Proof. Since $\hat{s}_j = \mathbb{E}_n[P_{nj}]$ and $\partial P_{nj}/\partial \delta_k = P_{nj}(\mathbb{1}\{j = k\} - P_{nk})$,

$$\frac{\partial \ln \hat{s}_j}{\partial \delta_k} = \frac{\mathbb{E}_n[P_{nj}(\mathbb{1}\{j = k\} - P_{nk})]}{\hat{s}_j}.$$

Hence $\partial T_j/\partial \delta_k = \mathbb{E}_n[P_{nj} P_{nk}]/\hat{s}_j \ge 0$ for $k \ne j$, and $\partial T_j/\partial \delta_j = 1 - \mathbb{E}_n[P_{nj} - P_{nj}^2]/\hat{s}_j$, which equals $\mathbb{E}_n[P_{nj}^2]/\hat{s}_j \ge 0$. Summing over k and using $\sum_k P_{nk} = 1 - P_{n0}$ gives the row sum $\mathbb{E}_n[P_{nj}(1 - P_{n0})]/\hat{s}_j$, which is below one because every traveller has a positive probability of choosing the outside option. By the mean value theorem, with the derivatives evaluated at a point between δ and δ',

$$\left|T_j(\delta) - T_j(\delta')\right| \le \sum_k \frac{\partial T_j}{\partial \delta_k} \left|\delta_k - \delta_k'\right| \le \rho \max_k \left|\delta_k - \delta_k'\right|,$$

so T is a contraction, and the contraction mapping theorem gives existence, uniqueness and the geometric rate. A fixed point satisfies $\ln s = \ln \hat{s}(\delta^\star; \theta)$. S. T. Berry, Levinsohn, and Pakes (1995) show that the iteration can be confined to a bounded set. ∎

For plain logit ($\sigma_\alpha = 0$) every traveller has the same $P_{n0} = \hat{s}_0$, so near the fixed point the modulus is simply $1 - s_0$: the larger the outside share, the faster the iteration. Figure 7.5 runs it on the YUL–YVR shares of Example 7.3, with $\sigma_\alpha = 0.002$ and 1,000 simulated travellers, starting from $\delta = 0$. The step size falls geometrically, and

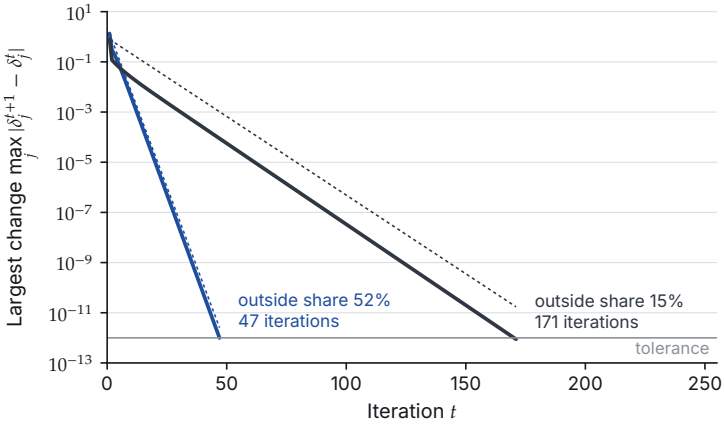


Figure 7.5 The BLP contraction on the YUL–YVR shares of Example 7.3, with a random fare coefficient (standard deviation 0.002, 1,000 simulated travellers). The step size falls geometrically; the dotted lines have the slopes implied by the largest Jacobian row sums (0.56 and 0.87). A narrower market definition, with a 15 percent outside share, needs almost four times as many iterations.

the ratio of successive steps settles at 0.55, just under the largest row sum of the result at the fixed point (0.56), which bounds it. With the market defined more narrowly, so that the outside share is 15 percent, the rate is 0.86 (bound 0.87) and the iteration needs 171 steps instead of 47 to reach a tolerance of 10^{-12}. The fixed point, $\delta^\star = (-0.745, -1.200, -2.133)$, differs from the plain-logit inversion because the random coefficient changes who buys what.

The full BLP estimator nests this inner loop inside a search over θ (S. T. Berry, Levinsohn, and Pakes 1995; Train 2009, ch. 13):

1. Draw the simulated travellers once and keep them fixed.
2. For a trial θ, run the contraction in every market to obtain $\delta(\theta)$.
3. Regress $\delta(\theta)$ on the attributes and the fare by 2SLS. This concentrates out β and α and leaves the residuals $\xi(\theta)$, the unobserved qualities implied by θ.
4. With instruments Z and a weight matrix W, evaluate the generalized method of moments objective

$$G(\theta) = \xi(\theta)' Z W Z' \xi(\theta).$$

5. Search over θ to minimize the objective, repeating steps 2 to 4.

Three practical lessons come with the method. The inner loop must be solved tightly, because a loose tolerance in the contraction feeds numerical error into the outer search; Dubé, Fox, and Su (2012) show how badly this can go and propose a constrained-optimization

alternative. The dispersion parameters θ govern substitution, so they need instruments that change the choice set, such as the number and characteristics of rival products. And the current reference implementation is the open-source PyBLP package of Conlon and Gortmaker (2020), which encodes these best practices. For airlines, S. T. Berry, Carnall, and Spiller (1996) applied the framework to US markets with two types of traveller who differ in their sensitivity to fares, to separate the cost advantages of hub operations from the higher markups that hub carriers earn; S. Berry and Jia (2010) used a related model to explain the industry's losses in the 2000s. Their estimates are historical, as the handbook warns (Network Planning Handbook 2026); the method is current.

7.7 Control functions

The BLP approach absorbs unobserved quality into a constant for each product in each market. Two situations call for something else: shares of zero, which leave the logarithm in (7.8) undefined, and fares that vary across individual travellers within a market, as they do in booking data where each traveller faces the fare open on the day of purchase (Train 2009, sec. 13.4). The control-function approach handles both. Model the fare as a function of instruments plus an error; the part of the utility error that is correlated with the fare is a function of that error; estimate the error in a first stage and add it to the choice model, where it "controls" for the endogeneity (Rivers and Vuong 1988; Petrin and Train 2010).

RESULT 7.6 Control functions

(a) Linear model. Let $y = \beta x + w'\beta_w + u$ with first stage $x = \pi' z + \pi_w' w + \mu$, where w are exogenous controls and z instruments. The OLS coefficient on x in the regression of y on x, w and the first-stage residual $\hat{\mu}$ equals the 2SLS estimate, and the t-statistic on $\hat{\mu}$ tests whether x is exogenous.

(b) Logit (Petrin and Train). Let utility be $U_{nj} = V(f_{nj}, x_{nj}; \beta_n) + \vartheta_{nj} + \varepsilon_{nj}$, with fare equation $f_{nj} = h(z_{nj}; \gamma) + \mu_{nj}$, $(\vartheta_{nj}, \mu_{nj})$ jointly normal and independent of the instruments, and ε_{nj} independent extreme value. Then $\mathbb{E}[\vartheta_{nj} \mid \mu_{nj}] = \omega\, \mu_{nj}$ and

$$U_{nj} = V(f_{nj}, x_{nj}; \beta_n) + \omega\, \mu_{nj} + \sigma_\zeta \zeta_{nj} + \varepsilon_{nj},$$

with ζ_{nj} standard normal and independent of the fare. Replacing μ_{nj} by the first-stage residual gives a mixed logit in which the fare is no longer endogenous.

Proof. For (a), partial out w from every variable (Frisch-Waugh-Lovell). The first stage splits x into a fitted part $\hat{x} = \hat{\pi}' z$ and the residual $\hat{\mu}$, which are orthogonal by construction. The coefficient on x in the regression on x and $\hat{\mu}$ equals the slope of y on the part of x orthogonal to $\hat{\mu}$, which is $\hat{x}$; regressing y on $\hat{x}$ is the second stage of 2SLS. If x is exogenous the residual has no explanatory power for u, which is the logic of Hausman's test (J. A. Hausman 1978). For (b), recall that for jointly normal variables the conditional mean is linear, $\mathbb{E}[\vartheta \mid \mu] = \omega\mu$ with $\omega = \mathrm{Cov}(\vartheta, \mu)/\mathrm{Var}(\mu)$, and the deviation $\vartheta - \omega\mu$ is normal, with constant variance σ_ζ^2 and independent of μ, hence of the fare once the instruments are conditioned on. Substituting gives the stated utility. ∎

On the route of Example 7.1, adding the first-stage residual to the naive regression gives a fare coefficient of −1.545, identical to 2SLS, and a residual coefficient of +1.93 with a naive t-statistic of 9.3: exogeneity is rejected decisively. The positive sign is the fingerprint of revenue management: departures whose fares were unexpectedly high for their fuel price also carried unexpectedly many passengers. The second-step standard error of the fare coefficient (0.21) is too small because it treats the residual as data; the correct value (0.32) is the 2SLS one. In nonlinear models, bootstrap the two steps together.

The price of the method is a set of assumptions about how fares are set. The fare equation needs an additive error that captures all of the fare's correlation with the utility error, the instruments must be

independent of both errors, and the conditional distribution of the utility error given the fare error must be specified correctly. The first assumption is the delicate one for airlines. Under markup pricing or revenue management the fare depends on demand through the elasticity or the seats remaining, not only through an additive cost error, so the first-stage residual mixes cost and demand shocks nonlinearly and a normal conditional distribution can only be an approximation (Train 2009, sec. 13.4.1). In exchange, the control function needs no contraction, works with individual booking data and zero shares, and runs in any logit or mixed logit software.

7.8 Natural experiments and difference-in-differences

Instruments are one way to find fare variation unrelated to demand. Natural experiments are another: events that change fares or market structure on some routes and not on others, for reasons that have nothing to do with those routes' demand shocks. For a network planner the common ones are a low-cost carrier's entry, a new nonstop, a change in airport charges or taxes, and a slot reallocation. The workhorse estimator is difference-in-differences (Card and Krueger 1994; Angrist and Pischke 2009).

DEFINITION 7.3 Difference-in-differences

Observe an outcome y, such as the log fare or log passengers, on a treated group of routes T (where the event happened) and a control group C, before and after the event. The **difference-in-differences** estimate is

$$\hat{\tau} = (\bar{y}_{\mathrm{T,post}} - \bar{y}_{\mathrm{T,pre}}) - (\bar{y}_{\mathrm{C,post}} - \bar{y}_{\mathrm{C,pre}}).$$

It estimates the average effect of the event on the treated routes under **parallel trends**: without the event, the outcome on treated routes would have changed by the same amount as on control routes.

The logic is one line. Write $\Delta y(0)$ for the change a route would have experienced without the event. The treated routes' change is the effect plus their own $\Delta y(0)$, so

$$\mathbb{E}[\hat{\tau}] = \tau_{\mathrm{ATT}} + \big(\mathbb{E}[\Delta y(0) \mid \mathrm{T}] - \mathbb{E}[\Delta y(0) \mid \mathrm{C}]\big),$$

where τ_{ATT} is the average effect on the treated, and the bracket vanishes under parallel trends. With many routes and periods the same comparison is a regression with route and period fixed effects,

$$y_{rt} = a_r + b_t + \tau d_{rt} + e_{rt}, \tag{7.10}$$

where $d_{rt} = 1$ for treated routes after the event. Replacing τd_{rt} by a separate coefficient τ_k for each quarter k relative to the event (omitting $k = -1$) gives an **event study**: the coefficients before the event test parallel trends, those after it trace the effect over time.

WORKED EXAMPLE 7.4 A low-cost carrier enters ten routes

A low-cost carrier starts flying 10 of the 30 domestic and transborder markets that Laurentian serves, nonstop or over Montreal, in the ninth of 16 quarters. The simulation picked the ten at random, so parallel trends hold by construction; in real data that is the claim to defend. Table 7.4 shows the two-by-two computation for Laurentian's log average fare. On the entry routes the log fare fell by 0.079 from before to after; on the other routes it rose by 0.044 with the trend. The difference-in-differences is -0.123, a fare reduction of 11.6 percent, and the regression (7.10) gives the same -0.123 (standard error 0.011).

Figure 7.6 shows the dynamics. Before entry the coefficients for the fare stay within 0.03 of zero, which supports parallel trends. Laurentian matches progressively: its log fare falls by about 0.06 in the quarter of entry and by 0.13 on average from the second quarter on (true value 0.14). Its passengers fall by about 5 percent (difference-in-differences −0.049, standard error 0.012) while total market passengers rise by 25 percent (0.224, standard error 0.017).

What not to compute. Dividing Laurentian's passenger effect by its fare effect gives $-0.049/(-0.123) = +0.40$, an upward-sloping demand curve. The ratio is a Wald estimator that uses entry as an instrument for Laurentian's fare, and entry fails the exclusion restriction: it adds a competing product, which takes Laurentian passengers at any fare. Dividing the market's growth by Laurentian's fare change (−1.82) fails for the same reason, since the entrant's own seats and fares stimulate the market. The natural experiment identifies the effect of entry, which is what a planner needs for a competitive scenario. It does not identify a fare elasticity.

Table 7.4 Laurentian's average log fare before and after low-cost entry, simulated; fares in parentheses are geometric means.

Routes	Before entry	After entry	Change
Entry routes (10)	5.805 (CAD 332)	5.726 (CAD 307)	−0.079
Other routes (20)	5.705 (CAD 300)	5.749 (CAD 314)	+0.044
Difference-in-differences			−0.123

Four threats deserve a planner's attention.

Anticipation. Goolsbee and Syverson (2008) found that incumbents cut fares when Southwest established a presence at both endpoints of a route, before it began flying the route itself. If Laurentian cuts its fares two quarters before entry (by 3 and then 6 percent in the simulation), the quarter before entry is no longer a clean baseline: the event-study effect from the second quarter on shrinks from −0.13 to −0.07, and the simple difference-in-differences to −0.112. Date the baseline before any announcement or presence at the endpoints.

Selection. Entrants and new routes go where growth is expected, which violates parallel trends by design. Check the leads, choose controls that resemble the treated routes before the event, and for a single route consider a synthetic control, a weighted combination

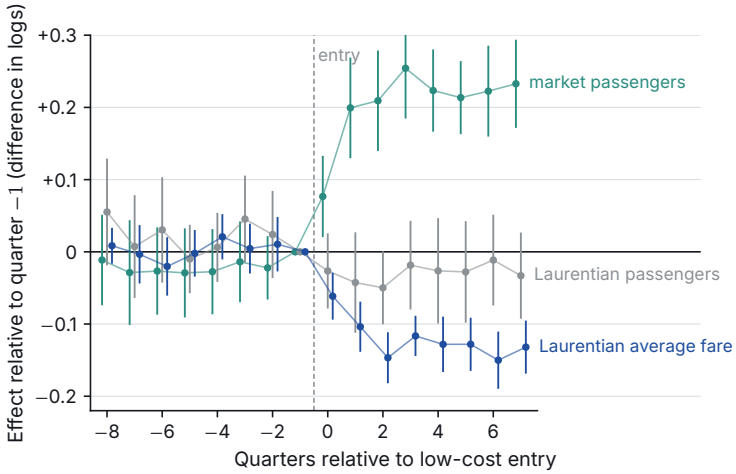


Figure 7.6 Event study of low-cost entry on 10 of 30 simulated Laurentian markets. Dots are effects relative to the quarter before entry, as differences in logs, with 95 percent confidence intervals from standard errors clustered by route. Before entry the entry and other routes move together; after entry Laurentian's fare falls, its passengers fall slightly and the market grows.

of untreated markets that reproduces the treated market's history (Abadie, Diamond, and Hainmueller 2010).

Spillovers. A new YUL–SVQ nonstop diverts passengers who used to connect over Madrid, so YUL–MAD is not a clean control: the control group must not be affected by the event.

Staggered timing. When routes are treated at different dates and the effect grows over time, the two-way fixed effects regression (7.10) implicitly compares late-treated routes with early-treated ones and can even get the sign wrong; estimators that compare each cohort only with routes not yet, or never, treated avoid the problem (Goodman-Bacon 2021; Sun and Abraham 2021).

IN PRACTICE **Natural experiments and their data in 2026**

Canadian markets have produced natural experiments of exactly this kind. The Competition Bureau's 2025 market study reports that Porter's jet expansion and Flair's entry put more competition on major routes, with the share of Air Canada's major routes facing both WestJet and at least one smaller carrier rising from 11 to 16 percent since 2019, and that one new competitor on a route between two cities lowers fares by 9 percent on average (Competi-

tion Bureau Canada 2025). The fare effect in Example 7.4 is of the same order, but it is illustrative. Studying them needs route-level traffic and fare data, and the public sources have changed. The US Department of Transportation's quarterly 10 percent ticket sample (DB1B), behind many academic studies of US airline demand, ended in June 2025; from July 2025 the US origin and destination survey is monthly with a 40 percent sample (DB1C/OD40), a break that rules out splicing old and new series mechanically (Network Planning Handbook 2026). Statistics Canada's domestic origin and destination survey is listed as inactive, with releases suspended for confidentiality reasons, and commercial O&D data built from MIDT and other inputs are adjusted estimates rather than a census (Network Planning Handbook 2026). Before reading a change in traffic as a response to fares or entry, check that it is not a change in what the data cover.

7.9 Before you use someone else's elasticity

Most elasticities a planner uses were estimated by someone else: a revenue management system, a vendor's itinerary-choice model, a consultant, an academic study, an internal analysis from a few years ago. The handbook's advice is to ask how a coefficient was identified and whether a proposed scenario stays within the data's support (Network Planning Handbook 2026). Made specific, that means eight checks.

1. **Find out what moved the fares.** If the answer is the airline's own pricing and revenue management, the estimate is biased toward zero unless an instrument or an experiment was used. Ask which, and ask for the first-stage F.
2. **Pin down which elasticity it is.** A market elasticity (all carriers' traffic when all fares move) differs from a carrier elasticity (one airline's traffic when only its fares move), and both differ from the response between fare products (Chapter 21). In the logit model of Example 7.3, a uniform fare change across all YUL–YVR products has elasticity $-\alpha\, s_0 \bar{f} = -1.29$, where $\bar{f}$ is the share-weighted average fare (CAD 418), while Laurentian's own elasticity is -1.95.
3. **Check what was held constant.** A coefficient estimated with frequency, capacity and rivals' fares held fixed does not describe a scenario in which they change too, such as a fare cut that rivals match.

4. **Check the data's support.** A fare cut of 40 percent on a route whose fares never moved by more than 10 percent is an extrapolation, whatever the model.
5. **Ask how demand was measured.** Loads on full flights and bookings in closed classes are censored (Chapter 9), and a regression on them understates the response at high fares.
6. **Ask how the market was defined.** In a logit model the market elasticity is proportional to the outside share, which follows from the assumed market size. With the same fares and inside shares rescaled to a 15 percent outside share, the YUL–YVR market elasticity falls from −1.29 to −0.38.
7. **Ask when, and under which fare structure.** Elasticities estimated before branded fares, before the pandemic or under another carrier's fare rules describe a different market (Chapter 17).
8. **For vendor and RM models, ask whether the parameter was estimated on your markets, calibrated to reproduce shares, or set by default.** Revenue management methods such as fare adjustment need an estimate of price sensitivity (Fiig et al. 2010); when that estimate comes from the airline's own booking history, it inherits the problem of this chapter (Chapter 21).

Compilations of published estimates, such as the review prepared by Gillen, Morrison, and Stewart (2004) for the Department of Finance Canada and the InterVISTAS report for IATA (2007), are useful as a range and as a check on orders of magnitude. The MIT lecture notes cite empirical market elasticities of −0.8 to −2.0, with business demand less elastic than leisure (Belobaba 2006d); these are historical values that teach the shape of the problem, not parameters for a 2026 route.

> **PITFALL**
>
> One elasticity for every market. The handbook warns against importing a universal fare elasticity or frequency-share curve into every market (Network Planning Handbook 2026). The mix of business and leisure travel, distance, competition and the outside option all differ from one market to the next, and so do the elasticities.

An elasticity is a property of a market and of the variation that identified it. Ask for both before you use it.

7.10 Summary

- Airline fares are endogenous. Revenue management raises fares when demand is strong, better products cost more and sell more, and fares, frequency and capacity are all chosen with demand in view.
- The OLS slope of log traffic on log fare converges to $\eta + R/\kappa$, a weighted average of the demand elasticity and the slope of the pricing path. On revenue-managed routes the naive elasticity is near zero or positive.
- In logit share models fares rise with unobserved quality (by s_j/α under logit markups), so the fare coefficient is biased toward zero and every money value β_k/α is overstated.
- An instrument must move the fare (relevance, checked with the first-stage F) and reach traffic only through the fare (exclusion, argued from how the market works). IV trades bias for variance.
- Airline instruments include fuel cost per seat, rival products, Hausman prices and operational capacity changes. Each fails in airline-specific ways: capacity responses to fuel, entry into growing markets, demand shocks shared through legs and bid prices.
- Berry's inversion turns logit demand into a linear IV regression, $\ln s_j - \ln s_0 = \delta_j$. For random coefficients, the BLP contraction recovers δ at a rate set by the outside share.
- Control functions add the first-stage residual to the model. In linear models this reproduces 2SLS; in choice models it requires assumptions about how fares are set.
- Natural experiments such as low-cost entry identify the effect of the event. They identify a fare elasticity only if the event affects demand through fares alone, which entry does not.
- Before using an elasticity, ask what moved the fares, which elasticity it is, what was held constant and how the market was defined.

Table 7.5 Key formulas of the chapter.

Quantity	Formula
Naive OLS limit	$\operatorname{plim} \hat{\eta}_{\mathrm{OLS}} = \eta + R/\kappa$
IV with one instrument	$\hat{\eta} = \widehat{\operatorname{Cov}}(z, \ln D)/\widehat{\operatorname{Cov}}(z, \ln f)$
Two-stage least squares	$\hat{\beta} = (X' P_Z X)^{-1} X' P_Z y$
Logit markup	$f_j = c_j + 1/(\alpha(1 - s_j))$, $\mathrm{d}f_j/\,\mathrm{d}\xi_j = s_j/\alpha$
Berry inversion	$\ln s_j - \ln s_0 = x_j'\beta - \alpha f_j + \xi_j$
Logit elasticities	own $-\alpha f_j(1 - s_j)$; market $-\alpha s_0 \bar{f}$
BLP contraction	$\delta^{t+1} = \delta^t + \ln s - \ln \hat{s}(\delta^t; \theta)$
Difference-in-differences	$\hat{\tau} = (\bar{y}_{\mathrm{T,post}} - \bar{y}_{\mathrm{T,pre}}) - (\bar{y}_{\mathrm{C,post}} - \bar{y}_{\mathrm{C,pre}})$

7.11 Further reading

Train's chapter on endogeneity (2009, ch. 13) is the library's treatment and the source of this chapter's account of BLP and control functions; read it with the handbook's warning about identification (Network Planning Handbook 2026). S. T. Berry (1994) and S. T. Berry, Levinsohn, and Pakes (1995) are the founding papers, Nevo (2000) is a practitioner's guide, and Conlon and Gortmaker (2020) set out current best practice together with the PyBLP software, with Dubé, Fox, and Su (2012) on numerical accuracy. Petrin and Train (2010) develop the control function for choice models. For airlines, S. T. Berry, Carnall, and Spiller (1996) and S. Berry and Jia (2010) estimate random-coefficients demand on US data, Williams (2022) models fares that respond to demand shocks, and Goolsbee and Syverson (2008) is a model natural experiment. Angrist and Pischke (2009) and Wooldridge (2010) cover IV and difference-in-differences, and Goodman-Bacon (2021) and Sun and Abraham (2021) explain the pitfalls of staggered adoption for difference-in-differences and event studies.

CHAPTER

8

Forecasting Demand

IN THIS CHAPTER

- Which decision needs a forecast of which quantity, at which level of detail and how far ahead.
- Benchmarks first: the naive and seasonal naive forecasts every model must beat.
- Exponential smoothing, ARIMA and dynamic regression with calendar effects, written out in full.
- Honest evaluation: rolling origins, scaled errors, bias, and errors weighted by what they cost.
- Forecasts that add up across markets: bottom-up, top-down and minimum-trace reconciliation.
- Booking curves and pickup forecasting, and when additive or multiplicative pickup is right.
- Markets without history: gravity models, analogues and stimulation; prediction intervals and scenarios for decision makers.

DRAWS ON Hyndman and Athanasopoulos, Forecasting: Principles and Practice (2021); Belobaba, Introduction to PODS passenger choice model (MIT 16.75J, 2006); Belobaba, Introduction to revenue management (MIT 16.75J, 2006); Belobaba, O&D control: what have we learned? (MIT 16.75J, 2006); handbook

Every number in a network plan rests on a forecast. The route proposal for Seville needs a forecast of the market and of Laurentian's share of it, two years ahead. The winter schedule needs monthly demand by direction and day of week for each route, a season ahead. Fleet assignment needs the mean and variability of demand for each departure. Revenue management needs, every night, a forecast of the bookings still to come for every flight and fare class. These are different forecasts, made at different levels of detail and horizons, and a planner who confuses them will build careful analysis on the wrong quantity.

This chapter is about prediction, not explanation. Chapter 4 explained why demand responds to price, time and frequency, and

Chapter 7 shows how to estimate those responses causally. A forecast has a narrower and more measurable goal: to be close to what happens. The handbook draws the line sharply: a model can predict well while implying misleading causal elasticities, so predictive accuracy is not evidence of causal understanding (Network Planning Handbook 2026). The two meet when a planner asks what will happen if Laurentian changes something, which needs both a forecast of the baseline and a causal estimate of the change.

The methods come from the forecasting textbook by Hyndman and Athanasopoulos (2021), which the handbook recommends as the methods companion for this work, applied throughout to simulated Laurentian data whose data-generating process is stated, so that every claim about accuracy can be checked against the truth.

8.1 What to forecast, and at which level

Airline demand can be forecast at four levels of detail, and each feeds different decisions (Figure 8.1). The **market** level is total travel between two cities by all carriers and paths. The **airline O&D** level is the airline's share of it, broken down by cabin, point of sale and week. The **itinerary and fare product** level splits that demand among the airline's paths and booking classes for each departure date. The **leg** level aggregates the itineraries that use each flight, by departure date and by days before departure.

The levels are linked by models, not only by sums. The airline's O&D demand is market demand times a share (Chapter 4); itinerary demand comes from a choice model (Chapters 5 and 6); leg demand is the sum of the itinerary demands that use the leg, (4.1). The last link has a consequence that is easy to miss: legs are not a level of the O&D hierarchy. A connecting passenger appears on two legs, so leg forecasts cannot be added up to O&D totals, and the handbook stresses that travellers counted by O&D and passengers counted by leg measure different things and do not belong in a single summing hierarchy (Network Planning Handbook 2026).

Three principles apply at every level.

Forecast demand, not bookings. Bookings and loads are censored by capacity and by revenue management, which closes fare classes before flights fill. Every forecast in this chapter is a forecast of unconstrained demand, obtained after unconstraining the history (Chapter 9).

Forecast what the decision needs, at its horizon. A seasonal schedule decision needs monthly or weekly demand a season ahead,

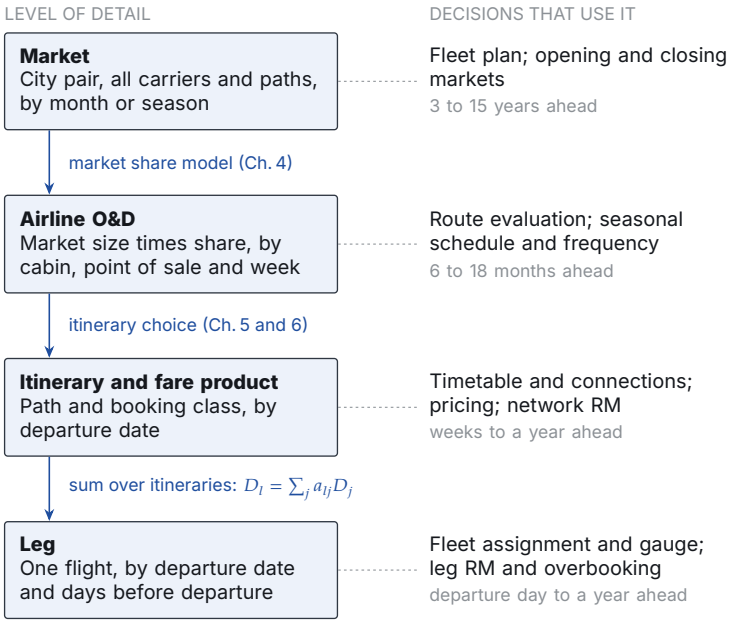


Figure 8.1 The forecasting hierarchy. Each level is derived from the one above by a model (market share, itinerary choice) or an accounting identity (the sum over itineraries that use a leg), and each serves different decisions at different horizons.

not next week's bookings (Network Planning Handbook 2026). Accuracy at a one-month horizon says little about accuracy at twelve.

Use only what was knowable at the time. A forecast for a decision made in March can use only data available in March: first-release traffic estimates, the competitor schedules published by then, the economic data as first reported. A backtest that uses revised data or schedules published later flatters the model (Network Planning Handbook 2026). Keep demand by travel date separate from bookings by booking date, because the same booking appears in both views at different times (Network Planning Handbook 2026).

Notation: y_t is the quantity in period t (passengers in a month, bookings for a departure), $\hat{y}_{T+h|T}$ is the forecast made at origin T for h periods ahead, and m is the seasonal period (12 for monthly data).

8.2 Benchmarks and exponential smoothing

The simplest forecasts are not straw men. They are the benchmarks against which every model is judged, and they are surprisingly hard to beat for seasonal airline series.

DEFINITION 8.1 **Naive and seasonal naive forecasts**

The **naive** forecast repeats the last observation, $\hat{y}_{T+h|T} = y_T$. The **seasonal naive** forecast repeats the observation from the same season of the last observed year:

$$\hat{y}_{T+h|T} = y_{T+h-m(k+1)}, \qquad k = \lfloor (h-1)/m \rfloor,$$

so that every January is forecast by the most recent January.

The handbook's advice is to begin with the seasonal naive forecast and to accept a more complex model only when it does better out of sample, on a task that matters for the decision (Network Planning Handbook 2026). A model that cannot beat "same month last year" has no business in a route proposal.

Simple exponential smoothing forecasts with a weighted average of past observations whose weights decline geometrically. In error-correction form, with level ℓ_t and smoothing parameter $0 < \alpha < 1$,

$$\ell_t = \ell_{t-1} + \alpha\,(y_t - \ell_{t-1}), \qquad \hat{y}_{T+h|T} = \ell_T,$$

and substituting backwards gives

$$\ell_T = \sum_{j=0}^{T-1} \alpha(1-\alpha)^j\, y_{T-j} + (1-\alpha)^T \ell_0.$$

A large α tracks recent observations closely; a small α averages over a long history.

Airline series have a trend and a strong season, so the workhorse is the **Holt–Winters** method, which smooths a level ℓ_t, a trend b_t and seasonal terms s_t, with parameters α, β^* and γ between 0 and 1. In its additive form (Hyndman and Athanasopoulos 2021):

$$\begin{aligned} \ell_t &= \alpha\,(y_t - s_{t-m}) + (1-\alpha)(\ell_{t-1} + b_{t-1}), \\ b_t &= \beta^*(\ell_t - \ell_{t-1}) + (1-\beta^*)\, b_{t-1}, \\ s_t &= \gamma\,(y_t - \ell_{t-1} - b_{t-1}) + (1-\gamma)\, s_{t-m}, \\ \hat{y}_{T+h|T} &= \ell_T + h\, b_T + s_{T+h-m(k+1)}. \end{aligned} \tag{8.1}$$

Each line is a weighted average of a new estimate and the previous one: the level is updated with the deseasonalized observation, the trend with the latest change in level, and each seasonal term with the latest deviation from trend. When seasonal swings grow with the level, as they do for airline traffic, the seasonal terms should be multiplicative. The simplest way to achieve that is to apply (8.1) to

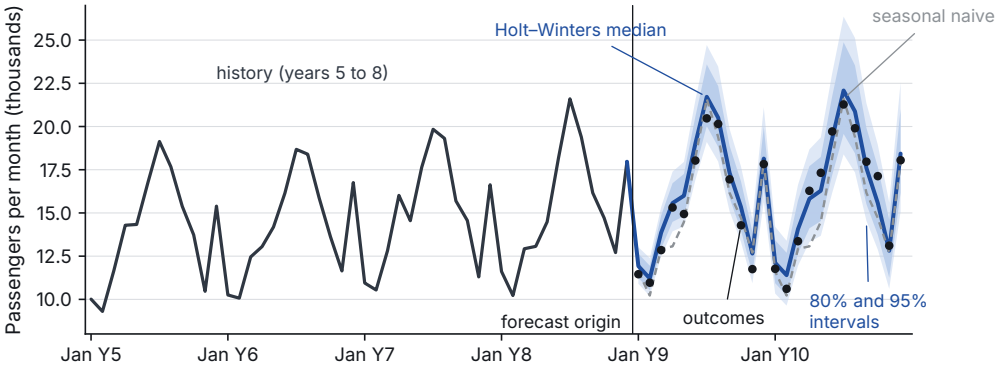


Figure 8.2 Simulated monthly Montreal–Paris O&D passengers for Laurentian. Forecasts made at the end of year 8 for 24 months: the Holt–Winters median with 80 and 95 percent prediction intervals from 5,000 simulated sample paths, the seasonal naive benchmark, and the outcomes that actually followed.

the logarithm of the series, which turns a multiplicative season and a constant growth rate into additive terms. The parameters and initial states are chosen to minimize the sum of squared one-step errors over the history. Written as a state space model with explicit errors, the same recursions give a likelihood, information criteria for choosing among variants, and prediction intervals (Hyndman and Athanasopoulos 2021).

WORKED EXAMPLE 8.1 Forecasting Montreal–Paris two years ahead

Data. Laurentian's monthly O&D passengers between Montreal and Paris, both directions, are simulated for ten years from

$$y_t = 12{,}000 \times 1.03^{(t-1)/12} \times S_{m(t)} \times e^{0.07 E_t} \times e^{u_t},$$

where S_m are monthly seasonal factors (0.71 in February to 1.315 in July), $E_t = 1$ in the month that contains Easter, which falls in March in years 3, 6 and 8 and in April otherwise, and $u_t = 0.6\,u_{t-1} + \varepsilon_t$ with $\varepsilon_t \sim \mathcal{N}(0, 0.035^2)$. Annual traffic grows from 148,759 passengers in year 1 to 183,036 in year 8.

Fit. Holt–Winters on the logarithms of months 1 to 96, with the initial states set from the first two years (a common shortcut) and only the three parameters estimated, gives $\alpha = 0.40$, $\beta^* = 0.00$ and $\gamma = 0.22$, with a residual standard deviation of 0.046, about 4.6 percent. The zero trend parameter means that

the growth rate is never updated: it stays at its starting value, 1.5 percent a year from the first two years, half the true 3 percent. The level absorbs most of the difference, but forecasts fall increasingly short as the horizon lengthens (Figure 8.3).

Forecast. Figure 8.2 shows the forecasts for years 9 and 10. The upper end of the 80 percent interval is 12 percent above its lower end one month ahead, 21 percent at twelve months and 30 percent at 24 months: uncertainty compounds with the horizon. Over the 24 months, the mean absolute error is 922 passengers for the seasonal naive forecast and 681 for Holt–Winters, mean absolute percentage errors of 5.7 and 4.4 percent. Of the 24 outcomes, 23 fall inside the 80 percent interval, more than the 19 or 20 one would expect, a first hint that the intervals are slightly too wide for this series.

Easter. In both forecast years the seasonal naive forecast makes its largest error in April, because it copies April of year 8, which had no Easter: for April of year 9 it forecasts 13,066 passengers against an outcome of 15,315. Holt–Winters forecasts 15,560, because its April seasonal term averages several years, most of which had Easter in April. Neither method knows when Easter falls; the next section adds that knowledge.

8.3 ARIMA and dynamic regression

ARIMA models describe the autocorrelation of a series rather than its components. Write B for the backshift operator, $By_t = y_{t-1}$. A series is made stationary by differencing, $(1-B)\,y_t = y_t - y_{t-1}$, or by seasonal differencing, $(1-B^m)\,y_t = y_t - y_{t-m}$, and the differenced series is modelled with autoregressive (AR) terms in its own past and moving-average (MA) terms in past shocks. The seasonal ARIMA$(p,d,q)(P,D,Q)_m$ model is

$$\phi(B)\,\Phi(B^m)\,(1-B)^d(1-B^m)^D\,y_t = c + \theta(B)\,\Theta(B^m)\,\varepsilon_t,$$

where ϕ, Φ, θ and Θ are polynomials of degrees p, P, q and Q and ε_t is white noise (Hyndman and Athanasopoulos 2021). Two special cases connect it to the methods above: simple exponential smoothing is an ARIMA$(0,1,1)$, and the seasonal naive forecast is the forecast of an ARIMA$(0,0,0)(0,1,0)_{12}$, a seasonal random walk. The most famous seasonal model in the literature, ARIMA$(0,1,1)(0,1,1)_{12}$ applied to logarithms, is called the **airline**

model because Box and Jenkins fitted it to monthly international airline passenger totals (Box and Jenkins 1970).

Airline demand also has structure that is known in advance and that neither smoothing nor ARIMA can see: **calendar effects**. Easter moves between March and April; Christmas and the New Year straddle two months; school breaks differ by province and country; Canadian and US Thanksgiving fall in different months; a month with five Fridays and five Sundays carries more weekend traffic than one with four. A **regression with calendar effects** puts these into the model directly. For log passengers $z_t = \ln y_t$:

$$z_t = \beta_0 + \beta_1 t + \sum_{j=2}^{12} \delta_j M_{jt} + \beta_E E_t + u_t,$$

where M_{jt} are month dummies and E_t is the Easter dummy. Economic regressors such as income, exchange rates or fares can be added, but they then have to be forecast themselves, so their benefit depends on how well they can be predicted at the horizon of the decision.

The errors u_t of such a regression are usually autocorrelated: a strong month tends to follow a strong month. Ordinary least squares then still gives unbiased coefficients but understates their uncertainty and wastes the information in recent errors. A **dynamic regression** models the error itself as an ARIMA process, for instance $u_t = \phi\, u_{t-1} + \varepsilon_t$, and forecasts with

$$\hat{z}_{T+h|T} = x'_{T+h}\hat{\beta} + \hat{\phi}^{\,h}\, \hat{u}_T,$$

so that the most recent surprise fades out geometrically (Hyndman and Athanasopoulos 2021). On the simulated Montreal–Paris series, this model estimates the Easter coefficient at 0.077 against a true 0.070, the trend at 0.033 a year in logarithms against a true 0.030, and ϕ at 0.58 against 0.6. Over months 97 to 120 its mean absolute error is 518 passengers, against 681 for Holt–Winters and 922 for the seasonal naive forecast.

Two cautions keep this result in proportion. The regression wins partly because the simulated world was built in its image: growth is constant, the Easter effect is fixed and the errors are exactly autoregressive. On real data the ranking is an empirical question, which is what the next section is for. And a regression that forecasts well does not deliver causal effects: if fares enter as a regressor, their coefficient mixes the response of demand to fares with the response of revenue management to demand (Chapter 7).

8.4 Measuring accuracy honestly

A forecast is useful only if its errors are known, and the way errors are summarized changes which model looks best.

DEFINITION 8.2 Forecast accuracy measures

For forecast errors $e_t = y_t - \hat{y}_t$ over a set of test periods:

- the **mean error** (bias) is $\text{ME} = \text{mean}(e_t)$; positive values mean the forecasts are too low;
- the **mean absolute error** is $\text{MAE} = \text{mean}\,|e_t|$ and the **root mean squared error** is $\text{RMSE} = \sqrt{\text{mean}(e_t^2)}$;
- the **mean absolute percentage error** is $\text{MAPE} = \text{mean}\,|e_t / y_t|$, the average error relative to the outcome;
- the **mean absolute scaled error** is $\text{MASE} = \text{mean}\,|e_t| \,/ Q$, where Q is the in-sample MAE of the seasonal naive forecast, $Q = \text{mean}_t\,|y_t - y_{t-m}|$ over the training data, with $m = 1$ for a series without seasonality (Hyndman and Koehler 2006).

MASE below one means the method beats the in-sample seasonal naive benchmark; it is scale-free, so it can be averaged across a large and a small market, and it is defined even when some outcomes are zero. Each measure also rewards a different forecast, which matters more than it seems.

RESULT 8.1 Which point forecast each measure rewards

Let $Y > 0$ have density f, with $\mathbb{E}[1/Y]$ finite. Among constant forecasts c:

1. $\mathbb{E}[(Y-c)^2]$ is minimized by the mean of Y;
2. $\mathbb{E}\,|Y-c|$ is minimized by the median of Y;
3. $\mathbb{E}\,|Y-c|/Y$ is minimized by the median of the density $g(y) \propto f(y)/y$, which is no larger than the median of Y.

So RMSE rewards the mean, MAE the median, and MAPE a forecast biased low.

Proof. For 1, $\mathbb{E}[(Y-c)^2] = \mathrm{Var}(Y) + (\mathbb{E}Y - c)^2$. For 2, the derivative of $\mathbb{E}\,|Y-c|$ with respect to c is $\Pr(Y < c) - \Pr(Y > c)$, which is zero at the median. For 3, $\mathbb{E}[|Y-c|/Y] = \int |y-c|\, f(y)/y \,\mathrm{d}y = \mathbb{E}[1/Y]\,\mathbb{E}_g\,|Y-c|$, where $g = f/(y\,\mathbb{E}[1/Y])$ is a density; by part 2 the minimizer is the median of g. Since $g/f \propto 1/y$ decreases in y, g puts relatively more weight on small values, so its distribution function lies above that of f and its median is no larger. Gneiting (2011) develops these results in general. ∎

MAPE has a second problem: it explodes when outcomes are small. Both problems are serious for airlines, which forecast thousands of thin flows, and the handbook warns that MAPE becomes unstable around small or zero denominators (Network Planning Handbook 2026).

WORKED EXAMPLE 8.2 Rolling-origin evaluation, and why MAPE misleads on thin flows

Rolling origins. On the simulated Montreal–Paris series, place the forecast origin at each month T from 60 to 108, refit every method on months 1 to T only, and forecast 1 to 12 months ahead: 49 origins and up to 588 forecasts per method. This is **rolling-origin evaluation**, the time-series version of cross-validation: every forecast uses only information available at its origin (Hyndman and Athanasopoulos 2021). Figure 8.3 summarizes the errors. One month ahead, the mean absolute errors are 2,621 passengers for the naive forecast, 742 for the seasonal naive, 507 for Holt–Winters and 398 for the dynamic regression; twelve months ahead, 857, 700 and 538 for the last three. In MASE terms the seasonal naive sits at 1.0 to 1.17,

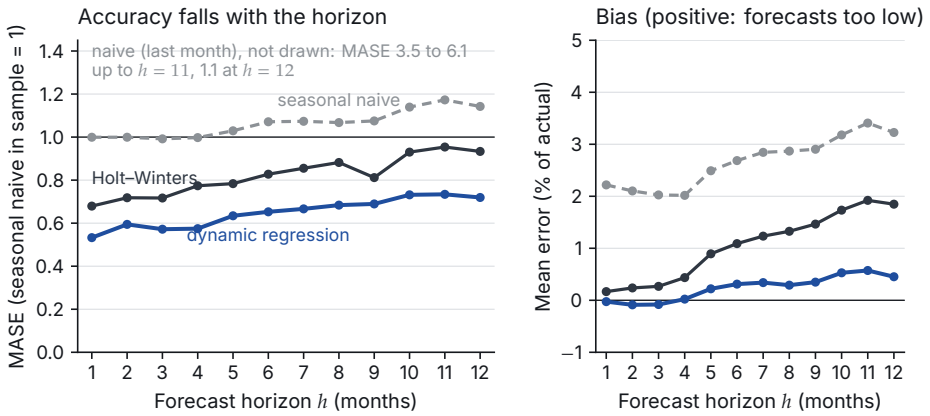


Figure 8.3 Rolling-origin evaluation of three methods on the simulated Montreal–Paris series, 49 origins. Left: mean absolute scaled error by forecast horizon; the line at 1 is the in-sample seasonal naive benchmark. Right: bias by horizon, as a percentage of the actual value; positive values mean the forecasts were too low.

Holt–Winters rises from 0.68 to 0.93 and the regression from 0.53 to 0.72. Errors grow with the horizon for the three seasonal methods, and the seasonal naive forecast is biased low by 2 to 3.4 percent because it ignores a year of growth. On March and April targets, where Easter moves, the seasonal naive MAPE is 8.9 percent, Holt–Winters 5.9 percent and the regression with its Easter dummy 4.2 percent.

A thin flow. Now take Laurentian's flexible-fare passengers from Quebec City to Seville, about four per departure: suppose they are Poisson with mean 4. The perfect forecast, 4, has a MAPE of 56 percent over the departures with at least one passenger, and MAPE is undefined for the 1.8 percent of departures with none. A forecast of 3 scores better, 46 percent, and is the MAPE-optimal constant, exactly as Result 8.1 predicts: MAPE rewards under-forecasting. A revenue management system tuned on MAPE would protect too few seats for its highest fares. MASE does not have this defect: the perfect forecast has MASE 0.70 against the naive benchmark, and the biased forecast of 3 does worse.

Four refinements turn this procedure into a decision tool. **Break errors down** by market size, season, direction, cabin and horizon, as the handbook recommends (Network Planning Handbook 2026): an average can hide a model that is excellent on large markets and poor on the marginal ones. **Track bias separately** from average error.

Over-forecasting marginal routes is costlier than errors of the same size on established ones (Network Planning Handbook 2026), because it is what gets weak routes approved. **Weight errors by what they cost.** A new model can lower the network-wide error while doing worse on the few capacity-constrained, high-revenue routes where decisions are actually made (Network Planning Handbook 2026); weighting errors by revenue, or by the value of the decisions they inform, exposes that. And **check the intervals**, not only the point forecasts: across all 49 origins and 12 horizons, the Holt–Winters 80 percent intervals contained the outcome 88 percent of the time (82 to 92 percent, depending on the horizon), which confirms that they are somewhat too wide for this series.

> **PITFALL**
>
> Look-ahead in backtests. A backtest that uses the latest, revised version of the data instead of what was available at each origin, or a competitor schedule filed after the forecast date, measures a forecast nobody could have made. Store data by vintage and rebuild each origin's information set exactly (Network Planning Handbook 2026).

8.5 Forecasts that add up

Forecasts made separately at different levels rarely agree. Laurentian's forecast of total Montreal–Europe traffic, produced by a model of the aggregate, will not equal the sum of its route forecasts. Network decisions need **coherent** forecasts, which respect the aggregation constraints, and the best way to obtain them uses all the forecasts rather than discarding some.

DEFINITION 8.3 Hierarchy, summing matrix and coherence

Let b_t be the vector of the n_b most disaggregated series and y_t the vector of all n series in a hierarchy, aggregates included. The **summing matrix** S, of size $n \times n_b$, expresses every series as a sum of bottom-level series: $y_t = S\,b_t$. A set of forecasts $\tilde{y}$ is **coherent** if $\tilde{y} = S\,\tilde{b}$ for some $\tilde{b}$.

For a Montreal–Europe total split into Paris, Lisbon and Seville,

$$\begin{pmatrix} y_{\text{tot}} \\ y_{\text{CDG}} \\ y_{\text{LIS}} \\ y_{\text{SVQ}} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}}_{S} \begin{pmatrix} y_{\text{CDG}} \\ y_{\text{LIS}} \\ y_{\text{SVQ}} \end{pmatrix}.$$

Given **base forecasts** $\hat{y}$ for all four series, produced independently, a linear reconciliation chooses an $n_b \times n$ matrix G and forms $\tilde{y} = S\,G\,\hat{y}$: G turns all the base forecasts into bottom-level forecasts, and S adds them up. The classical methods are special cases. **Bottom-up** uses $G = [\,0 \mid I\,]$, ignoring the total's forecast. **Top-down** uses $G = [\,p \mid 0\,]$, splitting the total's forecast in fixed proportions p and ignoring the route forecasts. Hierarchical forecasting of tourism demand, a close cousin of air travel, compares several ways of choosing those proportions (Athanasopoulos, Ahmed, and Hyndman 2009). Both methods throw information away.

RESULT 8.2 Minimum-trace reconciliation

Let the base forecasts be unbiased with error covariance matrix W, and consider reconciliations $\tilde{y} = SG\hat{y}$ that keep unbiased forecasts unbiased, which requires $GS = I$. The covariance of the reconciled errors is $SGWG'S'$, and its trace, the sum of the error variances of all series, is minimized by

$$G^* = \left(S'W^{-1}S\right)^{-1} S'W^{-1}. \tag{8.2}$$

With $W = \sigma^2 I$ this is ordinary least squares, $G = (S'S)^{-1}S'$; with W diagonal it is weighted least squares.

Proof. Any G with $GS = I$ can be written $G = G^* + D$ with $DS = 0$, since $G^*S = I$. Then $GWG' = G^*WG^{*\prime} + DWD' + G^*WD' + DWG^{*\prime}$,

and the cross terms vanish because $G^{*}WD' = (S'W^{-1}S)^{-1}S'D' = (S'W^{-1}S)^{-1}(DS)' = 0$. Hence $SGWG'S' = SG^{*}WG^{*\prime}S' + SDWD'S'$, and the last term is positive semidefinite, so its trace is non-negative and is zero when $D = 0$. ∎

This is the MinT method of Wickramasuriya, Athanasopoulos and Hyndman (2019), who estimate W from the in-sample one-step errors of the base forecasts, shrinking its off-diagonal elements toward zero for stability. For a single aggregation constraint, as in our example, the reconciliation has a transparent form. Write the constraint as $c'y = 0$ with $c = (1, -1, -1, -1)'$; the reconciled forecasts are then

$$\tilde{y} = \hat{y} - Wc\,\left(c'Wc\right)^{-1} c'\hat{y}, \tag{8.3}$$

the generalized least squares projection of the base forecasts onto the coherent set. With diagonal W, the incoherence $d = c'\hat{y}$ is shared among the series in proportion to their error variances: the series we trust least moves most.

WORKED EXAMPLE 8.3 Reconciling Montreal–Europe forecasts

Laurentian's base forecasts of next August's O&D passengers are 31,000 for the Montreal–Europe total, 19,800 for Paris, 6,900 for Lisbon and 3,500 for Seville. The routes sum to 30,200, so the forecasts are incoherent by $d = 800$. Backtests give error standard deviations of 1,500, 1,100, 600 and 500.

Table 8.1 Reconciled forecasts of August O&D passengers. All numbers are illustrative.

Method	Total	Paris	Lisbon	Seville
Base forecasts	31,000	19,800	6,900	3,500
Bottom-up	30,200	19,800	6,900	3,500
Top-down, forecast proportions	31,000	20,325	7,083	3,593
OLS	30,800	20,000	7,100	3,700
MinT, diagonal W	30,558	20,038	6,971	3,549
MinT, correlated errors	30,686	20,088	7,079	3,519

Bottom-up and top-down each keep one level and overrule the other. Top-down also has a structural problem here: Seville is a new route with no history, so historical proportions do not exist, and the table uses the proportions of the base forecasts instead.

OLS spreads the incoherence equally: $(S'S)^{-1}S'$ has rows such as $(0.25, 0.75, -0.25, -0.25)$, so each route rises by $800/4 = 200$ and the total falls by 200. MinT with the diagonal W uses (8.3): the variances are 2.25, 1.21, 0.36 and 0.25 million, summing to 4.07 million, so the total falls by $800 \times 2.25/4.07 = 442$, Paris rises by 238, Lisbon by 71 and Seville by 49. If the base errors really were independent with these variances, the reconciled total would have a standard deviation of about 1,000 passengers instead of 1,500: reconciliation improves the aggregate by borrowing strength from the routes. The last row uses a full covariance matrix in which the Paris and Lisbon errors are correlated (0.5) and each route's error is correlated with the total's; the answer moves again, which is why MinT estimates W from data rather than assuming it.

For the planner the lesson is that a single coherent set of numbers is a modelling choice with consequences. Bottom-up lets three route analysts overrule the network forecast; top-down lets the network forecast overrule them; MinT weighs each forecast by how good it has been.

The same machinery handles **grouped** structures, in which the same total is split along several dimensions at once (O&D by cabin and by point of sale), and **temporal** hierarchies, in which weekly forecasts must add up to monthly ones. It does not handle the leg dimension, for the reason given in Section 8.1: leg demand is a different measure, related to O&D demand through the incidence matrix and, once capacity binds, through spill.

8.6 Booking curves and pickup

Revenue management forecasts a different object: for each future departure, the demand still to come, given the bookings already on hand. Bookings accumulate along a **booking curve**, and its shape differs sharply by segment. In the PODS simulator's inputs, leisure travellers have made 75 percent of their final bookings 21 days before departure, while business travellers have made only 35 percent and book most heavily in the final two weeks (Belobaba 2006g). RM systems revise these forecasts at fixed checkpoints before departure, and the basic RM system that Belobaba describes forecasts the bookings still to come with simple **pickup** methods (Belobaba 2006l).

DEFINITION 8.4 Pickup forecasts

Let $B(\tau)$ be the bookings on hand τ days before departure and $D = B(0)$ the final (unconstrained) demand. From a history of comparable flown departures $h = 1, \dots, H$, the **additive pickup** forecast is

$$\hat{D} = B(\tau) + \frac{1}{H} \sum_h [B_h(0) - B_h(\tau)],$$

the **multiplicative pickup** forecast is

$$\hat{D} = B(\tau) \, \frac{\sum_h B_h(0)}{\sum_h B_h(\tau)},$$

and the **historical** forecast ignores bookings on hand, $\hat{D} = \frac{1}{H} \sum_h B_h(0)$. A **regression** forecast $\hat{D} = c_0 + c_1 B(\tau)$, fitted on the history, contains all three as special cases: $c_1 = 1$ (additive), $c_0 = 0$ (multiplicative) and $c_1 = 0$ (historical).

Which is right depends on why departures differ. If they differ only by chance, the bookings already made carry no information about the bookings to come, and the additive forecast is right. If they differ mainly because some departures are more popular than others, a departure that is ahead of its usual curve will probably stay ahead, and the multiplicative forecast is closer. A simple model makes this precise.

RESULT 8.3 Pickup under gamma-Poisson demand

Suppose final demand is $D \mid \nu \sim \text{Poisson}(\mu\nu)$, where the departure's popularity $\nu \sim \text{Gamma}(\kappa, \kappa)$ has mean one and variance $1/\kappa$, and each passenger has booked by τ independently with probability $F = F(\tau)$. Then

$$\mathbb{E}[D \mid B(\tau) = x] = x + \mu(1 - F)\,\frac{\kappa + x}{\kappa + \mu F}. \tag{8.4}$$

The best forecast is linear in bookings on hand. As $\kappa \to \infty$ (no popularity differences) it becomes the additive forecast $x + \mu(1-F)$; as $\kappa \to 0$ (large popularity differences) it becomes the multiplicative forecast x/F.

Proof. Given ν, Poisson thinning makes $B(\tau) \sim \text{Poisson}(\mu\nu F)$ and the remaining demand $D - B(\tau) \sim \text{Poisson}(\mu\nu(1-F))$, independent of each other. The gamma prior is conjugate to the Poisson likelihood, so $\nu \mid B(\tau) = x \sim \text{Gamma}(\kappa + x, \kappa + \mu F)$, with mean $(\kappa + x)/(\kappa + \mu F)$. Hence $\mathbb{E}[D - B(\tau) \mid x] = \mu(1-F)\,(\kappa + x)/(\kappa + \mu F)$. The limits follow directly. ∎

The result also explains why the regression form is a good default: it estimates the intercept and slope that (8.4) says are optimal, without having to know κ. In Belobaba's account of PODS simulations, replacing simple pickup with regression forecasts of bookings to come, together with better unconstraining, could matter as much for revenue as the choice of optimization method (Belobaba 2002).

WORKED EXAMPLE 8.4 How many passengers will the Seville flight have?

Data. Simulate YUL–SVQ departures in the shoulder season with $\mu = 150$ and $\kappa = 16$, so that popularity varies by 25 percent between departures and final demand has a standard deviation of about 40, the running case's values. A fifth of the passengers book on the PODS business curve and the rest on the leisure curve, so that 28 days before departure a fraction $F = 0.2 \times 0.28 + 0.8 \times 0.61 = 0.544$ of final demand has booked. The history is 40 flown departures. A current departure has 96 bookings 28 days out.

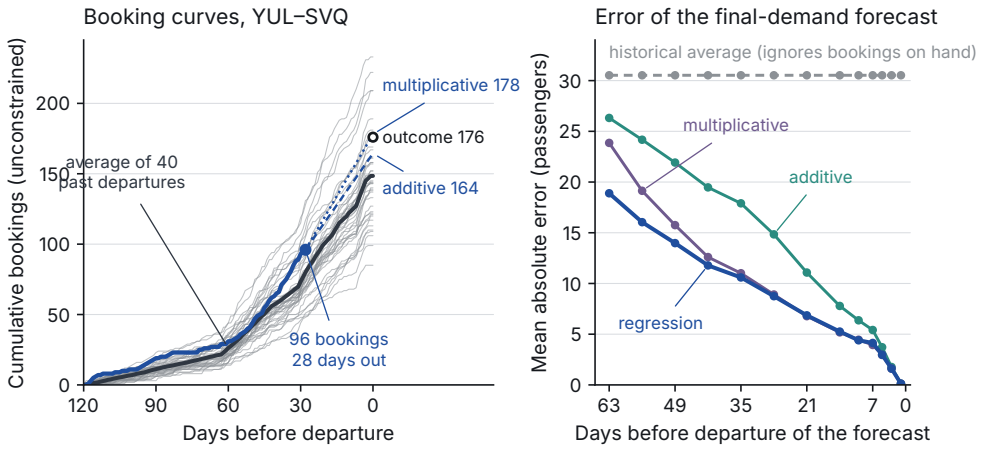


Figure 8.4 Left: booking curves of 40 simulated past YUL–SVQ departures, their average, and a current departure with 96 bookings 28 days before departure, projected to departure by additive and multiplicative pickup. Right: mean absolute error of four forecasts of final demand, by the day on which the forecast is made, over 3,000 simulated departures.

Forecasts. On average the 40 past departures had 80.0 bookings at 28 days and 148.5 at departure. Additive pickup adds the average pickup of 68.5: $\hat{D} = 164$. Multiplicative pickup applies the ratio $148.5/80.0 = 1.855$: $\hat{D} = 178$. The regression fitted on the history is $\hat{D} = 20.1 + 1.604\,B(28)$, giving 174, close to the conditional expectation (8.4), $11.2 + 1.701 \times 96 = 174.5$. The departure eventually has 176 passengers.

Accuracy. Over 3,000 simulated departures, 28 days out, the mean absolute errors are 30.5 passengers for the historical forecast, 14.8 for additive pickup, 8.9 for multiplicative pickup, 8.7 for the regression and 8.5 for the true conditional expectation. Popularity differences dominate here, so multiplicative beats additive. Sixty-three days out, when few bookings are on hand, the multiplicative ratio is 6.9 and amplifies noise: its error is 23.9 against 18.9 for the regression. If the departures differed only by chance, the ranking would reverse: with $\kappa \to \infty$, additive pickup has an error of 6.7 against 8.9 for multiplicative pickup 28 days out.

The planner should take two things from the example. Forecasts improve sharply as departure approaches, which is why capacity decisions made months ahead, such as gauge swaps, rely on far noisier numbers than revenue management's last-week decisions (Chapter 22). And a forecast of total demand

by departure needs the history to be unconstrained first: if low fares closed early on the busiest past departures, their recorded bookings understate demand and every pickup estimate inherits the bias (Chapters 9 and 21).

8.7 Markets without history

Time series methods need history. A new route has none, and a thin market served only by connections has little that can be trusted. Planners then combine three kinds of evidence: cross-sectional models of market size, analogues from comparable launches, and share models.

The standard cross-sectional model is the **gravity model**, named for its analogy with Newton's law: traffic between two cities rises with their size and falls with the distance between them. In logarithms,

$$\ln D_{ij} = \beta_0 + \beta_1 \ln(P_i P_j) - \beta_2 \ln d_{ij} + \gamma' w_{ij} + u_{ij},$$

where D_{ij} is O&D traffic between cities i and j, P_i and P_j their populations (or incomes), d_{ij} the distance, and w_{ij} other characteristics such as cultural and language ties, tourism appeal or the presence of a nonstop. It is estimated by least squares across many city pairs. Two details matter. Write x_0 for the full vector of regressors of a new pair and X for the matrix of regressors of the sample. The prediction $\exp(x_0'\hat\beta)$ estimates the median of D, not its mean; under normal errors the mean is $\exp(x_0'\hat\beta + \sigma^2/2)$. And the prediction interval in logarithms, $x_0'\hat\beta \pm t\, s\sqrt{1 + x_0'(X'X)^{-1}x_0}$, where s is the residual standard error and t a Student quantile, becomes a multiplicative interval in levels, which is the honest way to report it.

WORKED EXAMPLE 8.5 A gravity model for Montreal–Seville

Data. Simulate annual O&D traffic for the 63 pairs between nine North American cities (six Canadian and three US) and seven European ones from

$$\ln D_{ij} = 19.87 + 0.8 \ln(P_i P_j) - 1.2 \ln d_{ij} + 0.6\,\mathrm{Tie}_{ij} + 0.22\,\mathrm{NS}_{ij} + u_{ij},$$

with rounded illustrative metropolitan populations P in millions (4.3 for Montreal, 1.5 for Seville), great-circle distances

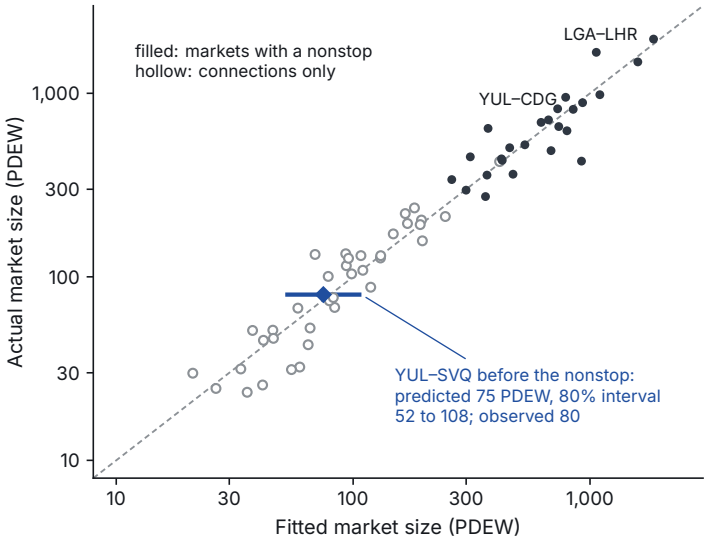


Figure 8.5 A gravity model fitted to 62 simulated city pairs between North America and Europe, actual against fitted market size on logarithmic scales. Pairs with a nonstop (filled) cluster at the top because nonstops were placed in the strongest markets. The bar shows the 80 percent prediction interval for Montreal–Seville before its nonstop, against the 80 PDEW assumed in Example 4.4.

d in kilometres, a language tie Tie_{ij} between Paris and Montreal, Quebec City or Ottawa, and $u_{ij} \sim \mathcal{N}(0, 0.35^2)$. A nonstop ($\text{NS}_{ij} = 1$) raises traffic by 25 percent ($e^{0.22} = 1.25$), but the simulated airlines placed nonstops in the 40 percent of markets they saw as strongest, so NS_{ij} is correlated with u_{ij}. Montreal–Seville has no nonstop and its true size is 80 PDEW, as in Example 4.4; it is left out of the estimation sample.

Estimates. Least squares on the 62 other pairs gives a population elasticity of 0.64 (standard error 0.04), a distance elasticity of −0.68 (0.23), a tie effect of 0.30 (0.17) and a nonstop effect of 0.79 (0.10), with a residual standard deviation of 0.28 and $R^2 = 0.95$.

Prediction. For Montreal–Seville without a nonstop, the model predicts a median of 54,650 passengers a year in both directions, 75 PDEW, with an 80 percent interval from 52 to 108 PDEW. The truth, 80, is comfortably inside. A gravity model is a useful cross-check on a market estimate from booking data, and its interval, a factor of about 1.45 either side of the median, is a

fair statement of how much a cross-section knows about one market.

The trap. The fit is excellent, yet every coefficient is wrong, because the nonstop dummy is endogenous. Its coefficient, 0.79, says that a nonstop multiplies traffic by 2.2; the truth is 1.25. Markets with nonstops are larger partly because of the nonstop and partly because airlines chose them for being large. A planner who read the coefficient as the stimulation from a new Seville nonstop would forecast 165 PDEW instead of 100. The selection biases the other coefficients too: across 2,000 simulated samples the population and distance elasticities average 0.71 and −1.02 instead of 0.8 and −1.2, and the nonstop coefficient averages 0.57 against a true 0.22; this particular sample adds sampling noise to the bias.

The second source of evidence is **analogues**: past launches that resemble the new route in distance, market size, mix of business and leisure travel, competition and frequency. For each analogue, measure the ratio of first-year traffic to the pre-launch market, the share the new service won, and the **ramp-up**, the months it took to reach mature traffic. The distribution of those ratios across analogues is the empirical counterpart of the stimulation parameter λ of Result 4.6, and the handbook insists that stimulation be supported by evidence from comparable launches and service changes rather than assumed (Network Planning Handbook 2026).

The third is the share model of Chapter 4, which converts a market estimate into the new service's traffic and separates its sources: stimulated demand, demand diverted from competitors, demand diverted from the airline's own services, and spill from capacity constraints, which the handbook asks to be kept apart (Network Planning Handbook 2026). The three sources of evidence fail in different ways, which is the argument for using all of them: booking data miss channels, gravity models miss market idiosyncrasies, and analogues are never exact.

> **PITFALL**
>
> Reading a cross-sectional coefficient as a causal effect. A nonstop dummy, a frequency variable or a fare in a cross-section of markets measures association, and airlines put nonstops, frequencies and fares where demand already is. Use such models

> to predict the size of markets like the ones in the sample, not to predict what a change in service will do (Chapter 7).

8.8 Uncertainty, scenarios and the decision

A forecast without a statement of its uncertainty invites the decision maker to treat it as a fact. A **prediction interval** gives a range that contains the outcome with stated probability; a **quantile forecast** gives the outcome that will be exceeded with stated probability. Both come out of the models above, from formulas or from simulated sample paths as in Figure 8.2, and both should be checked against outcomes, as the rolling-origin example did. The handbook recommends distributional forecasts where the cost of a bad outcome dominates the decision (Network Planning Handbook 2026), which describes most capacity decisions: an aircraft committed to a route that underperforms cannot be recovered quickly.

Statistical intervals capture only part of the uncertainty. The handbook lists five sources that a planner should keep apart: parameter uncertainty, macroeconomic scenarios, competitor response, launch ramp-up and operational feasibility, and observes that a single symmetric interval rarely represents all of them well (Network Planning Handbook 2026). The practical answer is to combine intervals for the statistical part with **scenarios** for the structural part. Scenarios must be internally coherent: lower demand tends to come with lower fares, and toggling every input independently produces combinations that cannot occur (Network Planning Handbook 2026).

WORKED EXAMPLE 8.6 Communicating the Seville forecast

The running case gives mature eastbound demand per departure of 175 passengers at the autumn and spring peaks, 150 in the summer shoulder season and 110 in winter, from November to March (the calendar of Chapter 12). Suppose analogues suggest that a new transatlantic route starts at 80 percent of mature demand and ramps up linearly over its first year. The uncertainty about the market as a whole, combining the market size estimate, the stimulation parameter and the share, is represented by a single factor common to all months, lognormal with a standard deviation of 0.20 in logarithms; month-specific noise adds a standard deviation of 0.06.

Break-even. With the running case's costs, the A321XLR's seven-hour trip costs CAD 80,500 before passenger costs, each passenger contributes CAD 600 (a net fare of 620 plus 30 of ancillary revenue, less 50 of passenger costs), and cargo contributes CAD 3,500, so the leg breaks even at $(80{,}500 - 3{,}500)/600 = 128$ passengers.

Fan chart. Figure 8.6 shows 20,000 simulated paths. The median starts at 140 in the launch month. The chance of beating break-even is 66 percent in the first September, 6 percent in the first November, 87 percent in the first April and 77 percent in the first August, and 93 and 23 percent in the second year's autumn and winter months. The 80 percent band spans a factor of 1.7 in every month, because it is dominated by the common market factor, which does not average out over time.

Scenarios. An adverse case with demand 20 percent lower and a net fare of CAD 590 raises the break-even load to 135 passengers while mature shoulder demand falls to 120; a favourable case with 15 percent more demand and a fare of CAD 640 lowers it to 124 against 172.

The message for the decision maker is not "the route carries 150 passengers". It is that autumn and spring are very likely profitable at the leg level once the route has ramped up, summer probably so, and winter probably not, unless connecting traffic, which this leg-level view treats crudely, changes the picture. That points to the decisions worth discussing, such as a winter downgauge or suspension, and to the assumption that would reverse the recommendation, here the level of the common market factor.

Three habits make forecasts useful to decision makers. State the decision threshold and the probability of crossing it, as in the example, rather than a single number. Name the assumption that reverses the recommendation; the handbook calls this the most useful sensitivity (Network Planning Handbook 2026). And keep a record: after launch, compare outcomes with the forecast as it was approved, not with a later revision, and separate a slow ramp-up from a persistent bias (Network Planning Handbook 2026). A forecasting process that never meets its own past forecasts does not learn.

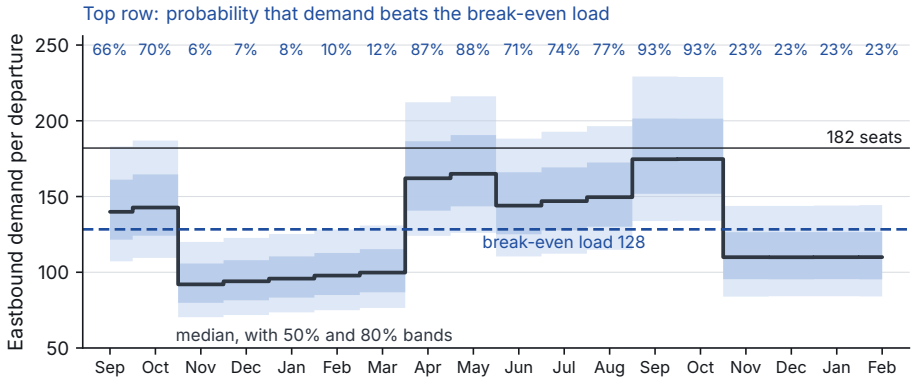


Figure 8.6 Fan chart of eastbound YUL–SVQ demand per departure for the 18 months after a September launch: median with 50 and 80 percent bands, against the leg-level break-even load and the 182 seats of the A321XLR. The top row gives the probability of beating break-even in each month. Illustrative assumptions.

IN PRACTICE Forecasting in 2026

Vendors of revenue management systems describe a move from the pickup and time series methods of this chapter toward Bayesian and machine-learning models, and toward forecasting willingness to pay rather than demand by booking class. PROS describes its RM forecasting as combining Bayesian statistics and artificial intelligence, and reports that adding willingness-to-pay data raises revenue by about 1 to 3 percent, a vendor figure without published methodology (PhocusWire n.d.). Fetcherr markets a generative "large market model" that ingests schedules, inventory, OAG data, weather and financial-market data to recommend fares (Boynton and Massy-Beresford 2025). Academic comparisons of reinforcement learning with dynamic programming for airline pricing are still experiments, and a 2025 editorial in the Journal of Revenue and Pricing Management calls for empirical validation of such methods (Yeoman 2025). Which forecasting engine a given airline runs is usually not public. On the planning side, the data themselves move: Cirium releases monthly O&D estimates about 55 days after the end of a month and restates them at 115 days (Cirium 2026a), so a backtest must record which vintage it used, and the US ticket sample changed from a quarterly 10 percent sample to a monthly 40 percent sample in July 2025, a break that should not be spliced naively (Network Planning Handbook 2026). Scenario analysis is also becoming more quantitative: Monte Carlo stress tests of the day-of-operations performance of future schedules were on the

2026 program of the AGIFORS scheduling and strategic planning group (AGIFORS 2026).

8.9 Summary

- Forecast unconstrained demand at the level and horizon the decision needs: markets for fleet and route decisions, airline O&D for seasonal schedules, itineraries and fare products for pricing and network RM, legs for fleet assignment and leg RM. Legs are not a level of the O&D hierarchy.
- Start from the seasonal naive benchmark. Holt–Winters smooths level, trend and season, preferably on logarithms for airline series; ARIMA models autocorrelation; dynamic regression adds calendar effects such as Easter that no smoothing method can see.
- Evaluate with rolling origins that use only what was knowable at each origin. Errors grow with the horizon; report them by horizon, market, season and direction, track bias separately, and weight errors by what they cost.
- RMSE rewards the mean, MAE the median and MAPE a forecast biased low; MAPE is also unstable for thin flows. MASE is scale-free and robust.
- Coherent forecasts come from reconciliation, $\tilde{y} = SG\hat{y}$. Bottom-up and top-down discard information; MinT (8.2) minimizes the total error variance and, with independent errors, shares any incoherence in proportion to error variances.
- Pickup forecasts project bookings on hand to departure. Under gamma-Poisson demand the best forecast is linear in bookings on hand, between additive pickup (departures differ by chance) and multiplicative pickup (departures differ in popularity).
- For new routes, combine gravity models, analogues and share models. Cross-sectional coefficients on nonstops or frequencies are not causal, and stimulation must come from comparable launches.
- Report intervals, quantiles and coherent scenarios, state the probability of crossing the decision threshold, name the assumption that would reverse the recommendation, and review outcomes against the approved forecast.

Table 8.2 Key formulas of the chapter.

Quantity	Formula
Seasonal naive	$\hat{y}_{T+h\|T} = y_{T+h-m(k+1)}$, $k = \lfloor (h-1)/m \rfloor$
Holt–Winters forecast	$\hat{y}_{T+h\|T} = \ell_T + h\, b_T + s_{T+h-m(k+1)}$
Dynamic regression	$\hat{z}_{T+h\|T} = x'_{T+h}\hat{\beta} + \hat{\phi}^h \hat{u}_T$
MASE	$\text{mean}\,\lvert e_t \rvert \,/\, \text{mean}\,\lvert y_t - y_{t-m} \rvert$
Reconciliation	$\tilde{y} = SG\hat{y}$, $G^* = (S'W^{-1}S)^{-1}S'W^{-1}$
One constraint	$\tilde{y} = \hat{y} - Wc\,(c'Wc)^{-1}c'\hat{y}$
Pickup (gamma-Poisson)	$\mathbb{E}[D \mid x] = x + \mu(1-F)(\kappa + x)/(\kappa + \mu F)$
Gravity model	$\ln D_{ij} = \beta_0 + \beta_1 \ln(P_i P_j) - \beta_2 \ln d_{ij} + \gamma' w_{ij}$

8.10 Further reading

The handbook's forecasting guidance (Network Planning Handbook 2026) is short and worth rereading after this chapter. Hyndman and Athanasopoulos (2021) is the methods companion: its chapters on evaluation, exponential smoothing, ARIMA, dynamic regression and reconciliation cover everything in Sections 8.2 to 8.5 with code. Hyndman and Koehler (2006) introduce MASE, Gneiting (2011) explains which point forecast each error measure rewards, and Wickramasuriya, Athanasopoulos and Hyndman (2019) derive MinT. Box and Jenkins (1970) is the origin of the airline model. For booking curves and RM forecasting, Belobaba's PODS and RM lectures (Belobaba 2006g, 2006l, 2002) give the airline context and Talluri and van Ryzin (2004b, ch. 9) the theory of estimation and forecasting for revenue management. Doganis (2019) discusses demand forecasting from the commercial side.

CHAPTER

9

Spill and Recapture

IN THIS CHAPTER

- Why a flight that is 85 percent full on average can still be turning away one customer in eight.
- The expected-spill formula under normal demand, derived step by step and read as a picture.
- Demand factor, load factor and spill rate, and the curves that tie them together.
- How to recover true demand from loads that capacity has capped.
- What one more seat is worth, and why the answer changes once revenue management controls the flight.
- Recapture, and why the cost of spill is a network quantity rather than a flight quantity.

DRAWS ON Belobaba, Airline demand analysis and spill modeling (MIT 16.75J, 2006); Boeing, Load factor analysis (1978); handbook Exercise B; Barnhart, Kniker and Lohatepanont (2002); Abramovich (2013)

Every network decision is a bet on demand that nobody observes. When a planner assigns a 137-seat A220 instead of a 169-seat 737 to the 17:00 departure to Vancouver, the airline commits seats months before any customer shows up, and the customers who would have flown on a bigger aircraft leave no trace in the airline's records. The flight closes, the booking engine stops offering it, and the would-be passengers buy another airline's ticket, travel another day, or stay home. This chapter is about those invisible customers. It gives you the vocabulary (demand, load, spill, recapture), the one formula that sits underneath most gauge and fleet decisions, the statistics needed to estimate demand that has been censored by capacity, and the economic logic that turns spill into dollars.

Spill is where the two halves of this book meet. The planner decides how many seats exist (Part III); revenue management decides who gets them (Part IV). A spill estimate that ignores how revenue

management fills the aircraft will systematically misprice capacity, a point we close the chapter with and return to in Chapter 22.

9.1 Demand, load and spill

Fix one flight leg, say Laurentian's Friday 09:00 departure from Montreal, and look at a long series of its departures operated with the same aircraft. For each departure distinguish three quantities.

DEFINITION 9.1 Demand, load and spill

The **unconstrained demand** D of a departure is the number of customers who would book it at the prevailing fares and schedule if seats were unlimited. The **capacity** C is the number of seats offered for sale. The **load** is the number of passengers actually carried,

$$Y = \min(D, C),$$

and the **spill** (or rejected demand) is the demand that could not be accommodated,

$$S = D - Y = [D - C]^{+} = \max(D - C, 0).$$

Two remarks prevent most early confusion. First, **demand here is leg demand**: it aggregates every origin-destination journey that wants a seat on this leg, local Montreal travellers and connecting travellers alike. How O&D demand is assembled into leg demand is the subject of Chapters 4 and 10; in this chapter we take the leg as given. Second, **spill is not a denied boarding**. Spill happens while the flight is on sale: a request arrives, no seat is available at any fare the customer accepts, and the customer goes elsewhere. A denied boarding happens at the gate, when more holders of confirmed reservations show up than there are seats, which is a consequence of overbooking (Chapter 19). A flight can spill without any overbooking, and it can deny boardings without having spilled anyone during the sale.

The difficulty is that the airline records Y, not D. The lecture example from the reading library (Belobaba 2006b) makes the point with five Friday departures of a 125-seat flight from Montreal to Los Angeles (Table 9.1).

Table 9.1 Five hypothetical Friday departures of a 125-seat YUL–LAX flight (adapted from Belobaba, 2006). The average load factor is 85.0 percent.

Date	Load	Capacity	Load factor	Spill?
1 Apr	92	125	73.6%	no
8 Apr	125	125	100.0%	likely
15 Apr	108	125	86.4%	no
22 Apr	83	125	66.4%	no
29 Apr	123	125	98.4%	possibly

On 1, 15 and 22 April the flight left with empty seats, so its load is its demand. On 8 April it left full: all we know is that demand was at least 125, and it was probably more. On 29 April it left with 123 passengers, two short of capacity; a flight that sells out a few days before departure and then loses two no-shows looks exactly like this, so spill was possible. Statisticians call such observations **censored**: for some departures we see the true value, for others only a lower bound.

Loads are what the airline sold, not what it could have sold. Every estimate of demand for a flight constrained by capacity is a model, never a measurement.

9.2 A probability model for one flight

Demand for "the Friday 09:00" varies from one departure to the next. Part of the variation is predictable: seasons, holidays, the day of the week, school breaks. Planners remove it by working with homogeneous sets of departures, or by modelling it explicitly (Chapter 8). What remains is variation that no forecast can remove, and we represent it with a probability distribution for D, with mean μ and standard deviation σ.

The size of that residual variation is summarized by a single ratio.

> **DEFINITION 9.2 k-factor**
>
> The **k-factor** (coefficient of variation) of demand is $k = \sigma / \mu$.

Empirical studies of airline data typically find k between 0.20 and 0.40, with leisure markets at the high end and business or mixed markets at the low end, because business demand is more regular

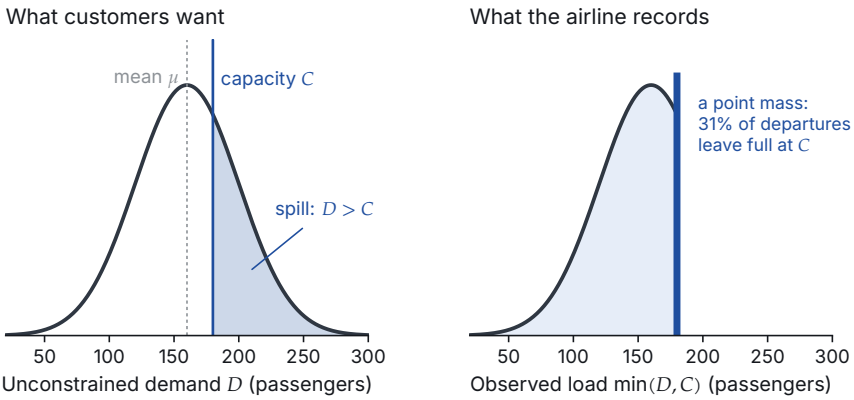


Figure 9.1 Left: the distribution of unconstrained demand for a departure with $\mu = 160$, $\sigma = 40$ and $C = 180$ seats; the shaded tail is the demand that cannot be carried. Right: the distribution of the load the airline actually records. Below capacity it follows demand; at capacity it has a point mass equal to the probability that demand reaches C, here 31 percent.

from week to week (Belobaba 2006b). Keep the order of magnitude in mind: with $k = 0.3$, a flight with mean demand of 150 passengers routinely sees departures with 105 and with 195.

The classical spill model assumes that D is **normally distributed**. The justification is pragmatic. Demand for a departure is the sum of many small, roughly independent booking decisions, so a central-limit argument makes a bell shape plausible; the normal family is pinned down by the two numbers planners can estimate; and the resulting formulas are explicit. The assumption has known flaws: demand is a non-negative integer, small flights are right-skewed, and the Poisson distribution that a purely random arrival process would produce has $k = 1/\sqrt{\mu}$, far smaller than observed (a flight with mean 150 would have $k \approx 0.08$). Real demand is **overdispersed** because the arrival rate itself varies with weather, events and the economy. Alternatives such as the gamma or negative binomial distributions handle skewness better, and Swan (2002) concludes that gamma shapes suit revenue management while the normal suits spill modelling; for flights with mean demand above about 50 the normal model is accurate enough for planning, and it is the language your colleagues will use.

Figure 9.1 shows what capacity does to the data. The load distribution is the demand distribution with its right tail folded onto a single value. The fold is the reason naive statistics on loads mislead: the mean of the loads is too low, and so is their standard deviation, because the departures with the highest demand are exactly the ones that hit the ceiling.

9.3 The expected spill formula

Planning decisions need the expected spill per departure, $\mathbb{E}[S]$, as a function of capacity. With a density f for demand,

$$\mathbb{E}[S] = \mathbb{E}[[D - C]^+] = \int_C^\infty (x - C) f(x)\, \mathrm{d}x.$$

For normal demand this integral has a closed form. Write the capacity in standard units,

$$z = \frac{C - \mu}{\sigma},$$

so that $z = 0$ means the aircraft has exactly as many seats as mean demand and $z = 1$ means it has one standard deviation of spare room.

RESULT 9.1 Expected spill and load under normal demand

If $D \sim \mathcal{N}(\mu, \sigma^2)$ and $z = (C - \mu)/\sigma$, then

$$\mathbb{E}[S] = \sigma\, L(z), \qquad L(z) = \varphi(z) - z(1 - \Phi(z)), \tag{9.1}$$

where φ and Φ are the standard normal density and distribution function. Consequently the expected load is $\mathbb{E}[Y] = \mu - \sigma L(z)$ and the expected load factor is $(\mu - \sigma L(z))/C$.

Proof. Substitute $x = \mu + \sigma t$, so that $f(x)\,\mathrm{d}x = \varphi(t)\,\mathrm{d}t$ and $x - C = \sigma(t - z)$:

$$\mathbb{E}[S] = \sigma \int_z^\infty (t - z)\, \varphi(t)\, \mathrm{d}t = \sigma \left[\int_z^\infty t\, \varphi(t)\, \mathrm{d}t \;-\; z \int_z^\infty \varphi(t)\, \mathrm{d}t \right].$$

The second integral is $1 - \Phi(z)$. For the first, use $\varphi'(t) = -t\,\varphi(t)$, so $\int_z^\infty t\,\varphi(t)\,\mathrm{d}t = [-\varphi(t)]_z^\infty = \varphi(z)$. Combining gives $\sigma[\varphi(z) - z(1 - \Phi(z))]$. The load follows from $Y = D - S$ and $\mathbb{E}[D] = \mu$. ∎

The function $L(z)$ is the **standard normal loss function**: the expected amount by which a standard normal variable exceeds z. It is tabulated in every operations textbook and available in any spreadsheet as `NORM.S.DIST(z,FALSE) - z*(1-NORM.S.DIST(z,TRUE))`. Three of its values are worth remembering. When the aircraft is sized at mean demand ($z = 0$), $L(0) = \varphi(0) \approx 0.40$: the flight spills on average 40 percent of a standard deviation, even though it is "the right size on average". When capacity is far below demand ($z \ll 0$), $L(z) \approx -z$, so $\mathbb{E}[S] \approx \mu - C$, the deterministic answer. When capacity exceeds mean demand by two standard deviations, $L(2) \approx 0.008$ and spill is negligible.

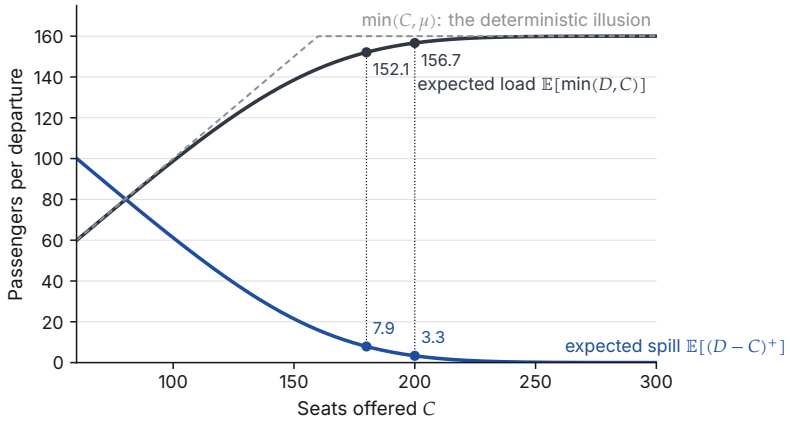


Figure 9.2 Expected load and expected spill as capacity varies, for $\mu = 160$ and $\sigma = 40$. The dashed line $\min(C, \mu)$ is what a planner who ignores variability would expect. The marked points are the two aircraft compared in Example 9.1.

How does expected load respond to one more seat? Differentiating (9.1) with $\mathrm{d}z/\,\mathrm{d}C = 1/\sigma$ and using $L'(z) = -(1 - \Phi(z))$ gives the most useful single result in capacity planning.

RESULT 9.2 The marginal seat

$$\frac{\mathrm{d}\,\mathbb{E}[Y]}{\mathrm{d}C} = \Pr(D > C) = 1 - \Phi(z), \qquad \frac{\mathrm{d}\,\mathbb{E}[S]}{\mathrm{d}C} = -\Pr(D > C).$$

The C-th seat is occupied only on departures whose demand exceeds $C - 1$, so on average it carries $\Pr(D \geq C)$ passengers. More generally, for any demand distribution,

$$\mathbb{E}[Y] = \mathbb{E}[\min(D, C)] = \int_0^C \Pr(D > x)\,\mathrm{d}x. \tag{9.2}$$

Equation (9.2) needs no normality: seat number x is used exactly when demand exceeds x, and expected load is the sum over seats of those probabilities. It also shows that expected load is **concave** in capacity. Each additional seat is used less often than the previous one, which is the economic heart of every gauge decision.

WORKED EXAMPLE 9.1 Twenty seats are not twenty passengers

This is the handbook's Exercise B, solved in full. Unconstrained demand on a departure is normal with $\mu = 160$ and $\sigma = 40$. Compare a 180-seat aircraft with a 200-seat aircraft.

For $C = 180$, $z = (180 - 160)/40 = 0.5$. From standard tables $\varphi(0.5) = 0.3521$ and $1 - \Phi(0.5) = 0.3085$, so $L(0.5) = 0.3521 - 0.5 \times 0.3085 = 0.1978$ and

$$\mathbb{E}[S] = 40 \times 0.1978 = 7.91, \qquad \mathbb{E}[Y] = 160 - 7.91 = 152.09,$$

an expected load factor of $152.09/180 = 84.5$ percent.

For $C = 200$, $z = 1$, $\varphi(1) = 0.2420$, $1 - \Phi(1) = 0.1587$, $L(1) = 0.0833$, so $\mathbb{E}[S] = 3.33$ and $\mathbb{E}[Y] = 156.67$.

The twenty extra seats carry $156.67 - 152.09 = 4.58$ additional passengers per departure, not twenty. Result 9.2 explains the number: the probability that demand exceeds x falls from 0.31 at $x = 180$ to 0.16 at $x = 200$, and the average of $\Pr(D > x)$ over those twenty seats is $4.58/20 = 0.229$.

Now value them. Suppose 25 percent of the newly carried passengers would otherwise have been recaptured on another Laurentian flight with the same contribution, and each truly incremental passenger contributes CAD 180. The upgauge is worth $4.58 \times 0.75 \times 180 \approx 618$ dollars per departure. If the larger aircraft costs CAD 900 more per trip, it is unattractive by about CAD 282. The arithmetic is simple; its value is that every assumption (normal demand, fixed fares, constant recapture, no stimulation, a single cabin) is visible and can be challenged.

PITFALL

Replacing demand by its mean. A planner who reasons "demand is 160, so 180 seats carry 160 passengers" overstates the load by 7.9 passengers and understates spill by the same amount. The error is largest exactly where decisions are made, when capacity is close to mean demand.

9.4 Load factor, demand factor and spill rate

Airline managers talk in ratios, because ratios compare flights of different sizes. Four of them organize the spill problem.

DEFINITION 9.3 Spill ratios

For a series of departures with capacity C:

- the **demand factor** is $\mathrm{DF} = \mu/C$, which can exceed 1;
- the **average load factor** is $\overline{\mathrm{LF}} = \mathbb{E}[Y]/C$, which cannot;
- the **spill factor** is $\mathrm{SF} = \mathbb{E}[S]/C$;
- the **spill rate** is $\mathrm{SR} = \mathbb{E}[S]/\mu$, the share of all demand that is turned away, or equivalently the probability that a randomly chosen would-be passenger cannot get a seat.

Because $\mu = \mathbb{E}[Y] + \mathbb{E}[S]$, the first identity is immediate: $\mathrm{DF} = \overline{\mathrm{LF}} + \mathrm{SF}$, and $\mathrm{SR} = \mathrm{SF}/\mathrm{DF}$. The second fact is less obvious and more powerful.

RESULT 9.3 Spill ratios depend only on the demand factor and k

For normal demand with k-factor k, let $z = (1 - \mathrm{DF})/(k\,\mathrm{DF})$. Then

$$\mathrm{SF} = k\,\mathrm{DF}\,L(z), \qquad \overline{\mathrm{LF}} = \mathrm{DF} - k\,\mathrm{DF}\,L(z), \qquad \mathrm{SR} = k\,L(z).$$

The proof is one line: divide (9.1) by C and note that $\sigma/C = k\mu/C = k\,\mathrm{DF}$ and $(C-\mu)/\sigma = (1-\mathrm{DF})/(k\,\mathrm{DF})$. The consequence is that a single table, indexed by demand factor, serves every aircraft size. This is what Boeing's 1978 spill model packaged as printed **spill tables** (Boeing Commercial Airplane Company 1978), and what Table 9.2 reproduces for $k = 0.35$ from the formulas above.

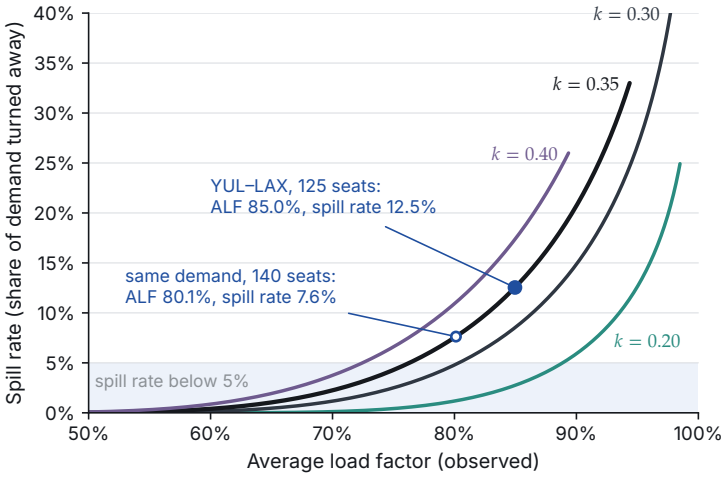


Figure 9.3 Spill rate as a function of the observed average load factor, for k-factors from 0.20 to 0.40. The shaded band marks spill rates below 5 percent. The two marked points are the YUL–LAX flight of Example 9.2 with 125 and with 140 seats.

Table 9.2 A spill table for normal demand with $k = 0.35$. Each row satisfies $\mathrm{DF} = \overline{\mathrm{LF}} + \mathrm{SF}$. The last column is $\Pr(D \geq C)$.

Demand factor	Average load factor	Spill factor	Spill rate	Departures full
0.70	0.687	0.013	1.9%	11%
0.75	0.726	0.024	3.2%	17%
0.80	0.761	0.039	4.9%	24%
0.85	0.792	0.058	6.9%	31%
0.90	0.818	0.082	9.1%	38%
0.95	0.841	0.109	11.5%	44%
1.00	0.860	0.140	14.0%	50%
1.05	0.877	0.173	16.5%	55%
1.10	0.891	0.209	19.0%	60%
1.20	0.914	0.286	23.9%	68%

The table answers two questions. Read left to right, it predicts the load factor and spill of a flight whose demand you know. Read from the load factor column back to the demand factor, it **unconstrains**: given an observed average load factor and an assumed k, it recovers mean demand. Figure 9.3 plots the second reading for several k-factors, and it is the picture every network planner should carry in their head.

The curves rise slowly and then explode. At an average load factor of 85 percent the spill rate is 2.8 percent if $k = 0.20$, 8.6 percent if

$k = 0.30$, 12.6 percent if $k = 0.35$ and 17.4 percent if $k = 0.40$. Belobaba notes the industry rule of thumb that spill rates much above 5 percent are "too high", since a lost-revenue share of that size is comparable to a typical airline operating margin (Belobaba 2006b). Read off the figure, the 5 percent line corresponds to an average load factor of 88.9 percent when $k = 0.20$ but only 76.3 percent when $k = 0.35$. A high average load factor is therefore a symptom, not a target: the same 85 percent can mean a well-sized flight in a stable business market and a badly undersized one in a volatile leisure market.

WORKED EXAMPLE 9.2 Five Fridays to Los Angeles

Return to Table 9.1. The mean load is 106.2 and the average load factor $106.2/125 = 85.0$ percent. The sample standard deviation of the loads is 18.6, an apparent k-factor of $18.6/106.2 = 0.175$. That number is not the k-factor of demand: two of the five observations are capped, which compresses the spread. Following the lecture, assume instead a typical $k = 0.35$ for total demand.

Unconstrain. Find the demand factor whose load factor is $106.2/125 = 0.8496$ in Result 9.3. Solving numerically gives DF = 0.971 (the lecture enters the printed Boeing table at the rounded 85.0 percent and reads 0.972; at exactly 85.0 percent the spill rate is the 12.6 percent quoted above). Hence

$$\mu = 0.971 \times 125 = 121.4, \qquad \sigma = 0.35 \times 121.4 = 42.5,$$

so each departure spills on average $\mathbb{E}[S] = 121.4 - 106.2 = 15.2$ passengers, a spill rate of $15.2/121.4 = 12.5$ percent.

Upgauge. With a 140-seat aircraft the demand distribution is unchanged (as long as passengers do not prefer one type over the other), so DF = $121.4/140 = 0.867$. Result 9.3 gives $\overline{\text{LF}} = 0.801$, an expected load of 112.2 (six more passengers per departure), spill of 9.3 and a spill rate of 7.6 percent. The lecture's table-based answers (load factor 0.802, spill 9.24, spill rate 7.6 percent) agree up to interpolation.

How much hangs on k. The value $k = 0.35$ is the lecture's assumption, not an estimate from these loads. Fitting the five loads directly with the censored normal model of Section 9.5, and treating the departures of 8 and 29 April as censored at 125, gives $\mu = 113.6$ and $\sigma = 26.4$: a k-factor of 0.23 and expected spill of 5.8 passengers per departure instead of 15.2. Five departures cannot discriminate between the two answers; a season of

loads on the same flight, or k-factors measured on comparable flights, can. Always report a spill estimate together with the k-factor it assumes.

The planner's question is now well posed: do the extra passengers per departure, at this route's net fares, pay for the extra trip cost of the 140-seat aircraft? Notice what the analysis did not do: it did not conclude "85 percent is high, so upgauge". It priced the answer.

9.5 Unconstraining: recovering demand from censored loads

The spill table needs an assumed k-factor. When you have a season of data you can estimate both μ and σ from the loads themselves, provided you treat the full departures correctly. Write the data as pairs (y_i, δ_i), $i = 1, \dots, n$, where $y_i = \min(d_i, C)$ is the observed load and $\delta_i = 1$ if departure i closed (so $d_i \geq C$ is all we know) and $\delta_i = 0$ otherwise. Three estimators are in common use.

Naive estimators. The sample mean and standard deviation of the y_i. Both are biased downward, and the bias grows with the share of closed departures. Never use them on constrained flights.

Censored maximum likelihood. A departure that did not close contributes its density, $\varphi((y_i - \mu)/\sigma)/\sigma$, to the likelihood; a departure that closed contributes the probability of closing, $1 - \Phi((C - \mu)/\sigma)$. The log-likelihood is

$$\ell(\mu, \sigma) = \sum_{i:\,\delta_i = 0} \left[\log \varphi\!\left(\frac{y_i - \mu}{\sigma} \right) - \log \sigma \right] + \sum_{i:\,\delta_i = 1} \log\!\left[1 - \Phi\!\left(\frac{C - \mu}{\sigma} \right) \right], \tag{9.3}$$

the Tobit model of econometrics with censoring from above. Maximize it numerically.

The EM algorithm. An elegant way to compute the same maximum is to alternate between guessing the missing demands and refitting (Dempster, Laird, and Rubin 1977). It needs two facts about a standard normal variable truncated from below at z: with the **inverse Mills ratio** $\lambda(z) = \varphi(z)/(1 - \Phi(z))$,

$$\mathbb{E}[Z \mid Z \geq z] = \lambda(z), \qquad \mathbb{E}[Z^2 \mid Z \geq z] = 1 + z\lambda(z).$$

Both follow from $\varphi'(t) = -t\,\varphi(t)$: integrate $t\,\varphi(t)$ directly and $t^2\varphi(t)$ by parts over $[z, \infty)$, then divide by $1 - \Phi(z)$.

RESULT 9.4 EM for censored normal demand

Start from any (μ, σ), for example the naive estimates, and repeat until convergence:

1. **E-step.** With $z = (C - \mu)/\sigma$, replace each closed departure by the conditional moments

$$m_1 = \mathbb{E}[D \mid D \geq C] = \mu + \sigma\,\lambda(z),$$
$$m_2 = \mathbb{E}[D^2 \mid D \geq C] = \mu^2 + \sigma^2 + \sigma\,(C + \mu)\,\lambda(z).$$

2. **M-step.** Average the completed data:

$$\mu \leftarrow \frac{1}{n}\Big(\sum_{\delta_i = 0} y_i + n_c\, m_1\Big),$$
$$\sigma^2 \leftarrow \frac{1}{n}\Big(\sum_{\delta_i = 0} y_i^2 + n_c\, m_2\Big) - \mu^2,$$

   where n_c is the number of closed departures.

No iteration decreases the log-likelihood (9.3), and the fixed point is the censored maximum likelihood estimate.

The second moment uses $\mathbb{E}[D^2 \mid D \geq C] = \mu^2 + 2\mu\sigma\lambda + \sigma^2(1 + z\lambda)$ and the identity $\sigma z = C - \mu$.

WORKED EXAMPLE 9.3 Unconstraining a season to Seville

Laurentian operates YUL–SVQ with a 182-seat A321XLR. Figure 9.4 shows a simulated peak season of 26 departures drawn from $\mathcal{N}(175, 42^2)$ and capped at 182. Eleven departures closed; the other fifteen loads average 141.1.

The naive estimates are $\hat{\mu} = 158.4$ and $\hat{\sigma} = 30.0$ (a standard deviation with divisor n, as in the M-step). Start EM there. The first E-step has $z = (182 - 158.4)/30.0 = 0.787$ and $\lambda(0.787) = 1.358$, so each closed departure is replaced by an expected demand of $158.4 + 30.0 \times 1.358 = 199.1$. Averaging gives $\mu = 165.6$ and $\sigma = 37.4$. The next two iterations give $(168.8, 41.3)$ and $(170.5, 43.4)$, and the algorithm settles at

$$\hat{\mu} = 172.7, \qquad \hat{\sigma} = 46.3,$$

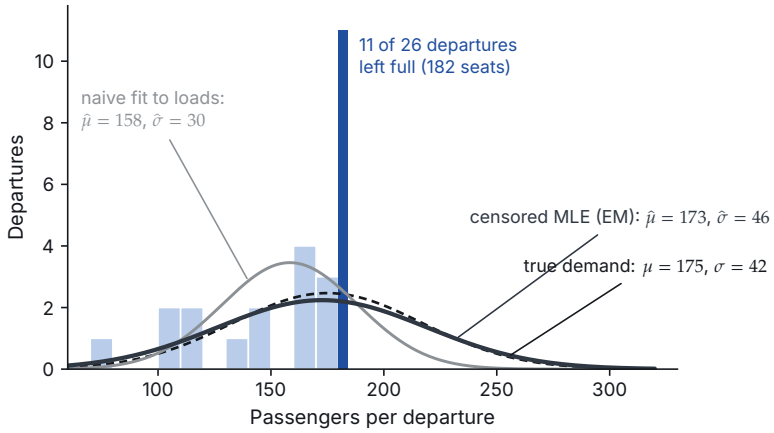


Figure 9.4 A simulated season of 26 departures of an A321XLR (182 seats) on YUL–SVQ, with true demand $\mathcal{N}(175, 42^2)$. Eleven departures closed. The naive fit to the loads (grey) badly understates both the level and the spread of demand; the censored maximum likelihood fit (solid dark curve) recovers the truth (dashed) up to sampling error.

the same values a direct numerical maximization of (9.3) returns.

The business consequence is large. Plugged into (9.1), the naive estimates imply 3.7 spilled passengers per departure; the censored estimates imply 14.2, a spill rate of 8.2 percent (the true model has 13.5). A route review based on the naive numbers would conclude that the XLR is comfortably sized for this season. It is not.

Sampling error is real: with 26 departures the standard error of $\hat{\mu}$ is roughly $46/\sqrt{26} \approx 9$ passengers, and more when many observations are censored. Pool comparable departures, and report intervals rather than points.

In practice the censoring is richer than a flight that closes. Revenue management closes fare classes one by one long before the flight fills, so low-fare demand is censored on many departures that leave with empty seats. Unconstraining is therefore done class by class and day by day along the booking curve, with methods such as the **booking profile** method, which extrapolates bookings after a closure from the booking curves of departures that stayed open, and projection detruncation, an EM-like iteration that uses a conditional median instead of the conditional mean (Weatherford and Pölt 2002). Errors

in this step feed back into the forecasts that set future availability, a mechanism behind the "spiral down" of Chapter 21.

9.6 What is a seat worth? Capacity under uncertainty

Suppose each carried passenger contributes r dollars (net fare minus passenger-variable cost) and each seat flown costs c dollars. If capacity could be chosen continuously, expected contribution would be

$$\pi(C) = r\,\mathbb{E}[\min(D, C)] - c\,C.$$

By Result 9.2 the derivative is $r\Pr(D > C) - c$, decreasing in C, so π is concave and the optimum sets the two terms equal.

RESULT 9.5 The newsvendor rule for capacity

The capacity that maximizes expected contribution satisfies

$$\Pr(D > C^*) = \frac{c}{r},$$

so that under normal demand

$$C^* = \mu + \sigma\,\Phi^{-1}\!\left(1 - \frac{c}{r}\right).$$

Add seats until the probability of filling the last one equals the ratio of its cost to its contribution.

The rule explains why airlines deliberately accept spill. Unless seats are free, the optimum always leaves some demand unserved, and the cheaper a seat is relative to a passenger's contribution, the more of the upper tail is worth covering.

Real gauge decisions are not continuous. Aircraft come in a few sizes, trip costs have large fixed components, and the comparison is between discrete options. The same machinery handles it: for aircraft type a with C_a seats and trip cost TC_a, expected contribution is

$$\pi_a(\mu) = r\Big[\mu - \sigma L\Big(\tfrac{C_a - \mu}{\sigma}\Big)\Big] - \mathrm{TC}_a,$$

and the best type is the one on the upper envelope of the π_a curves.

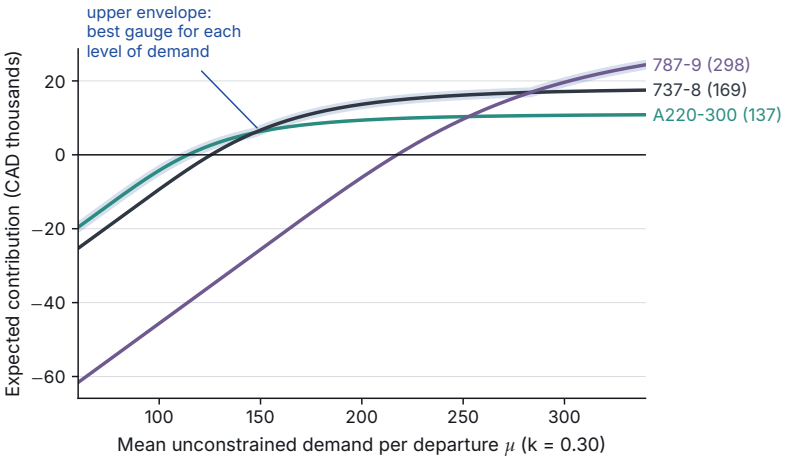


Figure 9.5 Expected contribution per departure on YUL–YVR for three Laurentian aircraft types as mean demand grows ($k = 0.30$, CAD 400 contribution per passenger, 5.2 block hours). The pale band marks the upper envelope: the best type for each level of demand.

WORKED EXAMPLE 9.4 Choosing the gauge for YUL–YVR

Laurentian's illustrative trip costs (block-hour and departure costs; passenger costs are netted out of r) for a 5.2-hour YUL–YVR sector are CAD 43,600 for the A220-300 (137 seats, CAD 318 per seat), CAD 49,300 for the 737-8 (169 seats, CAD 292 per seat) and CAD 85,600 for the 787-9 (298 seats, CAD 287 per seat). Each passenger contributes $r = 400$ and $k = 0.30$.

At $\mu = 150$ the A220 carries 124.8 passengers on average (load factor 91.1 percent) and contributes CAD 6,321; the 737-8 carries 140.0 (82.8 percent) and contributes CAD 6,688; the 787-9 loses CAD 25,602 at a 50 percent load factor. The 737-8 wins **with the lower load factor**: its extra 15.17 passengers are worth CAD 6,067, more than its CAD 5,700 extra trip cost. At $\mu = 190$ the gap widens: CAD 12,794 for the 737-8 against CAD 9,033 for the A220.

The curves in Figure 9.5 cross at mean demands of about 147 (A220 to 737-8) and 283 (737-8 to 787-9). The continuous newsvendor rule gives a cross-check. Moving from the A220 to the 737-8 costs $5{,}700/32 = 178$ dollars per extra seat, so the critical ratio is $1 - 178/400 = 0.555$ and $\Phi^{-1}(0.555) = 0.138$. At the crossover demand of 147 the ideal continuous capacity is

$147 + 0.138 \times 44 \approx 153$ seats, midway between the two narrow-bodies; at $\mu = 150$ it is $150 + 0.138 \times 45 \approx 156$, nearer the 737-8, which wins; both are far below the 787-9. Seats come in lumps, which is why fleet planners care about families of aircraft with closely spaced sizes.

A single route is not the end of the story. Aircraft are shared across the network, so putting the 737-8 on YUL–YVR means taking it from somewhere else. Fleet assignment (Chapter 14) solves the joint problem: it computes exactly these spill-adjusted contributions for every flight and every type, then chooses the assignment that maximizes total contribution subject to the number of aircraft available and the requirement that aircraft flows balance at every station.

9.7 Recapture

A spilled customer is not necessarily a lost customer. Some will take another Laurentian departure: an earlier flight the same day, the next day, or a connection over Toronto. The airline loses only the spill it fails to recapture.

DEFINITION 9.4 Recapture rate

The **recapture rate** b is the fraction of passengers spilled from an itinerary who travel instead on another itinerary of the same airline (or of a partner whose revenue counts, under the analysis's accounting boundary). The expected contribution lost to spill is then approximately $(1 - b)\, r\, \mathbb{E}[S]$.

The approximation hides a constraint: recaptured passengers need seats. They can only be recaptured onto departures with room, and when they take a seat they may in turn displace someone else. Recapture is therefore itself a capacity-constrained flow, which is why the itinerary-based fleet assignment model of Barnhart, Kniker, and Lohatepanont (2002) treats it explicitly, with a recapture rate b_{pr} from each itinerary p to each alternative r and seat constraints on both.

WORKED EXAMPLE 9.5 Recapture needs empty seats

Laurentian flies YUL–YVR twice a day with A220s (137 seats). The evening departure is the popular one, with demand $\mathcal{N}(160, 48^2)$; the morning departure has demand $\mathcal{N}(100, 30^2)$. Assume, for illustration, that 30 percent of spilled evening customers would accept the next morning's flight and that the two demands are independent.

The evening flight spills $\mathbb{E}[S] = 48\,L((137-160)/48) = 32.8$ customers per day. If seats were never a problem, $0.30 \times 32.8 = 9.8$ of them would be recaptured. The morning flight has 38.6 empty seats on average, but not every day; simulating both flights shows that 7.9 customers are actually recaptured on average, because some of the evening's worst days coincide with busy mornings. Net loss: $32.8 - 7.9 = 24.9$ customers, about CAD 10,000 of contribution per day at CAD 400 each. Ignoring the morning flight's capacity would understate the loss by about two passengers a day.

The general lesson is that **spill cost is a network quantity**. Whether a spilled passenger is lost depends on the alternatives the network offers and on how full they are, and adding capacity on one flight can pull passengers off another. This is the main reason network planners prefer itinerary-level models to leg-level arithmetic (Chapters 12 and 14).

9.8 Spill when revenue management controls the flight

The classical spill model assumes that bookings are accepted on a **first-come, first-served** basis until the aircraft is full, so the spilled customers are a random sample of all customers and each one is worth the average fare. No modern airline sells seats that way. Revenue management (Chapter 18) deliberately turns away early low-fare requests to protect seats for later, higher-fare customers. Three consequences matter for planning.

Spill is concentrated at the bottom of the fare ladder. The customers turned away are mostly those asking for the lowest fares, so the revenue lost per spilled passenger is well below the average fare.

The marginal seat is sold at the lowest open fare. On a flight managed by revenue management, one more seat does not let in an average customer. It raises the booking limit of the cheapest open class by one, so it earns roughly the lowest fare times the probabil-

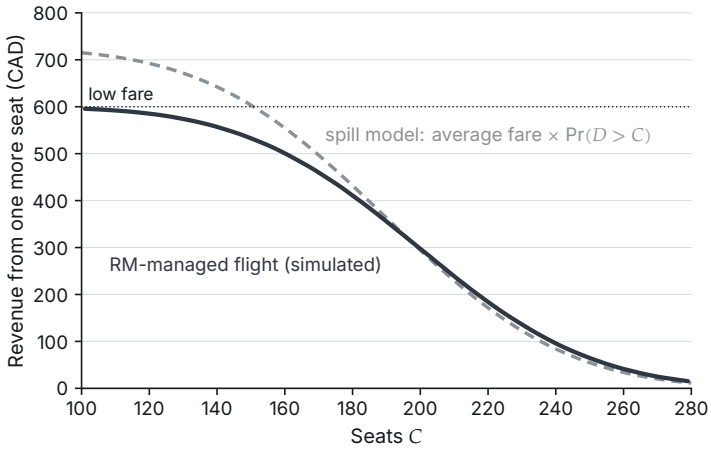


Figure 9.6 The value of one more seat on a two-fare flight. The dashed curve is the classical spill model: average fare times the probability of filling the seat. The solid curve simulates a flight controlled by revenue management, which protects 40 seats for the high fare; its marginal seat is mostly a low-fare seat.

ity that low-fare demand would have used it. Figure 9.6 compares this with the average-fare spill model on a two-class flight: a flexible fare of CAD 1,200 with demand $\mathcal{N}(40, 12^2)$ and a restricted fare of CAD 600 with demand $\mathcal{N}(150, 40^2)$. The spill model values the 150th seat at CAD 604 (the average fare of 726 times the probability of filling it); the managed flight earns about CAD 533 from it, CAD 505 from low-fare demand and the rest from high-fare customers on departures where low-fare demand falls short of its booking limit. The spill model overstates the value of the marginal seat wherever capacity is tight, which is exactly where gauge decisions are made; a little above mean demand (190 passengers) the two curves cross, and both values then fall towards zero.

More capacity can change who buys what. When an aircraft grows, revenue management opens cheaper classes more often, and some customers who would have paid more **buy down**. Simulation studies at MIT (Abramovich 2013) show that classical spill estimates can misstate the revenue value of capacity on RM-controlled networks, and identify settings in which the marginal revenue of additional seats is far lower than the spill model suggests, even negative in specific simulated scenarios. Treat this as a warning against mechanical capacity arithmetic, not as a claim that capacity normally destroys revenue.

The practical conclusion for a network planner is to value capacity with the same quantities revenue management uses: displace-

ment costs and **bid prices**, the opportunity cost of a seat on each leg (Chapter 20). A bid price is the RM system's estimate of the marginal value of capacity at the current state of the network, and it already reflects fare mix and connecting flows (and recapture, when the demand model includes customer choice). Chapter 22 shows how to use RM outputs to price a gauge change.

IN PRACTICE How spill is estimated today

In 2026 the logic of this chapter is unchanged, but it is embedded in software rather than spill tables. Network planning systems simulate origin-destination demand, itinerary choice, spill and recapture across the whole schedule; Lufthansa Systems, for example, documents O&D simulation with market-share and spill and recapture functions in its NetLine Plan product (Network Planning Handbook 2026; Lufthansa Systems 2026). Revenue management systems unconstrain demand at the level of fare product and day before departure, increasingly with forecasting engines that vendors describe as Bayesian and AI-based (PROS Holdings 2021). Where airlines price continuously instead of opening and closing fare classes (Szymański, Belobaba, and Papen 2021), demand is censored by price rather than by availability, and unconstraining becomes the estimation of a demand curve (Chapters 7 and 21). Your job is less to compute $L(z)$ by hand than to know which assumptions the tool is making, and to sanity-check its output with the formulas above.

9.9 Summary

- Airlines observe **loads**, which are demand capped at capacity. Any estimate of demand on a constrained flight is a model of censored data.
- Under normal demand, expected spill is $\sigma L(z)$ with $z = (C - \mu)/\sigma$ and L the standard normal loss function. Expected load is $\mu - \sigma L(z)$, always below $\min(C, \mu)$.
- The C-th seat is filled with probability $\Pr(D \geq C)$, so expected load is concave in capacity and extra seats carry far fewer passengers than their number.
- Spill ratios depend only on the demand factor and the k-factor. High average load factors imply rapidly rising spill rates, faster in volatile markets.

- Unconstrain with censored maximum likelihood or EM, never with averages of loads. Revenue management censors fare classes long before flights fill.
- Optimal capacity balances $r\Pr(D > C)$ against the cost of a seat. With discrete aircraft, compare spill-adjusted contributions; the best aircraft often runs at the lower load factor.
- Recapture reduces the cost of spill only where alternatives have seats, which makes spill cost a network quantity.
- On RM-controlled flights the marginal seat sells at the lowest open fare. Value capacity with bid prices and displacement costs rather than average fares.

Table 9.3 Key formulas of the chapter.

Quantity	Formula
Standardized capacity	$z = (C - \mu)/\sigma$
Loss function	$L(z) = \varphi(z) - z\,(1 - \Phi(z))$
Expected spill	$\mathbb{E}[S] = \sigma L(z)$
Expected load	$\mathbb{E}[Y] = \mu - \sigma L(z) = \int_0^C \Pr(D > x)\,\mathrm{d}x$
Marginal seat	$\mathrm{d}\mathbb{E}[Y]/\,\mathrm{d}C = 1 - \Phi(z)$
Spill ratios	$\mathrm{SF} = k\,\mathrm{DF}\,L(z)$, $\overline{\mathrm{LF}} = \mathrm{DF} - \mathrm{SF}$, $\mathrm{SR} = k\,L(z)$
Newsvendor capacity	$C^* = \mu + \sigma\,\Phi^{-1}(1 - c/r)$
EM conditional mean	$\mathbb{E}[D \mid D \geq C] = \mu + \sigma\,\lambda(z)$, $\lambda = \varphi/(1 - \Phi)$

9.10 Further reading

Belobaba's lecture note (Belobaba 2006b) is short and worth reading in full for its definitions and the spill table procedure; it is the source of Example 9.2. Swan (2002) examines which distributions fit airline demand, for revenue management and for spill. Weatherford and Pölt (2002) compare unconstraining methods on airline data, and Talluri and van Ryzin (2004b, ch. 9) cover estimation and forecasting for RM with censored data. Barnhart, Kniker and Lohatepanont (2002) introduce recapture into fleet assignment, and Abramovich (2013) shows by simulation why spill estimates must account for revenue management.

PART III

Designing the Network

This part is the core of the network planner's job: deciding which markets to serve, with which aircraft, how often and when, and checking that the resulting plan can be flown. It starts with the shape of the network and the aircraft that fly it, values a single route against the rest of the network, and then works down through schedule design, fleet assignment, aircraft routing and crew to the robustness of the plan on the day of operation.

CHAPTER

10

Network Structure and Connectivity

IN THIS CHAPTER

- Why most network airlines route passengers through hubs, and what that choice costs them.
- The combinatorics of connectivity: how n routes become $n(n+1)/2$ markets, and what one more spoke adds.
- Economies of density, scope and scale, a threshold for when a market deserves its own nonstop, and the econometric evidence.
- Hub banks, connecting and sixth-freedom traffic, and why a hub can serve markets far too thin for a nonstop.
- A worked count of what Seville adds to Laurentian's Montreal hub, and what its feed does to the feeder flights.
- Interline, codeshare, alliances and joint ventures; freedoms of the air, bilateral agreements and Canada's Blue Sky policy.
- Canadian market structure in the Competition Bureau's 2025 study, and why hub carriers earn fare premiums.

DRAWS ON Belobaba, Airline economics review, Airline schedule development, ePODS airline management game and Route planning and evaluation (MIT 16.75J, 2006); handbook; Competition Bureau, Cleared for take-off (2025); Air Canada Q2 2026 MD&A

A network planner rarely decides about one route in isolation. When Laurentian adds a flight from Montreal to Seville, it does not only open the Montreal–Seville market: it opens Quebec City–Seville, Boston–Seville, Toronto–Seville and every other market whose travellers can reach Montreal in time for the evening departure. It also puts more passengers on the feeder flights that bring them, and some of those flights were already full. The value of a route therefore depends on the network around it, and the shape of

that network, which cities connect to which and through where, is the first thing a planner has to understand.

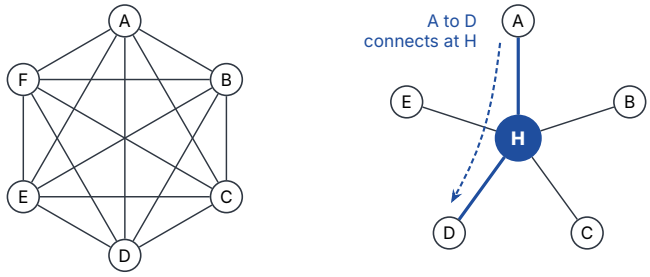
This chapter explains why networks look the way they do. It starts with the two archetypes, point to point and hub and spoke, counts the markets each structure connects, and derives the cost advantage of routing traffic through a hub from economies of density. It then turns to the traffic that hubs create (connecting, beyond and sixth-freedom flows), to the agreements that let airlines extend their networks through partners, and to the traffic rights that govern where they may fly at all. It closes with the structure of the Canadian market as the Competition Bureau described it in 2025, and with the fare premiums that hub carriers earn.

The chapter builds on the distinction between legs and origin-destination markets from Chapter 2 and on the demand concepts of Chapter 4. Its counts and thresholds are the raw material of Chapter 12, which turns network effects into dollars, and of Chapter 13, which designs the banks that make connections happen.

10.1 Point to point and hub and spoke

Recall the vocabulary of Chapter 2. A **flight leg** is one nonstop operation between two airports; an **origin-destination (O&D) market** is the set of travellers who want to go from one city to another; an **itinerary** is a sequence of legs that serves a market. Airline demand is generated by O&D markets, while supply comes as legs. One leg supplies seats jointly to many O&D markets, and one market is served by many itineraries, often of several airlines. Belobaba calls this the dichotomy of airline demand and supply, and it is why the profitability of a single market, or of a single leg, cannot be measured without allocation conventions (Belobaba 2006d).

How an airline chooses to link its cities determines how many markets each leg serves.



Point to point
15 routes serve 15 markets

Hub and spoke
5 routes serve 15 markets:
5 local and 10 connecting

Figure 10.1 Six cities served point to point need fifteen nonstop routes to connect every pair. Served through one of them as a hub, five routes connect the same fifteen markets: five local markets to and from the hub, and ten spoke-to-spoke markets whose travellers connect at H.

DEFINITION 10.1 Network structures

A **point-to-point network** serves each O&D market it sells with a nonstop flight between the two cities; passengers do not connect. A **hub-and-spoke network** concentrates flights at one or a few **hubs**: most routes link a hub to an outlying **spoke** city, and travellers between two spokes change planes at the hub. A **gateway** is a hub where domestic and regional flights connect to long-haul international flights. Real networks mix the two: hub carriers fly some nonstops that bypass their hubs, and point-to-point carriers carry some self-connecting passengers.

Figure 10.1 shows the trade-off with six cities. Through a hub, five routes cover all fifteen markets. Each spoke flight carries a mix of travellers: those going to the hub city itself and those continuing to four other spokes. Belobaba's lecture on schedule development names this **load consolidation**: one flight serving several O&D markets at the same time lets the airline offer higher frequency, larger aircraft, or both, and he calls it a fundamental reason for the economic success of hubs (Belobaba 2006f).

The price of consolidation is paid by the passengers who connect. Their trip is longer, it includes a change of aircraft that can be missed, and it competes with any nonstop offered by another airline. Belobaba states the economics in one comparison: routing passengers through a hub is more profitable when the cost saved by flying fewer,

larger and fuller flights exceeds the revenue lost from travellers who reject the connection in favour of a competitor's nonstop (Belobaba 2006i). The rest of this chapter makes both sides of that comparison precise. The cost side is economies of density (Section 10.3); the revenue side is the passenger's valuation of time and convenience, which Chapters 4 and 5 model.

A hub also changes what the passenger sees. Belobaba's total trip time model (Chapter 4) adds to the flight time a schedule displacement term that falls with frequency, so a connecting market with a dozen daily one-stop options can offer a lower total trip time than a single daily nonstop (Belobaba 2006d). This is why large hub networks compete in markets that also have nonstops.

10.2 The combinatorics of connectivity

The arithmetic of Figure 10.1 generalizes, and the general version is the most important number in network planning.

RESULT 10.1 Markets connected by one hub

A hub with n spokes, each linked to the hub by one route, connects

$$M(n) = \underbrace{n}_{\text{local}} + \underbrace{\binom{n}{2}}_{\text{connecting}} = \frac{n(n+1)}{2} \tag{10.1}$$

undirected O&D markets with n routes, or $n(n+1)$ directional markets. A point-to-point network that serves the same $n+1$ cities with nonstops needs $n(n+1)/2$ routes. Adding spoke $n+1$ adds $M(n+1) - M(n) = n+1$ markets: one local market and n connecting ones.

Proof. The network has $n+1$ cities, so there are $\binom{n+1}{2} = n(n+1)/2$ unordered city pairs. Each pair either contains the hub, in which case it is one of the n local markets served nonstop, or consists of two spokes, in which case it is served by the one-stop itinerary through the hub; there are $\binom{n}{2} = n(n-1)/2$ such pairs. A point-to-point network must fly one route per pair. The increment follows from $M(n+1) - M(n) = \frac{(n+1)(n+2) - n(n+1)}{2} = n+1$. ∎

The ratio of markets to routes in (10.1) is $(n+1)/2$, so leverage grows linearly with the size of the hub, and the marginal value of a

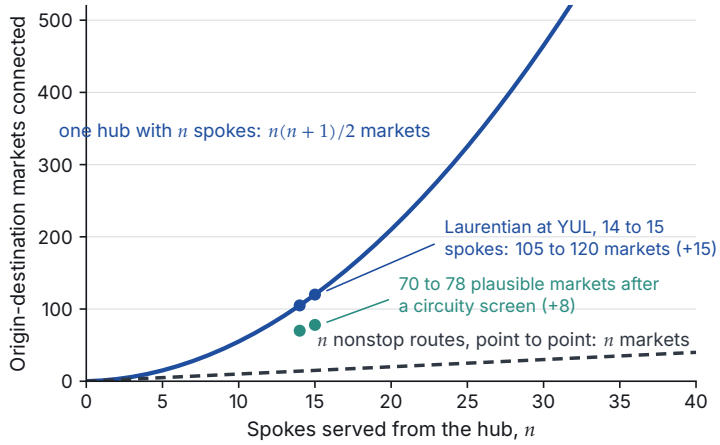


Figure 10.2 O&D markets connected by a single hub as spokes are added, against the markets served by the same number of point-to-point routes. The points mark Laurentian's illustrative YUL network before and after it adds Seville: 105 to 120 markets by the count of Result 10.1, and 70 to 78 after removing connections that backtrack (Example 10.2).

spoke grows with the number of spokes already there. A hub with 30 spokes connects 465 markets with 30 routes, and its 31st spoke brings 31 new markets. This increasing return to network size on the revenue side is why hub carriers defend their hubs and why a new spoke at a large hub is far easier to justify than the same route flown by an airline without one (Belobaba 2006i). Figure 10.2 draws the quadratic.

A hub turns n routes into $n(n+1)/2$ markets. That leverage, not the profit of any single route, is what a network planner manages.

With more than one hub the count depends on how the hubs are linked. Take two hubs, each with its own n_1 and n_2 spokes, joined by one route. Markets within each hub's catchment need at most one connection, and a spoke of hub 1 reaches hub 2 with one connection; but the $n_1 n_2$ markets between a spoke of hub 1 and a spoke of hub 2 need two connections, with long elapsed times and two chances to misconnect. This is why multi-hub airlines connect their larger spokes to more than one hub, and why alliances connect hubs of different airlines so that a traveller from a spoke of one carrier reaches a spoke of the other with a single change (Section 10.7).

Counting pairs gives an upper bound, not a forecast. The handbook warns against counting every geometrically possible connection as commercial value: a connection matters only if a legal, sellable itinerary exists with enough transfer time, a tolerable elapsed time and actual demand (Network Planning Handbook 2026). Two screens remove most of the illusion. The first is timing: an itinerary exists only if an arrival and a departure fall in the same connecting window, which is what banks are for (Section 10.4). The second is geometry. Define the **circuity** of a connection from i to j through hub H as the ratio of the distance flown to the direct distance,

$$\rho_{ij} = \frac{d_{iH} + d_{Hj}}{d_{ij}}.$$

A connection with ρ_{ij} close to 1 is a natural path; one with $\rho_{ij} = 1.8$ asks the traveller to fly almost twice the distance and will sell only at a deep discount, if at all. Where to set the cut-off is an analyst's choice, not an industry standard; this chapter uses 1.25.

> **PITFALL**
>
> Counting geometric markets as commercial ones. A hub with 40 spokes does not serve 820 markets. Between two spokes on the same side of the hub (Ottawa to Toronto through Montreal has a circuity of 1.8), between two spokes whose banks do not meet, and between cities whose travellers have a better non-

stop, the count is zero. Report connectivity after the timing and geometry screens, and say which screens you used.

10.3 Economies of density, scope and scale

Why do fuller flights and bigger hubs lower costs? Most of an airline's costs are incurred per flight, per departure or per block hour, whether the seats are full or not (crew, landing and navigation charges, ground handling, most of the fuel), and Chapter 3 shows that larger aircraft have a lower cost per seat. Carrying more passengers on the same routes therefore lowers the cost per passenger in two ways: the per-departure costs are shared by more travellers, and the airline can fly larger aircraft. Economists separate this effect of traffic density from the effect of network size.

DEFINITION 10.2 Economies of density, scale and scope

Let $C(Q, N)$ be an airline's cost of carrying traffic Q (for example, revenue passenger-kilometres) over a network of N points served, and write $\epsilon_{CQ} = \partial \ln C / \partial \ln Q$ and $\epsilon_{CN} = \partial \ln C / \partial \ln N$ for the cost elasticities.

- **Returns to density** are RTD $= 1/\epsilon_{CQ}$: the response of cost to more traffic on an unchanged network. RTD > 1 means economies of density.
- **Returns to scale** are RTS $= 1/(\epsilon_{CQ} + \epsilon_{CN})$: the response of cost when traffic and the number of points served grow together, so that density is unchanged.
- **Economies of scope** exist between two outputs, here two O&D markets, when producing them together costs less than producing them separately: $C(q_1, q_2) < C(q_1, 0) + C(0, q_2)$ (Panzar and Willig 1981).

The empirical benchmark is Caves, Christensen and Tretheway (1984), who estimated a cost function for US trunk and local service carriers over 1970–1981. Their estimates show substantial returns to density, about 1.24 on average, but approximately constant returns to scale: an airline that carries more passengers on the same network lowers its unit cost, while an airline that simply serves more cities at the same density does not. The cost difference between large and small carriers came from density, not size. Berry, Carnall and Spiller

(1996), estimating demand and costs jointly, found economies of density on longer routes but not on shorter ones. These are historical US estimates, and their magnitudes should not be transferred to a 2026 network; the mechanism is durable, and it explains what a hub is for. A hub is a device for buying density. By consolidating many thin O&D markets onto a few spoke routes, it raises the traffic on each leg without serving fewer markets.

Economies of scope follow from the same arithmetic. In a hub network every spoke flight produces a bundle of O&D markets jointly, and the departure cost it incurs is shared by all of them; producing Quebec City–Paris and Quebec City–Seville separately would require two sets of departures. The network cost function is subadditive in its markets because markets share legs.

10.3.1 When does a market deserve its own nonstop?

The cleanest way to see density at work is to ask when a market should leave the hub. Suppose the trip cost of a flight on a given sector is approximately linear in the number of seats,

$$\mathrm{TC}(s) = a + c\,s,$$

where a is the cost per departure that does not depend on aircraft size and c is the cost per seat. A market with Q travellers per day in each direction, flown nonstop once a day with an aircraft sized to carry them at load factor LF, costs $a + c\,Q/\mathrm{LF}$ per day and direction, plus a passenger-related cost c_{pax} per traveller. Write $\kappa = c/\mathrm{LF} + c_{\mathrm{pax}}$ for the cost per passenger carried, so the nonstop costs $a + \kappa Q$. Routed through the hub, the same travellers ride existing spoke flights. If those flights have room (or are upgauged to make room), they add no departures, only seats and passengers on two legs.

RESULT 10.2 The hub-versus-nonstop threshold

Let a nonstop in market m cost $a_m + \kappa_m Q$ per day and direction, and let routing the market's Q travellers over two hub legs cost $(\kappa_1 + \kappa_2)\, Q$. The hub routing is cheaper if and only if

$$Q < Q^* = \frac{a_m}{\kappa_1 + \kappa_2 - \kappa_m}. \tag{10.2}$$

Between two spokes with similar sectors, $\kappa_1 \approx \kappa_2 \approx \kappa_m = \kappa$ and $Q^* = a/\kappa$. Behind a gateway, where the long-haul seat is needed either way, $\kappa_2 = \kappa_m$ and $Q^* = a_m/\kappa_1$: the nonstop must pay a whole long-haul departure to save one short feeder seat. If a share λ of the market's travellers would fly a nonstop but not a connection, each worth contribution r, the threshold falls to $a_m/(\kappa_1 + \kappa_2 - \kappa_m + \lambda r)$.

Proof. The hub routing is cheaper when $(\kappa_1 + \kappa_2)Q < a_m + \kappa_m Q$, that is when $(\kappa_1 + \kappa_2 - \kappa_m)\, Q < a_m$, which gives (10.2) because the coefficient on Q is positive: two legs cost more per passenger than one. Counting the contribution lost from travellers who refuse to connect adds $\lambda r Q$ to the cost of the hub routing, which is Belobaba's comparison of cost savings with revenue loss, and the same algebra gives the second threshold. ∎

The result is stylized: aircraft come in discrete sizes, a daily frequency may be commercially too little, spill (Chapter 9) makes the nonstop's load uncertain, and the feeder seats are free only when the feeders have room. Its value is that it identifies the quantity that matters, which is not whether a market has enough passengers to fill an aircraft but whether it has enough to pay for an additional departure.

WORKED EXAMPLE 10.1 Two thresholds on Laurentian's network

Fit the linear trip cost to two Laurentian gauges with the case's cost model. On a 1.5-hour transborder sector, an A220-300 (137 seats) costs CAD 14,000 per trip before passenger costs and a 737-8 (169 seats) CAD 16,000. The slope is $c = 2{,}000/32 = 62.5$ dollars per seat and the intercept $a = 14{,}000 - 62.5 \times 137 = 5{,}437.5$ dollars per departure. With LF $= 0.80$ and a transborder passenger cost of CAD 30, $\kappa = 62.5/0.80 + 30 = 108.1$.

Between two spokes, such as Quebec City and Boston, Result 10.2 gives $Q^* = 5{,}437.5/108.1 = 50$ travellers a day each way. Below that, two seats on existing Laurentian flights cost less than a new daily departure. If 20 percent of the market would not accept a connection and each such traveller contributes CAD 230, the threshold falls to $5{,}437.5/(108.1 + 46) = 35$.

Behind the gateway, compare a Boston–Seville nonstop with a connection over Montreal. On a 7-hour sector the A321XLR (182 seats) costs CAD 80,500 per trip and the 787-9 (298 seats) CAD 115,500, so $c = 35{,}000/116 = 301.7$ and $a = 80{,}500 - 301.7 \times 182 = 25{,}586$. The connection needs the same long-haul seat plus a feeder seat on a 1.4-hour A220 sector, where the same fit gives $c = 59.4$ and $\kappa_1 = 59.4/0.80 + 30 = 104.2$. The threshold is $25{,}586/104.2 = 246$ travellers a day each way.

Laurentian's connecting markets through its evening bank (Example 10.2) carry 20 (Quebec City–Seville), 29 (Boston–Seville), 48 (Quebec City–Paris) and 54 (Boston–Paris) travellers per departure each way, a day's traffic if the bank operates daily: between a twelfth and about a fifth of the long-haul threshold. The Seville flight operates four times a week, so per day its markets are thinner still. On cost alone none of them could support a nonstop. All the numbers are illustrative, and the fitted line overstates the precision of comparing two different aircraft types; the order of magnitude is the lesson.

Figure 10.3 explains a pattern that every planner sees in the data. Long-haul nonstops from secondary cities appear only when the market is large or the aircraft is small: the A321XLR lowers the long-haul departure cost a_m, which is exactly what moves the threshold. Cirium's 2026 review of A321XLR deployments describes the aircraft as opening markets too thin for widebodies, at the price of less belly cargo and a cabin some passengers like less than a widebody's (Cirium 2026b). In the language of Result 10.2, a smaller long-haul aircraft lowers Q^* and lets some markets leave the hub; Chapter 11 returns to the aircraft itself.

10.4 Hub banks: how connections are made

A connection requires an arrival and a departure at the hub separated by at least the time needed to transfer the passenger and the

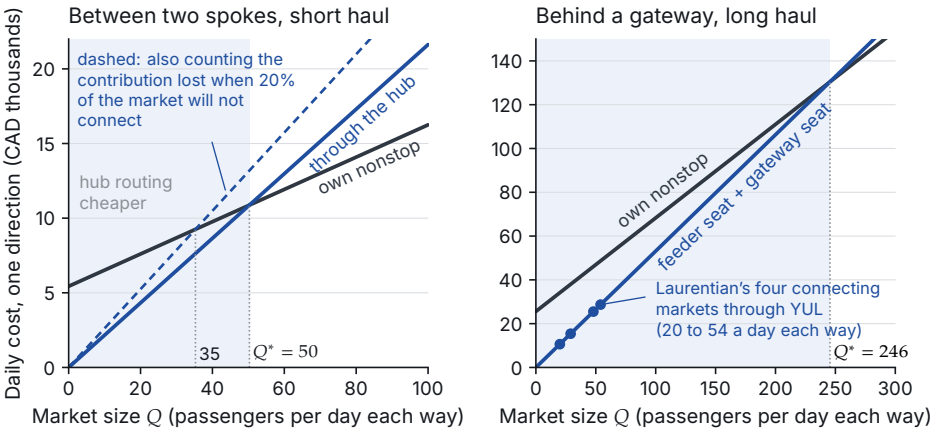


Figure 10.3 The daily cost of serving one market by its own nonstop and through the hub, as the market grows. Left: between two spokes on short sectors, the hub routing is cheaper below about 50 travellers a day each way, or 35 once the contribution lost to travellers who refuse to connect is counted. Right: behind a long-haul gateway the nonstop must pay a whole long-haul departure to save a feeder seat, and the threshold rises to about 246; Laurentian's four connecting markets sit far below it.

bags, and by not so much that the trip becomes unattractive. The handbook writes the screen directly (Network Planning Handbook 2026): arrival i connects to departure j if

$$\text{MCT}_{ij} \;\le\; t_j^{\text{dep}} - t_i^{\text{arr}} \;\le\; T_{\max}, \tag{10.3}$$

where MCT_{ij} is the applicable **minimum connecting time** for that pair of flights (it depends on the terminals and on whether the passenger crosses a border) and $T_{\max}$ is the longest connection the analyst considers sellable. Timestamps must be full dates and times in a common time base, not clock faces, because transatlantic flights cross midnight and time zones (Chapter 13). The upper limit is a commercial assumption, not an industry rule.

A **bank** (or connecting complex, or wave) is the schedule structure that makes (10.3) hold for many pairs at once: a group of arrivals scheduled close together, followed after at least the MCT by a group of departures. If n_a flights arrive in a bank and n_d depart, the bank creates up to $n_a n_d$ connections, less those that would return a passenger to the city they came from. If every spoke has one arrival and one departure in each of k banks a day, each directional spoke-to-spoke market gets up to k one-stop itineraries a day. Banks can be directional: the hubs of the ePODS teaching network, for instance, have two eastbound and two westbound connecting banks a day (Belobaba 2006h), which is how a hub between eastern and

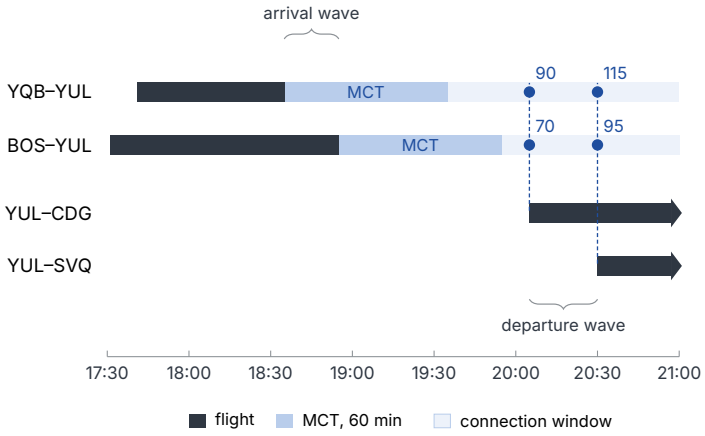


Figure 10.4 Four legs of Laurentian's evening bank at YUL. The Quebec City and Boston flights arrive at 18:35 and 18:55; each opens a connection window once the 60-minute MCT has passed. The Paris departure at 20:05 and the Seville departure at 20:30 fall inside both windows, so the bank creates four connections, with connecting times of 70 to 115 minutes. Times are illustrative except the 20:30 Seville departure, which mirrors the public schedule; the MCT is an assumption, not a YUL rule.

western spokes serves east-west traffic without wasting connections between two eastern cities.

Figure 10.4 shows the part of Laurentian's evening bank that the case data describe: two feeders and two transatlantic departures, four connections. The two feeders do not connect to each other, because both arrive; Quebec City–Boston passengers need a different bank. The shape of the bank is a trade-off. Tight banks shorten connections and multiply them, but they concentrate the demand for gates, security, baggage systems, crews and runways into short peaks; spreading the flights out eases the peaks and may lengthen or remove connections (Network Planning Handbook 2026). Bank design, connection windows and MCT are developed in Chapter 13.

10.5 Connecting traffic, the sixth freedom and thin markets

A hub changes who is on board. The flight from Montreal to Seville carries travellers from Montreal, but also from Quebec City and Boston, and in the other direction travellers continuing from Montreal to cities across North America. Planners describe this mix with a few standard categories.

DEFINITION 10.3 Local, connecting and sixth-freedom traffic

On a leg from hub H to destination B:

- **local** traffic travels only between H and B;
- **behind** (or feed) traffic originates at a city behind the hub and connects at H onto the leg, such as Quebec City to Seville over Montreal;
- **beyond** traffic continues past B on another flight;
- **flow** traffic is all traffic that is not local.

Sixth-freedom traffic is the part of connecting traffic that travels between two foreign countries through the airline's home country, such as Boston to Paris over Montreal on a Canadian airline. Quebec City to Paris over Montreal is connecting traffic but not sixth-freedom traffic.

Connecting travellers have three economic properties that recur throughout this book. First, when the legs they use have empty seats, they cost almost nothing to carry: a passenger-related cost on each leg and a share of the connection's handling. Second, they pay prices set in their own O&D market, where they can usually choose a competitor's nonstop or another airline's hub. Their fares therefore need not exceed the local fare of the long-haul leg by much: in Laurentian's illustrative fare table, the restricted Boston–Paris fare is CAD 599 while the restricted Montreal–Paris fare is CAD 549, although the Boston traveller uses an extra flight. Third, when the legs are full, a connecting passenger takes a seat that a local passenger could have bought, and the network gains only the difference. Valuing that displacement is the subject of Chapters 12 and 20.

Belobaba describes the planning consequence vividly: in an established hub, a new route to a small spoke can be justified with one or two connecting passengers per flight to each of 30 or more connecting destinations (Belobaba 2006i). He adds two warnings. The same incremental logic tends to overlook the displacement of other traffic on the connecting legs, and it makes it hard to stop flying an unprofitable spoke that feeds the rest of the network. Both warnings return in Chapter 12.

This is also the deepest reason a hub can serve markets that are far too thin for a nonstop. By Example 10.1, a market behind a long-haul gateway needs something like 250 travellers a day each way before a nonstop beats the connection on cost. Most city pairs never come

close. Through a hub, a market of five travellers a day is served at the cost of five feeder seats, as long as the feeder has room.

IN PRACTICE **Hubs, sixth freedom and partners at Air Canada in 2026**

Air Canada describes Toronto Pearson, Montreal-Trudeau and Vancouver as its international hubs (Air Canada 2025). Its second-quarter 2026 report attributes 9 percent year-on-year growth in sixth-freedom revenue to the US–Atlantic and US–Pacific corridors, mainly traffic from the United States to Europe, and describes regional flying by Jazz and PAL under the Air Canada Express brand as providing traffic feed to mainline and Rouge routes (Air Canada 2026e). At its December 2024 investor day the company set a goal of 2 percent of the market between the United States and international destinations, and said that 80 percent of that market could connect efficiently over Toronto, Vancouver or Montreal (Skies Magazine 2024). The Seville service announced on 14 October 2026 is described as timed to connect through Montreal, and the release notes that US customs preclearance in Montreal lets travellers continuing to the United States arrive as domestic passengers (Air Canada 2026g). Air Canada also says that its partnership with United brings about two million customers a year who connect to or from the United States onto its domestic network (Air Canada 2026h). These are company statements about the network as a whole, not evidence about the profitability of any route.

PITFALL

Calling every connection sixth-freedom traffic. The handbook lists this among the common errors (Network Planning Handbook 2026). Domestic-to-international connections (Quebec City–Paris) and international-to-domestic ones (Boston–Vancouver) depend on different traffic rights and different competitors from true sixth-freedom flows (Boston–Paris), and they respond differently to a change in the long-haul schedule.

10.6 What one more spoke adds: Seville at YUL

The worked example of this section puts the chapter's tools together on the case that runs through the book.

WORKED EXAMPLE 10.2 Adding Seville to Laurentian's Montreal hub

Counting markets. Suppose, for illustration, that Laurentian serves from YUL the fifteen other airports of the running case: Quebec City, Ottawa, Halifax, Toronto and Vancouver in Canada; Boston, New York LaGuardia and Chicago in the United States; and Paris, London, Lisbon, Seville, Madrid, Barcelona and Rome in Europe. Without Seville it has 14 spokes and, by Result 10.1, $14 \times 15/2 = 105$ markets: 14 local and 91 connecting. Adding Seville makes 120, so the new spoke adds 15 markets, one local and 14 connecting.

Most of the 14 are not markets. Computing the circuity of every connection from the airport coordinates removes the six European spokes at once: Lisbon to Seville over Montreal flies 33.6 times the direct distance, and even London to Seville flies 6.6 times. Halifax fails the screen too, with a circuity of 1.33, because Halifax lies east of Montreal and a Halifax traveller to Spain would fly west first. What remains are seven North American markets behind the gateway, Quebec City, Ottawa, Toronto, Vancouver, Boston, LaGuardia and Chicago to Seville, with circuities between 1.00 and 1.10. Seville adds 8 plausible markets, not 15. The same screen applied to the whole network leaves 70 plausible markets before Seville and 78 after (Figure 10.2).

Estimating feed. The market model of Example 4.4 gives the nonstop about 82 local Montreal–Seville travellers per departure, so the case's shoulder-season mean demand of 150 (the level that Chapter 12 applies to Seville's summer) includes about 68 connecting travellers: feed is 45 percent of the flight. The timing screen decides where they can come from, because only spokes with a flight arriving in the evening bank feed the 20:30 departure. The four-leg bank of the case data (Figure 10.4) includes two of the seven feeders, with a mean unconstrained demand per departure of 20 travellers from Quebec City and 29 from Boston. The Seville leg's own fare products, with mean demand 112, stand for everyone else on the flight: the 82 local travellers and about 30 who connect from the five spokes outside the model. The bank's total of 161 per departure lies between the case's shoulder and peak means, so its connecting share is a little above the 45 percent just computed. In the model's terms feed is $49/161 = 30$ percent of demand and, at the case fares, CAD 40,321 of the flight's CAD 121,709 of O&D ticket revenue, or 33 percent; counting the connections hidden in the leg's own products, about half of the flight connects. A small network model treats as local whatever arrives from outside it, and its user should know how much of the demand that is. Before Seville the bank served five of the model's markets; afterwards it serves eight.

What the feed does to the feeders. The connecting travellers need seats on the 137-seat A220s from Quebec City and Boston. Mean demand on the Quebec City flight rises from 121 to 141, and on the Boston flight from 124 to 153. Treating each leg's demand as normal, with variances added across the fare products that use it, the spill formula of Chapter 9 shows what happens. On the Quebec City flight, expected spill rises from 2.0 to 10.0 travellers per departure and the share of departures that leave full from 20 to 58 percent. On the Boston flight, expected spill rises from 2.8 to 18.6 and the share of full departures from 25 to 78 percent.

The two feeders together turn away about 24 more travellers per bank, half as many as the new feed. Some are Seville connectors, some are Quebec City and Boston locals, some are Paris connectors; which ones depends on how revenue management controls the feeders (Chapter 20). The Seville flight itself, with

mean demand 161 against 182 seats, spills 3.0 travellers per departure.

What the planner learns. The network effect of a spoke is a count of plausible markets times their demand, minus what the new traffic displaces on the legs it shares. Geometric counts nearly double the first term here, and the second term is not small: adding Seville turns two comfortable feeders into constrained ones, which makes them candidates for an upgauge (Chapters 9 and 14). Chapter 12 values both effects in dollars.

10.7 Alliances, codeshares, interline and joint ventures

An airline can extend its network without flying more aircraft by cooperating with other airlines. The forms differ in what they permit, in how revenue is shared and in how competition authorities treat them (Table 10.1), and the handbook insists on keeping them apart (Network Planning Handbook 2026).

Table 10.1 Forms of airline cooperation. Summarized from the handbook and the US Department of Transportation (Network Planning Handbook 2026; U.S. Department of Transportation 2024).

Form	What it permits	How revenue is shared	Competition treatment
Interline	One ticket across two or more airlines, with baggage checked through	Each carrier receives a share of the through fare under a prorate agreement	Ordinary commercial agreement
Codeshare	The marketing airline sells a partner's flight under its own flight number; no extra aircraft movement	Settlement between marketing and operating carrier, by prorate or a negotiated rate	Reviewed as a cooperative agreement; no joint pricing
Alliance	Multilateral cooperation: codeshares, loyalty reciprocity, lounges, coordinated connections	As for interline and codeshare	Membership alone does not permit joint pricing or capacity coordination
Joint venture	Coordinated fares, schedules, capacity and sales within a defined scope	Revenue or profit pooled and shared by formula, regardless of who operates the aircraft	Requires antitrust immunity or regulatory approval, usually with conditions

The joint venture is the deepest form. Its partners share revenue or profit on the markets in scope whichever partner operates the aircraft, a property called **metal neutrality**, so that each partner is indifferent about whose flight a customer takes and can plan capacity for the combined network. In the United States, antitrust immunity granted by the Department of Transportation lets the partners coordinate fares, services and capacity as if they were a single carrier in the relevant markets, subject to conditions in the order that grants it, after a competitive analysis and a public-benefits analysis (U.S. Department of Transportation 2024).

Why would a competition authority ever allow two airlines to set prices together? The economic answer is that for interline travellers the partners' flights are complements, not substitutes, and independent pricing of complements harms both the airlines and their customers.

RESULT 10.3 Double marginalization on an interline itinerary

Let demand for a two-carrier itinerary be $Q = A - P$, where $P = p_1 + p_2$ is the sum of the two carriers' segment prices and c_1, c_2 are their marginal costs per passenger, with $c = c_1 + c_2 < A$. If each carrier sets its segment price to maximize its own profit, the equilibrium through price is

$$P^{\text{sep}} = \frac{2A + c}{3},$$

while a joint venture that maximizes joint profit charges

$$P^{\text{JV}} = \frac{A + c}{2} = P^{\text{sep}} - \frac{A - c}{6}.$$

Joint pricing lowers the through fare, raises traffic from $(A - c)/3$ to $(A - c)/2$, raises the partners' combined profit and raises consumer surplus.

Proof. Carrier 1 maximizes $(p_1 - c_1)(A - p_1 - p_2)$; the first-order condition gives $p_1 = (A - p_2 + c_1)/2$, and symmetrically for carrier 2. Subtracting c_i and adding the two conditions shows that the two markups are equal at $p_i - c_i = (A - c)/3$, so $P^{\text{sep}} = c + 2(A - c)/3 = (2A + c)/3$ and $Q^{\text{sep}} = (A - c)/3$. A single decision-maker maximizes $(P - c)(A - P)$ and sets $P^{\text{JV}} = (A + c)/2$ with $Q^{\text{JV}} = (A - c)/2$. Joint profit is $(A - c)^2/4$ against $2(A - c)^2/9$ under separate pricing, and consumer surplus $Q^2/2$ rises with Q. ∎

This is Cournot's complements problem, and Brueckner (2001) uses it to analyze international alliances: cooperative pricing lowers fares for interline passengers, while on gateway-to-gateway routes where the partners would otherwise compete, cooperation can raise fares. That tension is why approvals typically come with conditions and carve-outs.

WORKED EXAMPLE 10.3 Pricing a beyond itinerary with a partner

Laurentian flies Montreal–Paris and a European partner flies onward from Paris. Take the illustrative through demand $Q = 1{,}600 - P$ (passengers per month), with marginal costs of CAD 300 on Laurentian's segment and CAD 200 on the partner's, so $c = 500$. Pricing separately, each carrier's markup is $(1{,}600 - 500)/3 = 366.7$, the through fare is CAD 1,233, 367 passen-

gers travel, and the two carriers earn CAD 268,889 together. A metal-neutral joint venture charges CAD 1,050, carries 550 passengers and earns CAD 302,500; consumer surplus rises from CAD 67,222 to CAD 151,250 a month. Both sides gain because the separate prices impose a negative externality on the partner: each carrier's markup reduces the traffic on which the other earns its own.

For the planner the lesson is practical: the revenue a route receives from partner feed depends on the cooperation agreement. The handbook warns that an ordinary prorate and a joint-venture revenue allocation do not produce the same answer for the same passenger (Network Planning Handbook 2026).

Air Canada's public partnership map shows the deeper forms of cooperation. Air Canada is a founding member of Star Alliance, created in 1997, and its partner page names joint ventures with United and the Lufthansa Group on transatlantic travel, with United on most flights between Canada and the United States, and with Air China on travel between Canada and China (Air Canada n.d.b; Network Planning Handbook 2026). The page itself carries a 2022 copyright notice, so treat it as a starting map. The transborder joint business agreement with United has operated since 2022; on 16 October 2026 the two carriers announced that they would renew and expand it, subject to regulatory approval, and that Air Canada had applied to Transport Canada for approval of the amended agreement (Air Canada 2026h). Transport Canada opened a public consultation, and the expansion would bring in ten city pairs that had been carved out of the cooperation (Travelweek 2026a).

The handbook's questions for the first week apply to any route that touches a partner: which geographic flows are in scope; how revenue and costs are settled; which partner feed can be assumed; what is incremental to the airline as opposed to the joint venture; and how the business case treats revenue on partner-operated legs. It recommends tracking three views where relevant: the operated flight's economics, the airline's incremental network economics, and the partnership's economics (Network Planning Handbook 2026).

PITFALL

Assuming that a codeshare or alliance membership permits joint pricing or capacity planning. Neither does (Network Planning Handbook 2026). Coordination of fares and capacity requires an approved joint venture, only within its approved scope, and only outside any carved-out markets. A planner who shares capacity plans with a partner outside that scope creates a legal problem, not a network benefit.

10.8 Freedoms of the air, bilateral agreements and Blue Sky

A commercially attractive route also needs the right to fly it. International scheduled services are governed by air transport agreements between states, which specify which airlines may fly, between which points, how often and with what freedom of pricing. The rights they exchange are described by a standard vocabulary.



Figure 10.5 Freedoms three to nine for an airline of home country H. Accent arrows carry the traffic each right concerns; grey dashed arrows are legs flown under another right. The examples are for a Canadian airline.

DEFINITION 10.4 Freedoms of the air

For an airline of home country H, Transport Canada defines seven freedoms (Transport Canada n.d.):

1. to fly over another country without landing;
2. to land in another country for non-commercial reasons, such as refuelling or repairs;
3. to carry traffic from H to another country;
4. to carry traffic from another country to H;
5. to carry traffic between two foreign countries on a service that starts or ends in H;
6. to carry traffic between two foreign countries through H;
7. to carry traffic between two foreign countries on a service that does not touch H.

Cabotage is the carriage of domestic traffic within a country by a foreign airline. The labels **eighth freedom** (domestic traffic carried on a service that starts or ends in the airline's home country, also called consecutive cabotage) and **ninth freedom** (stand-alone cabotage) are commonly used for its two forms. ICAO's Manual on the Regulation of International Air Transport describes all nine but calls every freedom beyond the fifth "so-called", because only the first five are recognized by treaty (International Civil Aviation Organization n.d.).

Figure 10.5 draws freedoms three to nine. A hub carrier's connecting traffic is built from third and fourth freedoms: a Boston–Paris traveller on a Canadian airline uses Canada–United States rights for the first leg and Canada–France rights for the second, and the sixth freedom names the combination. Fifth freedoms are rarer and more valuable, because they let an airline compete in a market between two foreign countries.

Canada's approach to negotiating these rights is its **Blue Sky policy** of 2006. Its objectives include encouraging competition and new international services, helping Canadian airlines compete in a more liberalized market and letting airports market themselves with minimal bilateral constraints; its principles include letting market forces set price, quality, frequency and the range of routes (Transport Canada n.d.). Its primary goal is reciprocal Open Skies-type agreements where they are in Canada's overall interest, which Transport Canada describes as including open third and fourth freedoms, no limits on airlines, frequencies or aircraft types, market-based pricing, flexible codesharing, unrestricted fifth and sixth freedoms, and seventh freedoms for all-cargo services. The policy excludes cabotage outright. Where a partner country will not agree to Open Skies, Canada may negotiate a smaller exchange of rights, and it may decline Open Skies where Canadian airlines would face discriminatory access or where competition would be significantly reduced (Transport Canada n.d.).

Three practical consequences matter for planning. First, traffic rights and airport slots are separate constraints: the right to serve a country does not guarantee usable arrival and departure times (Chapter 13). Second, a route outside an agreement's terms may still be possible through an extra-bilateral authority, which is a separate application, not an assumption. Third, licences, designations and operational permissions are distinct workstreams with their own owners (Network Planning Handbook 2026). The handbook's advice is to ask the responsible team for the actual treaty, licence, exceptions, seasonal limits and permitted commercial activities rather than infer them from the freedom categories (Network Planning Handbook 2026).

10.9 Market structure and competition in Canada

In June 2025 the Competition Bureau published a market study of Canada's domestic airline industry, Cleared for take-off (Competition Bureau Canada 2025). Its executive summary is short and worth reading in full; the numbers below are taken from it.

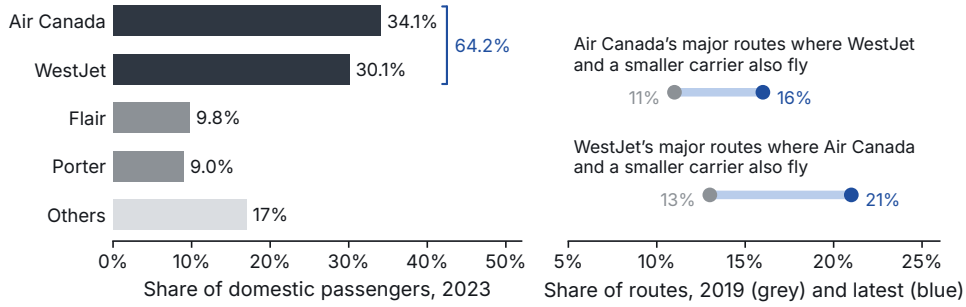


Figure 10.6 Canadian domestic market structure in the Competition Bureau's 2025 study. Left: estimated shares of domestic passengers in 2023; the two largest carriers hold 64.2 percent. Right: the share of each major carrier's main routes on which the other major and at least one smaller carrier (Flair or Porter) also fly, in 2019 and in the study's latest data. Redrawn from the executive summary (Competition Bureau Canada 2025).

Concentration. At major airports Air Canada and WestJet together carried between 56 and 78 percent of domestic passengers in 2023. Concentration fell by one tenth between 2019 and 2023 as Porter and Flair entered and grew, even though Lynx Air exited, but the Bureau calls average concentration extremely high and new competition fragile. WestJet's share is falling more at major eastern airports and Air Canada's more at major western ones, and the two compete directly on fewer routes than in 2019 (Competition Bureau Canada 2025). Nationally, the Bureau estimates 2023 domestic passenger shares of 34.1 percent for Air Canada, 30.1 for WestJet, 9.8 for Flair, 9.0 for Porter and 17 for all others, and finds that the four busiest airports (Toronto Pearson, Vancouver, Calgary and Montreal-Trudeau) handle 55 percent of domestic passengers.

Entry and fares. The Bureau's research finds that when one new competitor starts flying a route between two cities, average fares fall by 9 percent (Competition Bureau Canada 2025); the full report puts this at about CAD 27.80 per base fare. The share of Air Canada's major routes on which WestJet and at least one smaller carrier also fly rose from 11 to 16 percent after 2019, and the corresponding share for WestJet from 13 to 21 percent (Figure 10.6). The history of entry and exit, from Canada 3000 and JetsGo to Lynx Air and Canada Jetlines in 2024, is why the Bureau calls these gains fragile.

Barriers. The summary lists the challenges new airlines face: small scale and infrequent service, the difficulty of building a brand, high start-up costs with slow revenue growth, lack of access to interline agreements and alliances, and loyalty programs that favour incumbents. Fees under Canada's user-pay model make up about 30 percent of what passengers pay traditional full-service airlines

(airport fees about 20 percent, NAV CANADA charges 5, the Air Travellers Security Charge 3 and fuel tax 1) and a higher share for ultra-low-cost carriers. Slot allocation rules favour incumbents at busy airports in peak hours, airport infrastructure and incentive decisions can tilt access, and foreign ownership limits and the ban on cabotage restrict access to capital (Competition Bureau Canada 2025). The Bureau also notes that incumbents' competitive responses, lower prices and more seats, are a normal part of competition, while the Competition Act restricts predation and related conduct that harms competition by driving out rivals; telling the two apart requires careful analysis.

Recommendations. The Bureau makes ten recommendations in three areas (Competition Bureau Canada 2025). To prioritize competition in aviation policy: remove the Minister of Transport's power to override the Competition Act's review of airline mergers and competitor collaborations; remove barriers that keep smaller airports from competing with major hubs; publish airline traffic, quality and slot data; and consider reviewing the airport oversight and funding model. To use international capital: raise the single-investor foreign ownership limit to 49 percent; allow up to 100 percent foreign ownership of a new class of domestic-only Canadian airlines; and work with other countries to remove foreign-competition restrictions, including on cabotage, from air agreements. To support northern and remote access: coordinate leadership through a national working group, tailor regulation to the North, and use government investments and tools, including the promotion of interlining, to foster competition. The first recommendation bears directly on airline joint ventures, which the Minister of Transport reviews (Travelweek 2026a) (Section 10.7).

WORKED EXAMPLE 10.4 How concentrated is the domestic market?

The Herfindahl-Hirschman index (HHI) sums squared market shares in percent. From the Bureau's national shares, the four named carriers contribute $34.1^2+30.1^2+9.8^2+9.0^2 = 2{,}246$. The 17 percent held by others adds between almost nothing (if it is spread over many small carriers) and $17^2 = 289$ (if it were one firm), so the national HHI lies between 2,246 and 2,535. The two-firm concentration ratio is 64.2 percent and the four-firm ratio 83.0 percent.

These numbers understate concentration in the markets that matter. Travellers buy trips between two cities, not a national product, so the relevant antitrust market is the O&D market, possibly widened to nearby airports, and many routes have one or two carriers. The Bureau's full report says average concentration is more than three times the level at which the Competition Act's merger provisions consider markets concentrated (Competition Bureau Canada 2025), which is consistent with route-level and airport-level indices far above the national figure. At Montreal-Trudeau, the Bureau's airport chart shows Air Canada with roughly three fifths of domestic passengers, the highest share of any large airport in the chart (this figure is read off the chart, not published as a number). For an industrial economist, the lesson is the one learned in merger review: measure concentration in the market where substitution happens.

10.10 Hub premiums

Fares on routes to and from an airline's hub tend to be higher than fares in comparable markets elsewhere. The phenomenon is old and well documented, and the explanations matter for a planner because they determine what a hub's revenue advantage depends on.

DEFINITION 10.5 Hub premium

The **hub premium** of an airline is the difference between its average fare in markets that touch its hub and its average fare in otherwise comparable markets, usually measured by a regression of fares on market characteristics and a hub indicator or the airline's airport share.

Borenstein (1989) showed with US data that an airline's dominance at the endpoint airports of a route is strongly associated with higher fares, and interpreted the association as market power. Berry, Carnall and Spiller (1996) estimated a differentiated-products model in which travellers differ in their sensitivity to price. In their estimates the hub carrier's ability to raise prices is concentrated on tickets bought by price-insensitive business travellers, who prefer the carrier that dominates their home airport, with an average premium of about 20 percent; the high hub fares do not shelter other

airlines, and the cost side shows economies of density on longer routes. The mechanisms usually proposed for the preference are familiar to any planner: the hub carrier has the most departures and the best schedule from its hub city, its loyalty program is the one local travellers collect in, its corporate contracts are with local employers, and its control of gates and slots makes entry harder. The handbook cautions that the historical estimates should not be transferred to current markets (Network Planning Handbook 2026).

A one-line result from discrete choice shows how these sources of preference become prices.

RESULT 10.4 Logit markups rise with share

Suppose travellers in a market choose among airlines and an outside option with logit probabilities, utility $V_j = \xi_j - \alpha p_j$ for airline j (where ξ_j collects schedule, loyalty and product quality and $\alpha > 0$ is price sensitivity), and $V_0 = 0$ for not flying. If airline j sells one product with marginal cost c_j and share s_j, its profit-maximizing price satisfies

$$p_j - c_j = \frac{1}{\alpha\,(1 - s_j)}. \tag{10.4}$$

Proof. With market size M, profit is $(p_j - c_j)\,M s_j(p)$ where $s_j = e^{V_j}/(1+\sum_k e^{V_k})$. The logit derivative is $\partial s_j/\partial p_j = -\alpha s_j(1-s_j)$ (Chapter 5). The first-order condition $s_j - (p_j - c_j)\,\alpha s_j(1 - s_j) = 0$ gives (10.4). ∎

Anything that raises ξ_j for the hub carrier, such as frequency, loyalty or airport presence, raises its share and therefore its markup, even with identical costs. The markup is also inversely proportional to price sensitivity, so the same increase in share is worth far more among business travellers. Figure 10.7 draws the two effects.

WORKED EXAMPLE 10.5 Price and mix effects in a hub premium

Take two segments with marginal cost CAD 300 per traveller: business travellers with $1/\alpha = 450$ dollars and leisure travellers with $1/\alpha = 150$. A carrier with 25 percent of each segment charges business travellers $300 + 450/0.75 = 900$ dollars and leisure travellers $300 + 150/0.75 = 500$. At its hub its presence lifts its business share to 55 percent but its leisure share only to

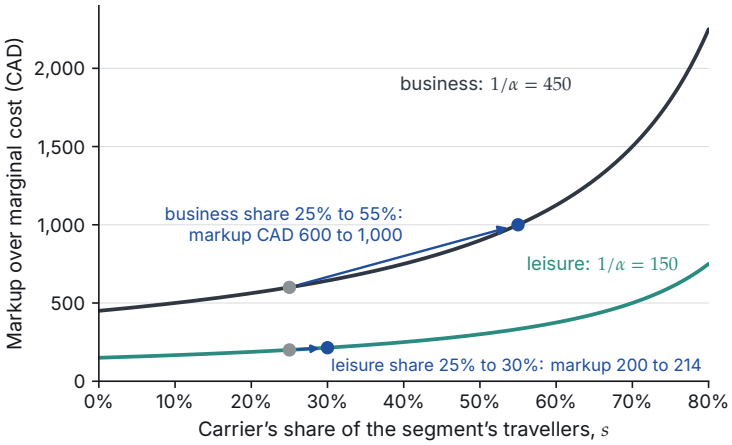


Figure 10.7 The profit-maximizing markup of a single-product logit firm against its share of a traveller segment, for price-insensitive and price-sensitive travellers. Hub presence that lifts the carrier's business share from 25 to 55 percent raises its business markup from CAD 600 to CAD 1,000; a smaller lift in leisure share barely moves the leisure markup. All values are illustrative.

30 percent, the pattern Berry, Carnall and Spiller found in direction if not in these illustrative magnitudes. Its fares become $300 + 450/0.45 = 1{,}300$ and $300 + 150/0.70 = 514$: a premium of 44 percent on business fares and 3 percent on leisure fares.

The measured average premium depends on the passenger mix. With 30 percent business travellers in both markets, the average fare rises from CAD 620 to CAD 750, a premium of 21 percent. If the hub also attracts more business travellers, say 40 percent, the average fare is CAD 829 and the measured premium 34 percent, of which more than a third comes from mix rather than price. A regression of average fares on a hub indicator that does not control for mix confounds the two, which is one more reason fares cannot be treated as exogenous in demand estimation (Chapter 7).

For a network planner, hub premiums have three implications. They are part of the revenue that makes thin spokes viable, so a hub's economics cannot be read from cost data alone. They should not be imported into the valuation of a market where the airline lacks the same presence: Laurentian's Montreal-origin fares say little about what it could earn starting a route from Toronto. And they are fragile where their source is the absence of competition: the Bu-

reau's estimate that one additional competitor lowers fares by 9 percent (Competition Bureau Canada 2025) is a reminder that some of the premium can disappear with a single entrant.

10.11 Summary

- Demand is generated by O&D markets and supply by legs; a hub consolidates many markets onto each leg and lets an airline fly higher frequencies and larger aircraft.
- A hub with n spokes connects $n(n+1)/2$ markets with n routes, and its next spoke adds $n+1$ markets. Timing, circuity and demand screens cut the geometric count sharply; in Laurentian's network the circuity screen alone removes a third of it.
- Airline costs show strong economies of density but roughly constant returns to scale; a hub is a device for buying density.
- A market deserves its own nonstop only when it can pay an extra departure: $Q^* = a_m/(\kappa_1 + \kappa_2 - \kappa_m)$. Behind a long-haul gateway the threshold is several hundred travellers a day, which is why hubs serve thin markets and why smaller long-haul aircraft let some markets leave the hub.
- Banks create connections; feed is valuable when seats are empty and displaces other travellers when they are not. Adding Seville to Laurentian's hub adds 8 plausible markets, not 15; connecting travellers are about 45 percent of the flight's shoulder-season demand, and on the two modelled feeders they add roughly 24 spilled travellers per bank.
- Interline, codeshare, alliance and joint venture permit different things; only an approved joint venture permits joint pricing, and joint pricing of complementary flights lowers fares by removing double marginalization.
- Traffic rights come from bilateral agreements; Canada's Blue Sky policy seeks Open Skies-type agreements but excludes cabotage. Rights, slots and licences are separate constraints.
- The Competition Bureau found Canada's domestic market highly concentrated (Air Canada and WestJet hold 56 to 78 percent of domestic passengers at major airports) and estimated that one more competitor lowers fares by 9 percent.
- Hub premiums come from schedule, loyalty and airport presence, concentrated on price-insensitive travellers; measured premiums mix price and passenger-mix effects.

Table 10.2 Key formulas of the chapter.

Quantity	Formula
Markets connected by one hub	$M(n) = n(n+1)/2$, increment $n+1$
Circuity of a connection	$\rho_{ij} = (d_{iH} + d_{Hj})/d_{ij}$
Returns to density and scale	$\text{RTD} = 1/\epsilon_{CQ}$, $\text{RTS} = 1/(\epsilon_{CQ} + \epsilon_{CN})$
Hub-versus-nonstop threshold	$Q^{*} = a_m/(\kappa_1 + \kappa_2 - \kappa_m)$, with $\kappa = c/\text{LF} + c_{\text{pax}}$
Viable connection	$\text{MCT}_{ij} \le t_j^{\text{dep}} - t_i^{\text{arr}} \le T_{\max}$
Interline prices	$P^{\text{sep}} = (2A + c)/3$, $P^{\text{JV}} = (A + c)/2$
Logit markup	$p_j - c_j = 1/(\alpha(1 - s_j))$

10.12 Further reading

Belobaba's lectures on airline economics and on route planning (Belobaba 2006d, 2006i) set out the dichotomy of demand and supply and the hub economics this chapter formalizes, and his schedule development lecture introduces load consolidation (Belobaba 2006f). The handbook's sections on connections, partnerships and market access (Network Planning Handbook 2026) are the best short guide to what to ask in a first week, and the Competition Bureau's executive summary (Competition Bureau Canada 2025) is the essential Canadian reference. For the economics, read Caves, Christensen and Tretheway (1984) on density and scale, Borenstein (1989) and Berry, Carnall and Spiller (1996) on hub premiums, and Brueckner (2001) on alliances. Transport Canada's Blue Sky policy page (Transport Canada n.d.) defines the freedoms and Canada's negotiating stance. The Global Airline Industry (Belobaba, Odoni, and Barnhart 2015) and Doganis (2019) cover alliances and network strategy at textbook length.

CHAPTER

11

Fleet Planning

IN THIS CHAPTER

- Fleet planning as a staged decision under uncertainty: the evaluation chain, the horizons and what the network analyst contributes.
- The top-down capacity-gap model and the bottom-up route model, with their equations and their blind spots.
- How airlines evaluate an aircraft: economics, range and payload, commonality, availability and financing.
- Payload-range diagrams from first principles, redrawn from the A321 manual and reproduced by a two-parameter fuel model.
- The A321XLR on Montreal–Seville: distance, winds, why payload can bind westbound in winter, and the approvals behind extended diversion operations.
- Lease versus buy, the economics of commonality, and what turnaround time does to utilization.
- A Laurentian fleet decision for thin transatlantic markets, with its sensitivity analysis.

DRAWS ON Belobaba, Airline fleet planning models (MIT 16.75J, 2006); Airbus A321 aircraft characteristics, airport and maintenance planning (rev. 33, 2024); handbook; Air Canada Q2 2026 MD&A

A fleet is the most expensive and least reversible commitment an airline makes about its network. Aircraft are ordered years before delivery, financed over a decade or more and flown for two or three decades. The analyst who designs next summer's schedule inherits fleet decisions taken years earlier: which markets can be served at all, at what trip cost and with how many seats. Fleet decisions also depend on network planning in return, because every fleet plan rests on assumptions about routes, frequencies and demand that come from the planning team.

This chapter treats fleet planning from the analyst's seat. It starts with the evaluation process and its horizons, then sets out the two classical families of fleet planning models, top-down and bottom-up.

It then turns to the aircraft itself: the criteria by which airlines compare types, the payload-range diagram that summarizes what an aircraft can carry how far, and the A321XLR's mission to Seville as a worked case, including the winds that make the westbound flight the harder one and the approvals that a twin-engine aircraft needs over the ocean. Financing, commonality and turnaround time complete the economics. The chapter ends with a fleet decision for Laurentian that brings together the trip-cost model of Chapter 3 and the spill model of Chapter 9. Fleet assignment, which takes the fleet as given and assigns types to flights, is the subject of Chapter 14.

11.1 The fleet planning problem

A fleet is the number of aircraft an airline operates and the mix of types that make it up. Each type has its own technical capability, notably how much payload it can carry how far, so fleet composition shapes the airline's costs, its finances and, above all, the routes it can serve (Belobaba 2006k). The sums are large. In 2006 a 150-seat narrowbody cost USD 40 to 60 million and a long-range 747-400 more than USD 180 million; depreciation affected the balance sheet for 10 to 15 years, and some aircraft were flown economically for more than 30 (Belobaba 2006k). The prices are historical; the proportions are not.

> **DEFINITION 11.1 Fleet plan**
>
> A **fleet plan** specifies, for each future period, the number of aircraft of each type the airline will operate, together with the timing of deliveries, lease returns, retirements and transfers between subfleets that produce those numbers. Choosing it is a **staging problem**: decisions are taken in sequence under uncertainty about demand, costs and competition, and each is constrained by the existing fleet, by the ease of disposing of older aircraft and by the delivery positions manufacturers still have available (Belobaba 2006k).

The lecture describes fleet planning as an evaluation chain, redrawn in Figure 11.1. Traffic and yield forecasts give revenues; a planning load factor converts the traffic forecast into a required capacity in seat-kilometres; dividing by what one aircraft produces gives the number of aircraft by type; the aircraft imply an investment, and with it depreciation, lease payments, insurance and in-

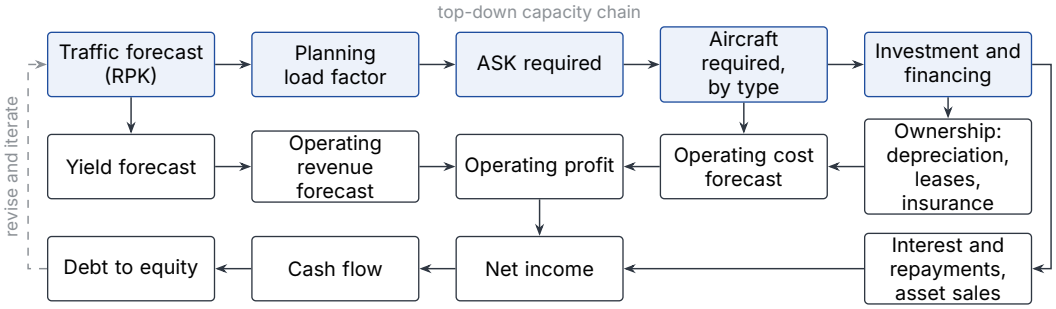


Figure 11.1 The fleet planning evaluation process (adapted from Belobaba, 2006). The top row is the capacity chain used by top-down models; the lower rows carry the plan through costs, revenues and financing to the balance sheet.

terest; operating costs and revenues give operating profit, and the financing flows carry the plan through to net income, cash flow and the debt-to-equity ratio. The process is iterative and should involve many parts of the airline, because each output feeds back into the assumptions (Belobaba 2006k).

Fleet decisions span several horizons. At the strategic horizon of ten years or more, the airline chooses its fleet families and places orders. At the tactical horizon of two to five years, it adjusts delivery timing, extends or returns leases, retires older aircraft and moves aircraft between brands or bases. At the seasonal horizon, the fleet is fixed and the question becomes which type flies which flight (Chapter 14); in operations, which individual aircraft flies which rotation (Chapter 15). The handbook frames the multi-year question as which regions and aircraft capabilities deserve scarce capital, and asks for scenarios, fleet availability, a demand outlook and airport and regulatory feasibility as evidence (Network Planning Handbook 2026).

Table 11.1 shows what such a plan looks like in practice: Air Canada's actual and planned fleet for selected types, as published in its second-quarter 2026 MD&A.

Table 11.1 Air Canada's operating fleet at 31 December 2025 and 30 June 2026, and planned at 31 December 2026 and 2027, for selected types (Air Canada 2026e). Planned transfers are subject to the necessary approvals.

Type	Dec 2025	Jun 2026	Plan Dec 2026	Plan Dec 2027
Boeing 787-9	32	32	32	32
Boeing 787-10	0	0	2	8
Airbus A321XLR	0	2	9	19
Airbus A321, mainline	21	21	26	34
Airbus A220-300	41	46	57	65
Boeing 737 MAX 8, mainline	47	31	0	0
Boeing 737 MAX 8, Rouge	0	21	52	52
Mainline and Rouge, all types	256	262	267	288

The table is a staging plan in miniature. Narrowbody capacity grows through A220 and A321XLR deliveries, the 737 MAX 8s move from mainline to the leisure brand, and the 787-10 joins the widebody fleet. The MD&A reports that the 787-10 order was reduced from 18 to 14 aircraft in November 2025, that eight A350-1000s were ordered in February 2026 for delivery from 2030, and that committed capital expenditure stood at CAD 12.7 billion on 30 June 2026 (Air Canada 2026e). None of these numbers is a forecast for a particular route; together they bound what the network can do for years.

11.2 Top-down and bottom-up models

The lecture distinguishes two approaches (Belobaba 2006k). A **top-down** (macro) model works with aggregates: total traffic for the airline or a region, an average stage length, an average utilization. A **bottom-up** (micro) model builds future networks and schedules route by route and derives the fleet from them.

The top-down capacity gap The top-down model follows the capacity chain of Figure 11.1. Let RPK_t be the traffic forecast for year t and $\mathrm{LF}^{\mathrm{plan}}$ the planning load factor, the average load factor the airline is prepared to plan for. Let N_{kt} be the number of aircraft of type k the airline will still have in year t after planned retirements and lease returns, and q_k the seat-kilometres one aircraft of type k produces in a year. Then

$$G_t = \frac{\mathrm{RPK}_t}{\mathrm{LF}^{\mathrm{plan}}} - \sum_k N_{kt}\, q_k, \qquad q_k = 365\, U_k\, C_k\, s_{b,k}, \tag{11.1}$$

where G_t is the **capacity gap**, U_k the utilization in block hours per day, C_k the seats and $s_{b,k} = d_k/\mathrm{BH}(d_k)$ the block speed at the type's typical stage length d_k (Chapter 3). Filling the gap with type k alone needs $\lceil G_t/q_k \rceil$ new aircraft; the types are then compared on the annual cost of producing G_t seat-kilometres, operating cost plus ownership cost per ASK.

RESULT 11.1 The capacity gap is leveraged on planning assumptions

Let $A_t = \mathrm{RPK}_t/\mathrm{LF}^{\mathrm{plan}}$ be the required capacity. The elasticity of the capacity gap with respect to the planning load factor is

$$\frac{\partial \ln G_t}{\partial \ln \mathrm{LF}^{\mathrm{plan}}} = -\frac{A_t}{G_t},$$

and its elasticity with respect to the traffic forecast is $+A_t/G_t$. Because the gap is usually a small fraction of the required capacity, both are large.

Proof. $\partial G_t/\partial \mathrm{LF}^{\mathrm{plan}} = -\mathrm{RPK}_t/(\mathrm{LF}^{\mathrm{plan}})^2 = -A_t/\mathrm{LF}^{\mathrm{plan}}$; multiply by $\mathrm{LF}^{\mathrm{plan}}/G_t$. The second statement follows the same way from $\partial A_t/\partial \mathrm{RPK}_t = A_t/\mathrm{RPK}_t$. ∎

WORKED EXAMPLE 11.1 Laurentian's capacity gap

Give Laurentian, for illustration, a 2027 fleet of 30 A220-300s, 20 737-8s, 6 A321XLRs and 10 787-9s, flown at 9, 10, 13 and 14 block hours a day on typical stages of 900, 2,000, 5,500 and 6,500 km. With the block-time function of Chapter 3, one aircraft of each type produces 237, 411, 667 and 1,193 million ASK a year, and the fleet 31.3 billion, enough for 25.6 billion RPK at a planning load factor of 82 percent. Traffic grows 4 percent a year and two 737-8 leases end each year from 2030.

By 2032 the required capacity is 38.0 billion ASK and the remaining fleet supplies 28.8 billion: the gap is 9.2 billion ASK (Figure 11.2). Filled with A321XLRs alone it needs 13.8, so 14 aircraft; with 787-9s alone, 7.7, so 8. The XLR's avoidable cost

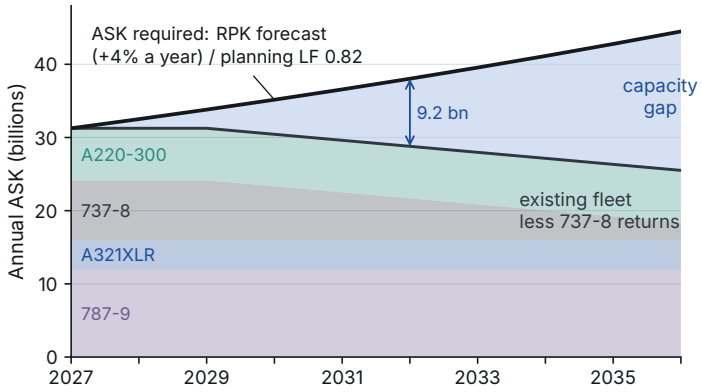


Figure 11.2 A top-down capacity gap for an illustrative Laurentian fleet: 30 A220-300s, 20 737-8s, 6 A321XLRs and 10 787-9s in 2027, with two 737-8 leases returned each year from 2030. Traffic grows 4 percent a year and the planning load factor is 82 percent.

at its typical stage is 8.17 cents per ASK and its illustrative ownership cost 1.22 cents; the 787-9's are 7.00 and 1.61 cents. The gap costs CAD 868 million a year with XLRs and 795 million with 787-9s, so the top-down answer is the widebody, by CAD 73 million a year.

Result 11.1 shows how fragile that answer is. The elasticity of the 2032 gap with respect to the planning load factor is $-38.0/9.2 = -4.1$: planning at 80 percent instead of 82 raises the gap to 10.2 billion ASK, planning at 85 percent shrinks it to 7.9 billion, a difference of four XLRs or two 787-9s. Traffic growth of 3 or 5 percent a year instead of 4 moves the 2032 gap to 7.4 or 11.1 billion ASK.

The top-down model is fast and transparent, which is why the lecture describes it as the more common approach: detailed scenarios ten or fifteen years ahead are speculative and likely to be wrong, and political decisions can override the analysis anyway (Belobaba 2006k). Its weakness is structural. It treats seat-kilometres as interchangeable, so it cannot tell whether the seats it plans can be sold in the markets where they would be flown. A 787-9 produces cheap seat-kilometres only if it flies full enough, and in a market that generates 150 passengers a day it does not.

The bottom-up model A bottom-up model starts from the network. Future routes and schedules are generated, the airline's share of each origin-destination market is assumed, and demand and revenue are

allocated to individual flights; the outputs are aircraft assignments, operating statistics by route and a full financial projection for each fleet plan (Belobaba 2006k). The lecture describes the approach in words; written as an optimization over a set of routes r and types k, it is

$$\begin{aligned} \max_{x,N} \ & \sum_r \sum_k \pi_{rk}\, x_{rk} - \sum_k o_k N_k \\ \text{s.t.} \ & \textstyle\sum_k x_{rk} = f_r \quad \text{for every route } r, \\ & \textstyle\sum_r t_{rk}\, x_{rk} \le H_k N_k, \quad N_k \le \bar{N}_k \quad \text{for every type } k, \\ & x_{rk} = 0 \ \text{ if route } r \text{ is beyond the reach of type } k, \\ & x_{rk},\, N_k \in \mathbb{N}, \end{aligned} \tag{11.2}$$

where x_{rk} is the number of weekly departures of type k on route r, f_r the planned weekly frequency, N_k the number of aircraft of type k (limited by delivery positions $\bar{N}_k$), o_k the weekly ownership cost of one aircraft, t_{rk} the aircraft time a departure consumes (block plus turn), H_k the hours an aircraft is available per week, and

$$\pi_{rk} = r_r\, \mathbb{E}[\min(D_r, C_k)] + K_{rk} - \mathrm{TC}_{rk}$$

the expected contribution of one departure: contribution r_r per passenger times the spill-adjusted expected load of Chapter 9, plus cargo contribution K_{rk}, minus the capacity cost TC_{rk}, the trip cost of Chapter 3 without its passenger term, which r_r already nets out. The reach condition is where payload-range analysis enters (Section 11.4). In the linear relaxation, the dual value of the aircraft-time constraint is the value of one more hour of type k, which the model weighs against its ownership cost. Making the frequencies f_r decisions too, with demand per departure depending on frequency, turns the model into the integrated schedule design and fleet planning problem of Chapters 13 and 14.

The bottom-up model gives route-level detail and supports what-if analysis, but it needs detailed future scenarios and handles competitors' strategies poorly (Belobaba 2006k). In practice airlines use both: top-down to size the fleet and test it against macro scenarios, bottom-up to check that the aircraft fit the markets and to compare types for a specific mission. The fleet decision at the end of this chapter is a small bottom-up evaluation, and it reaches a different answer from Example 11.1.

The financial evaluation of fleet alternatives rests on discounted cash flows (Belobaba 2006k). The lecture recommends combining net present value with Monte Carlo simulation of the uncertain

inputs (fuel prices, exchange rates, traffic growth, yields) to obtain a distribution of outcomes rather than a single number, and it warns that the discount rate is often chosen with a bias: too low at government-supported airlines, too high at private airlines trying to allow for the industry's volatility.

> **PITFALL**
>
> Treating seat-kilometres as interchangeable. A top-down comparison that picks the type with the lowest cost per ASK assumes every seat it plans can be sold at the planning load factor. On thin routes a large aircraft's cheap seat-kilometres are cheap only on paper: spill, frequency and the trip cost decide the economics, and only a route-level evaluation sees them.

11.3 Evaluating an aircraft

The lecture groups airlines' evaluation criteria into technical and performance characteristics, economics of operation and revenue, marketing and environmental issues, and political and trade concerns (Belobaba 2006k). Table 11.2 rearranges them around the questions a network planner is asked.

Table 11.2 Criteria for evaluating an aircraft type, adapted from Belobaba (2006).

Criterion	What it measures	Typical evidence
Economics	trip cost, seat cost, CASK at planned stages; cabin, cargo and revenue potential	trip-cost model, spill-adjusted contribution
Range and payload	payload-range, runway and airport performance, seasonal winds	manufacturer charts, operator performance data
Commonality	shared type ratings, spares, tooling, crew pools	fleet and crew plans
Availability	delivery positions, lease market, engine reliability	manufacturer backlog, lessors
Financing	price, residual value, lease rates, export credit	net present value, lease versus buy
Environment and regulation	noise, emissions, extended diversion approvals	certification, operator approvals
Marketing and politics	product appeal of a new type; trade and ownership pressure	qualitative judgment

Aircraft are usually described by size and range, and types with similar capabilities compete with each other in airlines' evaluations (Belobaba 2006k). Historically the relation between size and range was almost linear: the longest routes needed the largest aircraft. The lecture already noted the break in that pattern in 2006: new 100 to 150-seat aircraft had much longer ranges, US transcontinental routes were flown by 737s and A320s, and 180-seat 757s flew thin long-haul routes such as Newark–Lisbon (Belobaba 2006k). The A321LR and A321XLR extend the same logic across the Atlantic.

Availability deserves emphasis in 2026. IATA estimated in June 2026 that the industry's order backlog had reached 18,100 aircraft, almost 60 percent of the active fleet, and that delivery shortfalls were being absorbed through higher utilization, delayed retirements and higher load factors (IATA 2026d). Engine reliability is part of availability too: problems with Pratt & Whitney geared turbofan engines grounded, at the peak, about 650 GTF-powered A320neos (Hamilton 2026). A fleet plan that assumes every aircraft on order arrives on time and stays in service is a plan with no downside case.

11.4 Payload-range diagrams

The payload-range diagram is the most important technical description of an aircraft for a network planner (Belobaba 2006k). It answers one question: how much can this aircraft carry over a given distance? The answer comes from a handful of weight limits and the physics of fuel burn.

> **DEFINITION 11.2 Design weights**
>
> The **operating empty weight** (OEW) is the weight of the aircraft ready for service, with crew and equipment but without payload or usable fuel. **Payload** PL is the weight of passengers, baggage and cargo. The **zero-fuel weight** OEW + PL must not exceed the **maximum zero-fuel weight** (MZFW), a structural limit; the **take-off weight**, zero-fuel weight plus fuel, must not exceed the **maximum take-off weight** (MTOW); the fuel on board must not exceed the **fuel capacity** $F_{\max}$, usable tank volume times fuel density; and the landing weight must not exceed the **maximum landing weight** (MLW).

The A321 manual lists these limits for every weight variant of the family (Airbus 2024). The A321neo is offered with MTOWs from 80 to 95 tonnes and MZFWs of 67.0, 73.3 or 75.6 tonnes; the long-range neo-ACF variants reach 97 tonnes; the A321neo-XLR reaches 101 tonnes, with a MZFW of 73.3 or 75.6 tonnes and an MLW of 77.3 or 79.2 tonnes. Usable fuel is 23,490 litres in a standard A321neo, up to 32,943 litres in a neo-ACF with three additional centre tanks, and 39,748 litres in the XLR with its permanent rear centre tank and optional forward tank. At the manual's density of 0.785 kg per litre, that is 18.4, 25.9 and 31.2 tonnes.

Fuel and range Fuel needed for a given range follows from the **Breguet range equation**. In steady cruise, lift equals weight W and thrust equals drag, $W/(L/D)$, where L/D is the lift-to-drag ratio; the engines burn fuel at a rate c per unit of thrust per hour. Weight therefore falls at the rate $\mathrm{d}W/\mathrm{d}t = -c\,W/(L/D)$, while the aircraft covers ground at speed V. Dividing, $\mathrm{d}R/\mathrm{d}W = -V(L/D)/(c\,W)$, and integrating from the start weight W_0 to the end weight W_1,

$$R = \Lambda \ln \frac{W_0}{W_1}, \qquad \Lambda = \frac{V\,(L/D)}{c}. \tag{11.3}$$

The **range factor** Λ, measured in nautical miles (nm), summarizes the aerodynamics and engines. Trip fuel for range R is $F_t = W_0(1 -$

$e^{-R/\Lambda}$): a fixed fraction of the take-off weight. An airline must also carry reserves. We model them as a fixed amount r_0 (final reserve, alternate fuel and other allowances) plus contingency fuel equal to a share ε of trip fuel, so that fuel at take-off is $F = r_0 + (1+\varepsilon)F_t$.

RESULT 11.2 The payload-range diagram

The maximum payload at still-air range R is the smallest of three limits, $\mathrm{PL}(R) = \min\{P_1, P_2(R), P_3(R)\}$. Writing $e_R = 1 - e^{-R/\Lambda}$,

$$\begin{aligned} P_1 &= \mathrm{MZFW} - \mathrm{OEW}, \\ P_2(R) &= \mathrm{MTOW} - \mathrm{OEW} - r_0 - (1+\varepsilon)\,\mathrm{MTOW}\,e_R, \\ P_3(R) &= \frac{F_{\max} - r_0}{(1+\varepsilon)\,e_R} - F_{\max} - \mathrm{OEW}. \end{aligned}$$

The curve has three segments: a **structural** plateau, a **weight-limited** segment on which payload is traded for fuel, and a **fuel-limited** segment on which tanks are full and payload is removed only to make the aircraft lighter. They meet at corner A (MZFW and MTOW both bind), corner B (MTOW and fuel capacity both bind) and corner C, the ferry range with no payload. With F_t the trip fuel and W_0 the take-off weight, the slopes are

$$\text{A to B:}\quad \frac{\mathrm{dPL}}{\mathrm{d}R} = -\frac{(1+\varepsilon)(\mathrm{MTOW} - F_t)}{\Lambda}, \qquad \text{B to C:}\quad \frac{\mathrm{dPL}}{\mathrm{d}R} = -\frac{W_0(W_0 - F_t)}{\Lambda F_t}.$$

Proof. P_1 is the structural limit. If take-off weight equals MTOW, the Breguet equation gives $F_t = \mathrm{MTOW}\,e_R$, so the zero-fuel weight is $\mathrm{MTOW} - r_0 - (1+\varepsilon)\mathrm{MTOW}\,e_R$, which gives P_2. If the tanks are full, $F_t = (F_{\max} - r_0)/(1+\varepsilon)$ is fixed, the take-off weight that flies exactly R is $W_0 = F_t/e_R$, and the zero-fuel weight is $W_0 - F_{\max}$, which gives P_3. For a given range, payload rises with take-off weight, so each expression is the largest payload compatible with its own constraint; all three constraints must hold, so the feasible maximum is the smallest. Differentiating P_2 gives $-(1+\varepsilon)\mathrm{MTOW}\,e^{-R/\Lambda}/\Lambda$, and $\mathrm{MTOW}\,e^{-R/\Lambda} = \mathrm{MTOW} - F_t$. For P_3, $\mathrm{d}(F_t/e_R)/\,\mathrm{d}R = -F_t e^{-R/\Lambda}/(\Lambda e_R^2) = -W_0(W_0 - F_t)/(\Lambda F_t)$, using $W_0 = F_t/e_R$ and $W_0 e^{-R/\Lambda} = W_0 - F_t$. ∎

At corner B the take-off weight is MTOW in both expressions, so the fuel-limited slope is $\mathrm{MTOW}/((1+\varepsilon)F_t)$ times the weight-limited

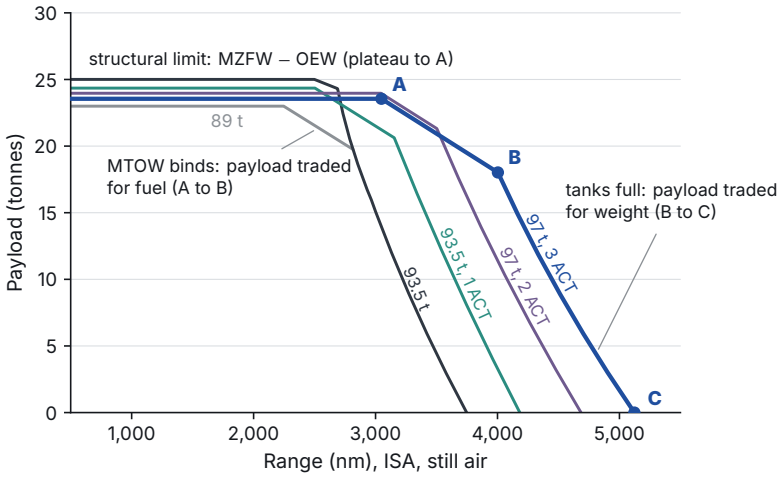


Figure 11.3 Payload-range of the A321neo at ISA in still air, redrawn from the Airbus A321 airport and maintenance planning manual (revision 33, June 2024, figure 3-2-1-991-021). Manufacturer data, given for information only. The 97-tonne, three-ACT curve is marked with its corners A, B and C.

one: about six times for an A321neo without additional tanks, and four to five times with them, because more fuel in the tanks means a larger F_t at the knee. That ratio is the most practical fact in the diagram: before the knee, each extra 100 nautical miles costs about half a tonne of payload; beyond it, several tonnes.

Figure 11.3 redraws the A321neo chart of the manual (Airbus 2024). The five curves are for MTOWs of 89 and 93.5 tonnes without additional tanks, 93.5 tonnes with one additional centre tank (ACT), and 97 tonnes with two and three ACTs. All show the three segments. The 97-tonne, three-ACT aircraft carries its structural maximum of about 23.5 tonnes to about 3,050 nm (A), trades payload for fuel at about 0.58 tonnes per 100 nm down to about 18.0 tonnes at 4,000 nm (B), and reaches a ferry range of about 5,120 nm (C). Each ACT costs 0.4 to 0.65 tonnes of structural payload, the weight of the tank and its systems, and adds about 2.5 tonnes of fuel and roughly 450 nm of ferry range. The manual notes that the curves are for information only and that approved values are in the operating manuals of each airline.

The model of Result 11.2 reproduces these curves closely. With contingency fixed at $\varepsilon = 0.05$ and the OEW of each variant inferred from its plateau (50.6 tonnes for the 93.5-tonne aircraft, rising to 52.1 tonnes with three ACTs), fitting the two remaining parameters to 38 points read from the five curves gives $\Lambda = 15{,}800$ nm and $r_0 = 3.2$

tonnes, with a root-mean-square error of 25 nm and a largest error of 64 nm. At a cruise speed of about 450 knots, that range factor corresponds to a ratio $(L/D)/c$ of about 35 hours, the order of magnitude expected of a modern narrowbody. The fitted slopes are close but somewhat flatter than the chart's: the model predicts 0.52 tonnes per 100 nm on the weight-limited segment of the three-ACT curve, against about 0.58 on the chart, and 3.2 tonnes per 100 nm just past the knee of the 93.5-tonne curve, falling to 1.6 near its ferry range, against roughly 4 and 2 on the chart.

Headline ranges deserve suspicion. The manual's introduction describes the A321LR as flying up to 4,000 nm with 206 passengers (Airbus 2024); the three-ACT curve offers about 18 tonnes at 4,000 nm, which implies about 87 kg per passenger including baggage, a light assumption. Network planners need the range with their own cabin, their own passenger and baggage weights, their airline's reserve policy and the winds on their routes.

11.5 The A321XLR on Montreal–Seville

Revision 33 of the A321 manual has no payload-range chart for the A321neo-XLR (Airbus 2024). It does give the XLR's weights and fuel capacity, and the model of Result 11.2, fitted on the A321neo, can turn them into an **illustrative** curve. The calculation below is this book's construction, not manufacturer or Air Canada data.

Winds and equivalent still-air distance A payload-range chart is drawn for still air. Real flights fly through wind, and over the North Atlantic the prevailing westerlies, strongest in winter, help eastbound flights and hinder westbound ones. A flight that covers a ground distance D along its track at true airspeed V with an average wind component w (positive for a tailwind) takes $D/(V + w)$ hours and burns the fuel it would burn flying $D\,V/(V + w)$ in still air. That quantity is the **equivalent still-air distance** (ESAD), and it is where the payload-range curve must be read:

$$\mathrm{ESAD} = D\,\frac{V}{V + w}. \tag{11.4}$$

WORKED EXAMPLE 11.2 An illustrative A321XLR mission to Seville

Distance. The great-circle distance from Montreal to Seville is 3,000 nm (5,556 km). Allowing 2 percent for routing, the track distance flown is about 3,060 nm. At 450 knots, Laurentian's scheduled block times of 7.0 hours eastbound and 7.9 hours westbound, less the 0.65 hours of taxi, climb and approach allowed in the block-time function of Chapter 3, imply average wind components of roughly 30 knots each way.

Aircraft. We take the manual's weight variant with MTOW 101 tonnes, MZFW 75.6 tonnes and MLW 79.2 tonnes, and its fuel capacity of 39,748 litres (31.2 tonnes) with the optional forward tank. The manual is not fully consistent about the rear tank, which its introduction gives as 12,900 litres and its fuel table as 13,100 litres; we use the table. It quotes no typical operating weight for the XLR, only reference weights for ground clearance (52.0 tonnes) and jacking (49.2 tonnes). We use 52.0 tonnes as a stand-in OEW for a 182-seat cabin and keep Λ and r_0 from the A321neo fit. With 95 kg per passenger including baggage (the typical mission rule shown in Belobaba's lecture (Belobaba 2006k)), 182 passengers weigh 17.3 tonnes.

The curve. Result 11.2 gives corner A at 3,709 nm with the structural payload of 23.6 tonnes, corner B at 4,841 nm with 17.8 tonnes and a ferry range of 6,100 nm. With 182 passengers and bags and no cargo, the still-air range is about 4,870 nm, close to the manual's headline of up to 4,700 nm (Airbus 2024), which is a sanity check rather than a validation. One more limit applies on the plateau: zero-fuel weight plus the reserves still on board at landing must stay below the MLW. With this model's reserves it would trim about 0.6 tonnes from the structural payload at A and 0.4 to 0.5 tonnes on the typical Seville missions below, but almost nothing on the strong-wind days; the payloads quoted below ignore it, so their cargo room is slightly optimistic.

The missions (Figure 11.4). Eastbound with a 30-knot tailwind, the ESAD is $3{,}060 \times 450/480 = 2{,}869$ nm; westbound against 30 knots it is $3{,}060 \times 450/420 = 3{,}279$ nm. Both lie on the plateau: the aircraft can carry its full structural payload, 182 passengers and 6.3 tonnes of cargo. On a strong winter day with a 90-knot average headwind, the westbound ESAD is $3{,}060 \times 450/360 = 3{,}825$ nm, past corner A, and the pay-

load limit falls to 23.0 tonnes. If winter weather also calls for 2 tonnes of extra fuel (holding, a more distant alternate, de-icing), the limit falls to 21.0 tonnes, leaving 3.7 tonnes for cargo. In this model the westbound payload limit starts to fall below the structural maximum, eating into cargo capacity, when the average headwind exceeds about 79 knots, or 37 knots with the extra 2 tonnes; passengers would be restricted only beyond 140 knots, which does not happen on this route.

A longer route. Montreal–Rome is 3,555 nm. Against a 90-knot headwind with 2 tonnes of extra fuel, its ESAD is about 4,530 nm, beyond the knee, and the payload limit is 16.0 tonnes: 1.3 tonnes, about 13 passengers with their bags, below a full cabin.

The planner's lesson is that range is not a circle on a map. A route well inside the headline range can still face seasonal payload limits, they appear westbound in winter, and because the curve steepens more than threefold past the knee on this XLR curve (about sixfold on the standard A321neo), a route whose winter westbound ESAD sits near corner B is fragile. The handbook asks for exactly this analysis: separate eastbound and westbound performance assessments by season, covering payload, baggage and cargo, winds and routing, reserves, runway conditions, alternates, diversion requirements and turnaround (Network Planning Handbook 2026).

Weight is not the only limit on cargo. The XLR's rear centre tank occupies the front of the aft hold, so the aircraft has eight container positions, or six with the optional forward tank, against ten on a standard A321neo (Airbus 2024). With 182 passengers' baggage on board, cargo is often limited by volume before weight. Industry analysis lists the loss of belly cargo, together with passengers' preference for widebodies, as the main trade-offs of the type (Cirium 2026b).

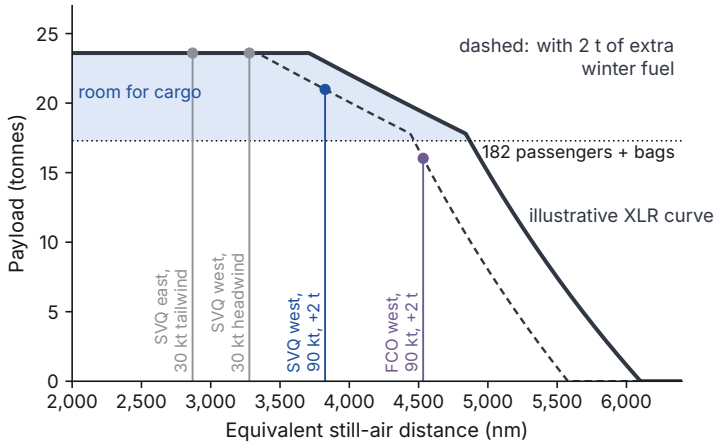


Figure 11.4 An illustrative A321XLR payload-range curve, this book's construction and not manufacturer data, since the manual has no XLR chart (manual weights and fuel capacity, Breguet model fitted to the A321neo charts, OEW assumed at 52.0 tonnes), with Montreal–Seville and Montreal–Rome missions at their equivalent still-air distances. The dashed curve carries 2 tonnes of extra winter fuel; the shaded band is the payload available for cargo above 182 passengers and their bags (17.3 tonnes).

PITFALL

Reading range from a circle. A route inside the manufacturer's headline range is not necessarily feasible with a full cabin in every season. Ask for the range with the airline's cabin, passenger and baggage weights and reserve policy, read at the westbound winter ESAD, and check how far that point sits from the knee of the curve.

Extended diversion operations A twin-engine aircraft crossing the North Atlantic spends part of its flight far from any airport where it could land after an engine failure; between Newfoundland and the Azores, for example, lie about 1,200 nm of ocean. Such flights fall under the rules for **extended diversion time operations** (EDTO), the current term for what was long called ETOPS, extended-range twin-engine operations. They apply when an aircraft is planned to fly farther than a threshold time from an adequate en-route alternate airport, and they impose layers of approval that a planner must not confuse: the aircraft type and engine combination must be approved for the diversion time, the operator must be approved by its regulator, and each operation must be planned with suitable en-route alternates within the approved diversion time and fuel to reach them.

The handbook notes that Canadian guidance now uses EDTO terminology and refers to Transport Canada's TP 6327, and warns against equating the certification of an aircraft type with the approval of a particular operator and route (Network Planning Handbook 2026). For a network planner, EDTO shapes routings and fuel, and so payload, and its feasibility check belongs to flight operations.

11.6 Financing and commonality

An airline can pay for aircraft with cash and retained earnings, with debt, with equity, or by leasing (Belobaba 2006k). The lecture notes that leasing can cost more but buys flexibility: an operating lease (typically three to seven years in 2006) lets an airline introduce a new type, renew its fleet more often or exit a type without selling aircraft, and it needs less capital up front. Up-front costs of a purchase include spare engines and parts, ground equipment and training as well as the aircraft itself. Air Canada's XLR program shows that the two routes are often combined: of its 30 A321XLRs, 15 are leased and 15 bought from Airbus, with purchase rights for 10 more between 2030 and 2032 (Air Canada 2026e). The same MD&A reports CAD 501 million of sale-and-leaseback proceeds in the first half of 2026, a third route by which owned aircraft become leased ones.

RESULT 11.3 Break-even lease rent

Compare buying an aircraft at price P, holding it T years and selling it for residual value V_T, with leasing it for T years at monthly rent L paid in advance and returning it at a cost E for return conditions. With annual discount rate ρ, discount factor $\nu = (1+\rho)^{-1}$ and monthly rate $i = (1+\rho)^{1/12} - 1$, the present costs are

$$\mathrm{PV}_{\mathrm{buy}} = P - V_T\,\nu^T, \qquad \mathrm{PV}_{\mathrm{lease}} = L\,\ddot{a} + E\,\nu^T,$$

where $\ddot{a} = (1+i)(1-(1+i)^{-12T})/i$ is the value of a monthly annuity of one paid in advance. Leasing is cheaper if and only if

$$L < L^{*} = \frac{P - (V_T + E)\,\nu^T}{\ddot{a}}.$$

The proof is the comparison of the two present values. Operating costs and maintenance reserves are assumed equal under both op-

tions and taxes are ignored; in practice both matter, and so does the value of flexibility.

WORKED EXAMPLE 11.3 Leasing or buying an A321XLR

Take, for illustration, a price of CAD 90 million, a 12-year holding period, a residual value of 45 percent (CAD 40.5 million), return conditions costing CAD 3 million and a discount rate of 8 percent. Then $v^{12} = 0.3971$ and $\ddot{a} = 94.31$. Buying costs $90 - 40.5 \times 0.3971 = 73.92$ million in present value. A lease at CAD 0.68 million a month costs $0.68 \times 94.31 + 3 \times 0.3971 = 65.32$ million, so leasing is cheaper by CAD 8.6 million. The break-even rent is $(73.92 - 1.19)/94.31 = 0.771$ million a month, a lease rate factor of 0.86 percent of the price per month.

Table 11.3 Break-even monthly lease rent L^* (CAD millions) for the illustrative A321XLR of Example 11.3. Leasing is cheaper when the offered rent is below the entry.

Discount rate	Residual 35%	Residual 45%	Residual 55%
6%	0.702	0.658	0.615
8%	0.809	0.771	0.733
10%	0.917	0.884	0.851

Table 11.3 shows how the answer moves. At a 6 percent discount rate the same CAD 0.68 million rent is above the break-even of 0.658 million and buying is cheaper by CAD 2.2 million. A higher expected residual value favours buying; a higher discount rate favours leasing. Leasing is attractive when the lessor's cost of capital is below the airline's, or when the lessor is more optimistic about residual values, and it transfers residual-value risk away from the airline. This is the lecture's warning about discount rates in concrete form: the choice of ρ can decide the comparison.

Commonality **Fleet commonality** is the sharing of type ratings, spare parts, tooling, maintenance programs and crew pools across an airline's aircraft. It reduces the costs of training, ground equipment and spare parts inventory that come with each new type (Belobaba 2006k). Some of these costs are one-off, but commonality also has a lasting benefit that is easy to quantify: pooling.

RESULT 11.4 Spares and the size of a subfleet

Suppose each of n aircraft of a type is unavailable on a given day with probability p, independently. To operate all $n - s$ scheduled lines with probability at least α, the subfleet needs s spare aircraft, where s is the smallest integer with $\Pr(B \le s) \ge \alpha$ for $B \sim \text{Bin}(n, p)$. For large n, $s \approx np + z_\alpha \sqrt{np(1-p)}$, so the spare ratio $s/n \approx p + z_\alpha \sqrt{p(1-p)/n}$ falls toward p, its excess over p shrinking as $1/\sqrt{n}$.

The first statement is the definition of the binomial quantile; the second is the normal approximation to the binomial. With $p = 5$ percent and $\alpha = 95$ percent, a subfleet of 4 aircraft needs 1 spare (25 percent of the subfleet), 8 aircraft need 2 (25 percent), 30 need 4 (13 percent) and 60 need 6 (10 percent). A small subfleet of a unique type carries a heavy spare burden, while a type that shares spares and crews with a larger family draws on a bigger pool. The same square-root logic applies to reserve crews and spare engines.

11.7 Turnaround time and utilization

Time on the ground is time not flying. Chapter 3 showed that utilization rises when turns are shorter, in whole departures. The A321 manual shows what determines turn time (Airbus 2024).

Table 11.4 Turn round times in the Airbus A321 manual (revision 33, sections 5-2-0 and 5-3-0), under the manual's assumptions about rates and equipment. Trucks are catering trucks; a full turn includes refuelling and all servicing, a transit turn uses stairs at two doors and no refuelling for the A321ceo/neo. Passenger flow means deplaning and boarding; catering in turn means one truck serving both galleys one after the other; rear catering means the second truck's work at the rear galley, between deplaning and boarding.

Aircraft	Turn	Seats	Equipment	Minutes	Critical path
A321ceo/neo	full	185	1 bridge, 1 truck	52	catering in turn
A321neo-ACF	full	202	1 bridge, 1 truck	51	catering in turn
A321neo-XLR	full	206	1 bridge, 2 trucks	41	passengers, rear catering
A321neo-XLR	full	244	1 bridge, 2 trucks	43	refuelling
A321ceo/neo	transit	220	2 stairways	23	passenger flow
A321neo-ACF	transit	202	2 stairways	38	catering in turn

A turn is a small project. Its activities (deplaning, cleaning, catering, cargo handling, refuelling, water servicing, boarding) run partly in parallel, and the turn time is the length of the longest chain of dependent activities, the **critical path**. In the manual's charts, passengers leave through one door at 20 a minute and board at 12 a minute, so when passenger flow is on the critical path each additional seat adds $1/20 + 1/12 = 0.133$ minutes, about 8 minutes per 60 seats. The 41-minute XLR turn of Figure 11.5 beats the 52 minutes of the standard A321 mainly because two catering trucks serve the front and rear galleys at the same time instead of one truck serving them in turn; refuelling with passengers on board keeps the XLR's 32,450-litre uplift off the critical path. With 244 seats, refuelling itself becomes critical and the turn takes 43 minutes. Boarding by stairs through two doors, without refuelling, brings a transit turn down to 23 minutes. The manual stresses that actual times depend on each operator's practices, resources, equipment and conditions (Airbus 2024), and airlines schedule standard turns longer than these technical minimums to absorb variability (Chapter 16).

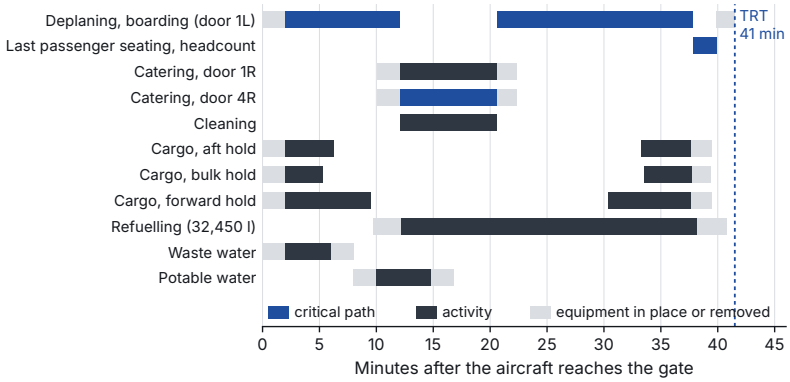


Figure 11.5 Full-servicing turn round of the A321neo-XLR with 206 seats and containerized cargo, redrawn from the Airbus A321 manual (revision 33, figure 5-2-0-991-011). The critical path runs through deplaning, rear-galley catering, boarding and the final headcount; refuelling, at 26 minutes, ends at about the time boarding does.

WORKED EXAMPLE 11.4 What eleven minutes are worth

An A321 shuttles 1.5-hour sectors within an 18-hour operating day. By Result 3.3, with 52-minute turns it can fly $\lfloor (18 + 0.867)/(1.5 + 0.867) \rfloor = \lfloor 7.97 \rfloor = 7$ sectors, 10.5 block hours. With 41-minute turns it flies $\lfloor 18.683/2.183 \rfloor = \lfloor 8.56 \rfloor = 8$ sectors, 12.0 block hours: 14.3 percent more utilization and 12.5 percent less ownership cost per block hour. With two-hour sectors the same saving changes nothing, because six sectors fit either way. Turn time matters in whole departures.

On long-haul flying the technical turn is rarely the constraint. Laurentian's Seville rotation leaves Montreal at 20:30, spends 7.0 hours to Seville and 3.25 hours on the ground there, and returns in 7.9 hours: the aircraft is back 18.2 hours after it left, with 5.8 hours to spare before another 20:30 departure, the arithmetic the handbook works through for the published schedule (Network Planning Handbook 2026). The schedule (departure banks, arrival windows, night restrictions) sets the turn, not the catering trucks. A daily rotation would give 14.9 block hours a day; at four flights a week, the aircraft flies only 8.5 block hours a day unless it is paired with another rotation on the other three days. Weekly patterns, not minutes on the ramp, decide long-haul utilization, and the fleet count with them.

11.8 A Laurentian fleet decision

Laurentian's fleet team must equip a program of thin transatlantic routes from Montreal: Seville, Lisbon, Madrid and Barcelona, between 2,827 and 3,192 nm away and well inside the XLR's reach on the analysis of Section 11.5. Two proposals are on the table, with roughly equal capital: four A321XLRs flying each market daily, or two 787-9s flying each market three or four times a week.

WORKED EXAMPLE 11.5 Four A321XLRs or two 787-9s?

Inputs. Demand per daily departure in each direction follows the case's Seville numbers by season: $\mathcal{N}(175, 42^2)$ for 16 peak weeks, $\mathcal{N}(150, 40^2)$ for 14 shoulder weeks and $\mathcal{N}(110, 35^2)$ for 22 winter weeks. These week counts are a generic calendar for the four markets; Chapter 12 builds Seville's own calendar, in which autumn and spring are the strongest months. Lisbon, Madrid and Barcelona are assumed 10, 15 and 25 percent larger, with net fares of CAD 600, 630 and 640 against Seville's 620. Every passenger brings CAD 30 of ancillary revenue and costs CAD 50. Block times are calibrated to the case, $0.65 + d/875$ hours eastbound and $0.65 + d/764$ westbound for a distance of d km. Trip costs come from Chapter 3; cargo contributes CAD 3,500 per XLR departure (the case) and an assumed CAD 12,000 per 787-9 departure; ownership costs CAD 8.16 million a year per XLR and 19.2 million per 787-9 (the illustrative figures of Chapter 3). All inputs other than the case's are illustrative.

Frequency and demand. Under option A (four XLRs) each aircraft flies one rotation a day and each market is served daily. Under option B (two 787-9s) each aircraft also flies one rotation a day, so each market gets 3.5 flights a week. Fewer flights lose customers who cannot travel on the days offered. We assume weekly demand varies with weekly frequency f as $(f/7)^{\theta}$, with a **frequency elasticity** $\theta = 0.4$: halving frequency loses 24 percent of demand ($0.5^{0.4} = 0.76$), and each 787-9 departure carries the remaining demand of two days. We keep the k-factor unchanged, which favours daily service: pooling two days' demand would make each consolidated departure less variable, and if the two days were independent it would divide k by $\sqrt{2}$ and add about CAD 4 million a year to an option that flies each market 3.5 times a week.

One market, one season. Take Seville in the shoulder season, eastbound. Under A, demand per departure is $\mathcal{N}(150, 40^2)$, the expected load is 145.2 (79.8 percent) and the contribution is

$$600 \times 145.2 + 3{,}500 - (11{,}000 \times 7.0 + 3{,}500) = 10{,}120$$

per departure, CAD 70,800 a week. Under B, mean demand per departure is $2{\times}150{\times}0.758 = 227.4$ with standard deviation 60.6; the 298-seat aircraft carries 223.7 on average (75.1 percent) and contributes

$$600 \times 223.7 + 12{,}000 - (15{,}500 \times 7.0 + 7{,}000) = 30{,}720$$

per departure, CAD 107,500 a week from 3.5 departures.

The program. Summing over markets, seasons and directions gives Table 11.5. Option A contributes CAD 16.3 million a year before ownership and CAD −16.4 million after; option B contributes 37.9 million before ownership and −0.5 million after. On these numbers the 787-9s win by CAD 15.8 million a year.

Table 11.5 Annual contribution of three fleet options for Laurentian's thin transatlantic program (both directions, frequency elasticity $\theta = 0.4$; illustrative inputs). A: four A321XLRs, each market daily; B: two 787-9s, each market 3.5 times a week; C: two A321XLRs, each market 3.5 times a week.

CAD millions a year	Option A	Option B	Option C
Seville	−0.69	5.27	6.47
Lisbon	4.30	9.30	7.69
Madrid	6.07	11.11	8.34
Barcelona	6.59	12.19	7.74
Contribution before ownership	16.27	37.86	30.24
Ownership	32.64	38.40	16.32
Contribution after ownership	−16.37	−0.54	13.92
Departures a year	2,912	1,456	1,456
Passengers a year (thousands)	424	330	249
Load factor	80.0%	76.1%	93.9%

Where the difference comes from. Comparing A with B mixes two changes: fewer departures and bigger aircraft. Separate them with a third option, C: two XLRs flying the same 3.5 weekly pattern as B. Going from C to A doubles frequency with the same gauge; it gains demand ($2^{0.4} = 1.32$ times as many customers) but doubles trip costs and ownership, and

costs CAD 30.3 million a year. Going from B to C keeps frequency and swaps the gauge; the XLRs spill more (Seville in the shoulder season: expected load 174.0 on 182 seats, a load factor of 95.6 percent) and lose the widebody's cargo, but their trip cost is CAD 35,000 lower and two XLRs cost CAD 22 million a year less to own; C beats B by CAD 14.5 million. Option C, which nobody proposed, is the best of the three: CAD 13.9 million after ownership.

Sensitivity. Figure 11.6 varies the parameter that matters most. C beats B for any frequency elasticity above 0.19 and beats A for any value below 1.15; A beats B only above 0.56. With demand 30 percent higher in every market, B becomes best for elasticities below 0.57, because the XLRs then spill too much, and A becomes best above 0.67, leaving C a narrow middle band. Other changes do not alter the ranking: CAD 6,000 instead of 12,000 of cargo per 787-9 departure, fuel 30 percent dearer (with fuel taken as 40 percent of the cost per block hour), a 5 percent premium-cabin uplift on the 787-9's fares or CAD 40 a passenger of network contribution all leave C in first place. A Monte Carlo simulation with θ uniform between 0.2 and 0.8, demand and fuel price lognormal around their base values (standard deviations 15 and 20 percent) and 787-9 cargo uniform between CAD 6,000 and 14,000 ranks C first in 86 percent of draws, B in 10 percent and A in the remaining 4 percent; A beats B in 40 percent.

What the planner learns. The question as posed, four XLRs or two 787-9s, was the wrong question: the comparison mixed a frequency decision with a gauge decision, and the best answer was a plan nobody had proposed. The 787-9 wins on cost per seat and cargo, the XLR on trip cost and ownership; at equal frequency, the XLR's lower trip cost beats the 787-9's lower seat cost in markets of this size, exactly the lesson of Chapter 3. Whether to fly these markets daily depends on how much demand daily service earns, a parameter planners rarely know well and should estimate (Chapters 4 and 13). Two cautions remain. None of these figures counts network effects, feed and cannibalization at Montreal, which Chapter 12 adds; and a recommendation should state what would reverse it: here, demand 30 percent above the base case together with a frequency elasticity below about 0.57.

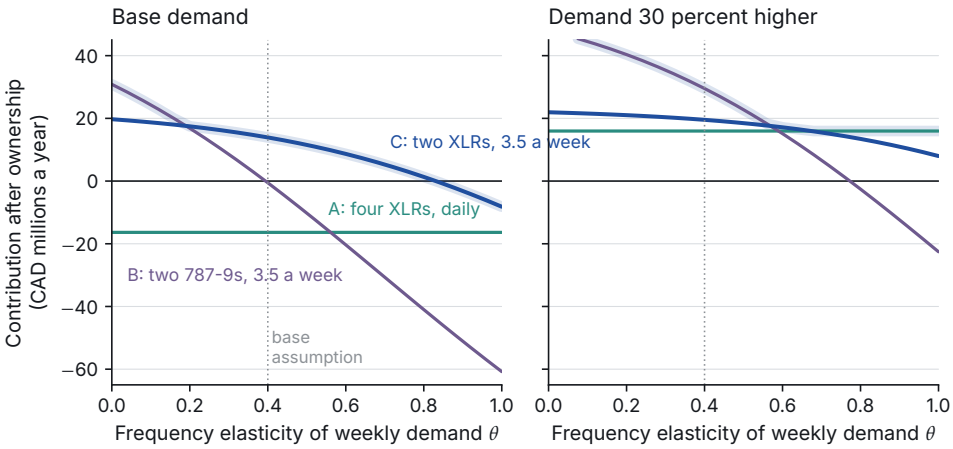


Figure 11.6 Annual contribution after ownership of the three fleet options as the frequency elasticity of demand varies, with base demand (left) and demand 30 percent higher in every market (right). The pale band marks the best option; the dotted line is the base assumption $\theta = 0.4$.

IN PRACTICE **Long-range narrowbodies in 2026**

Cirium counted 50 A321XLRs delivered and 44 in revenue service by the third quarter of 2026, with 28 airlines committed to 440 aircraft (Cirium 2026b). Cabins range from 138 seats at JetBlue to 239 at Wizz Air; Air Canada and Iberia fit 182. Cirium lists up to 30 percent lower fuel burn per seat than the 757 and access to markets too thin for widebodies as the type's advantages, and the loss of belly cargo and passengers' preference for widebodies as its trade-offs; it reports that United avoided some New York–Italy routes because of range (Cirium 2026b), the payload problem of Example 11.2. Air Canada's first XLR entered scheduled service on 9 June 2026 with 14 business and 168 economy seats, in a program of 30 aircraft (Network Planning Handbook 2026), and its announced uses favour moderate frequencies: Montreal–Seville from 8 September 2027 at three, then four, flights a week (Network Planning Handbook 2026), and for summer 2027 Toronto–Shannon three times a week and Montreal–Basel-Mulhouse four times a week (PAX News 2026). These schedules resemble option C of Example 11.5 more than option A.

11.9 Summary

- A fleet plan stages deliveries, leases, retirements and transfers under uncertainty. The evaluation chain runs from traffic forecasts

and a planning load factor to aircraft numbers, costs, profit and the balance sheet.

- The top-down model sizes the capacity gap from aggregate traffic, load factor, utilization and stage length. It is fast, but the gap is highly leveraged on planning assumptions (elasticity $-A_t/G_t$ with respect to the load factor), and it treats seat-kilometres as interchangeable.
- The bottom-up model allocates demand to future flights and chooses types and aircraft numbers route by route, with spill-adjusted contribution, aircraft time and reach constraints.
- Aircraft are judged on economics, range and payload, commonality, availability, financing, environment and regulation, and marketing; in 2026, availability (backlogs, engine reliability) is a first-order risk.
- The payload-range diagram is the minimum of a structural, a weight and a fuel limit. Past the knee, payload falls several times faster per mile than before it: four to six times on the A321neo curves. A two-parameter Breguet model reproduces the A321neo charts to within 65 nm.
- Range must be read at the equivalent still-air distance. On Montreal–Seville an illustrative XLR has a payload margin in both directions; winter westbound winds consume its cargo capacity first, and a route 20 percent longer would lose passengers on strong-wind days. Over the ocean, EDTO adds approvals of the aircraft type, the operator and each operation.
- Leasing beats buying when the rent is below L^*; the discount rate and the residual value decide. Commonality pools spares and crews: spare ratios fall with subfleet size, their excess over the daily unavailability rate shrinking with the square root of the size.
- Turn time matters in whole departures on short-haul flying; on long haul, schedules and weekly patterns set utilization.
- In Laurentian's fleet decision, two XLRs at 3.5 flights a week beat both four XLRs daily and two 787-9s across most plausible assumptions; the frequency elasticity of demand is the parameter that decides.

Table 11.6 Key formulas of the chapter.

Quantity	Formula
Capacity gap	$G_t = \mathrm{RPK}_t/\mathrm{LF}^{\mathrm{plan}} - \sum_k N_{kt} q_k$
Aircraft productivity	$q_k = 365\, U_k C_k s_{b,k}$
Gap leverage	$\partial \ln G_t / \partial \ln \mathrm{LF}^{\mathrm{plan}} = -A_t/G_t$
Breguet range	$R = \Lambda \ln(W_0/W_1)$, $\Lambda = V(L/D)/c$
Payload-range	$\min\{\mathrm{MZFW} - \mathrm{OEW},\ \text{weight limit},\ \text{fuel limit}\}$
Equivalent still-air distance	$\mathrm{ESAD} = D\,V/(V+w)$
Break-even lease rent	$L^* = (P - (V_T + E)\nu^T)/\ddot{a}$
Spare ratio	$s/n \approx p + z_\alpha \sqrt{p(1-p)/n}$

11.10 Further reading

Belobaba's lecture on fleet planning models (Belobaba 2006k) is short and sets out the evaluation process, the two families of models and the selection criteria used in this chapter; Chapter 7 of Belobaba, Odoni and Barnhart (2015) places fleet planning within the airline planning process. The Airbus A321 airport and maintenance planning manual (Airbus 2024) is the source of the weights, fuel capacities, payload-range curves and turn-round charts; read it with its own caveat that operators' approved data govern. The handbook (Network Planning Handbook 2026) frames the A321XLR's economics and the EDTO approval layers for a network planner, and Air Canada's MD&A (Air Canada 2026e) gives the current fleet plan and its financing. For the integration of fleet choice with schedule design, see Lohatepanont and Barnhart (2004).

CHAPTER

12

Route Evaluation

IN THIS CHAPTER

- Three views of route economics, and why one route can pass one view and fail another.
- Prorating a connecting fare by distance or by local fares, and why allocated revenue is not the marginal value of a leg.
- Feed, network contribution and displacement, and the costs a route decision can actually avoid.
- The opportunity cost of aircraft time as a shadow price, and how to avoid counting it twice.
- The handbook's Exercise A and Montreal–Seville case, reproduced and extended with spill-based loads by season.
- A full evaluation of Laurentian's YUL–SVQ by season and direction, with scenarios, a tornado chart, a sensitivity matrix and the threshold that reverses the decision.
- Ramp-up after launch, and the one-page memo that carries a recommendation.

DRAWS ON Belobaba, Route planning and evaluation (MIT 16.75J, 2006); handbook (route economics, Exercise A, the Montreal–Seville case, the decision memo); Chapter 9 spill model; Air Canada, Seville announcement (2026)

Sooner or later every network analyst is asked the same question about a route: should we open it, keep it, change it or drop it? The question sounds like a profit calculation, and every airline has a report that prints a profit for each flight. This chapter explains why that number cannot answer the question on its own. A flight's revenue depends on how connecting fares are split among legs; its costs depend on the horizon of the decision; its passengers may be taken from other flights or may push other passengers off; and the aircraft it uses could be doing something else. Belobaba states the economic criterion in one sentence: the potential for incremental profit in the short run, given the opportunity cost of taking the aircraft from another route (Belobaba 2006i). The handbook states

the planner's core decision in another: choose the feasible schedule and aircraft allocation that creates the greatest incremental network value, compared with the best realistic alternative (Network Planning Handbook 2026).

This chapter turns those two sentences into a method. It draws on almost everything before it: the costs of Chapter 3, the demand of Chapter 4, the spill model of Chapter 9 and the network effects of Chapter 10. It ends with a complete evaluation of Laurentian's Montreal–Seville route, the case that runs through the book, and with the memo that carries the recommendation. Fleet assignment (Chapter 14) and network revenue management (Chapter 20) provide the shadow prices and bid prices that this chapter uses as inputs.

12.1 Three views of route economics

Belobaba summarizes the measurement problem in two observations. Costs are driven by the fleet and the schedule and are incurred per flight and per network; revenues are driven by O&D markets and earned per itinerary (Belobaba 2006i). There are therefore two ways to measure a flight's profit. The ideal is to add, change or remove the flight, re-optimize the rest of the network, and measure the difference; this captures every interaction but requires a model of the whole operation. The common alternative is to allocate all revenues and costs to flights and treat each flight as independent; it is easy to explain but the allocation is always subjective and the network effects, which are important in most cases, disappear (Belobaba 2006i).

The handbook resolves the tension by asking for both, and one more.

DEFINITION 12.1 Three views of route economics

1. **Flight contribution**: the revenue allocated to the flight under a stated convention, minus its costs under a stated definition (avoidable, or including ownership, or fully allocated).
2. **Incremental network contribution**: the change in the whole network's revenue minus avoidable cost when the route is added or removed, relative to a dated baseline, counting feed, displacement and traffic taken from the airline's other services.
3. **The capital and strategic case**: the longer-term view that adds aircraft ownership, launch and station costs, fleet commitments and strategic effects such as competitive presence.

A route can pass one view and fail another (Network Planning Handbook 2026).

The lecture's sample network, taken from an article by Baldanza, makes the point with one flight (Belobaba 2006i). A leg from Syracuse to Omaha carries local passengers and feeds two onward flights, Omaha to San Francisco and Omaha to Los Angeles. Its local O&D revenue is USD 6,000; connecting revenue prorated to the leg is USD 1,500, and the share prorated to the onward legs USD 4,000. Variable operating costs are USD 4,500, aircraft ownership USD 2,000 and allocated overhead and non-operating costs USD 1,500. The network bears USD 700 of extra variable costs for the connecting passengers and USD 500 of opportunity cost for seats beyond Omaha that other passengers could have used. Table 12.1 combines these into nine measures.

Table 12.1 Nine profitability measures for one Syracuse–Omaha flight (adapted from Belobaba, 2006, who draws on Baldanza). Historical teaching numbers.

Measure	Calculation	USD
Variable leg profitability	6,000 + 1,500 – 4,500	3,000
Plus aircraft ownership	3,000 – 2,000	1,000
Fully allocated leg profitability	1,000 – 1,500	(500)
Variable, with network contribution	3,000 + 4,000 – 700	6,300
Variable, network contribution and opportunity cost	6,300 – 500	5,800
Ownership and network contribution	6,300 – 2,000	4,300
Ownership, network contribution and opportunity cost	4,300 – 500	3,800
Fully allocated, with network contribution	4,300 – 1,500	2,800
Fully allocated, network contribution and opportunity cost	2,800 – 500	2,300

The same flight earns anywhere from a loss of USD 500 to a profit of USD 6,300. None of these numbers is wrong; each answers a different question. Belobaba's lecture matches measures to decisions (Belobaba 2006i). For short-term schedule changes, ownership and overhead are fixed and the flight needs its network contribution to be judged at all, so the variable measure with network contribution applies. In the middle term aircraft can be moved between routes, so ownership enters. To judge a hub's profitability for a month, flights should be measured without network contribution, because adding every flight's network contribution counts the same connecting revenue twice. To judge whether a hub should exist at all, everything is variable, so the fully allocated measure applies.

A route's profit is not a property of the route. It is the answer to a question about a decision, and the decision fixes the revenue convention, the cost horizon and the alternative.

12.2 Revenue allocation and prorate

The handbook's rule for revenue is to follow the journey first and allocate second (Network Planning Handbook 2026). Start from passenger journeys by O&D, cabin, travel period and point of sale; forecast the net revenue the airline keeps, after partner settlements or

joint-venture treatment; leave out taxes collected for governments, which are not airline revenue; and agree with Finance and revenue management which ancillary, loyalty, surcharge and cargo items are already included before adding any line. Only then, if a flight view is needed, split connecting revenue among legs.

DEFINITION 12.2 Prorate

Let product j be sold at fare f_j on an itinerary that uses the set of legs $\ell(j)$. A **prorate** assigns to each leg $l \in \ell(j)$ a share $f_{jl} = w_{jl} f_j$ with weights $w_{jl} \geq 0$ that sum to one, so allocations add up to the fare. Two conventions are common:

- **mileage prorate**: $w_{jl} = d_l / \sum_{k \in \ell(j)} d_k$, where d_l is the distance of leg l;
- **straight-rate prorate**: $w_{jl} = f_l^{\text{loc}} / \sum_{k \in \ell(j)} f_k^{\text{loc}}$, where f_l^{loc} is a reference local fare on leg l, here the full unrestricted fare.

Between airlines, the same idea is applied through prorate agreements; inside one airline it is a reporting convention to agree with Finance.

Belobaba's lecture names both bases, distance and the ratio of full fares, and the option of assigning the whole connecting fare to the leg being analyzed (Belobaba 2006i). Each choice tells a different story about the same passenger.

WORKED EXAMPLE 12.1 Three ways to split a Boston–Paris fare

In Laurentian's case data a restricted Boston–Paris ticket over Montreal costs CAD 599 and a flexible one CAD 1,349. The legs are 254 miles (Boston–Montreal) and 3,432 miles (Montreal–Paris); the full local fares are CAD 389 and CAD 1,149.

Mileage prorate gives Boston–Montreal $254/3{,}686 = 6.9$ percent of the fare: CAD 41 of the restricted fare and CAD 93 of the flexible one. **Straight-rate prorate** gives it $389/1{,}538 = 25.3$ percent: CAD 152 and CAD 341. On the mileage convention the feeder earns less from a restricted Boston–Paris passenger than a quarter of its own cheapest local fare (CAD 189), so a report built on mileage prorate will show Boston–Montreal as a weak flight however much feed it supplies.

Network value. Chapter 20 shows how revenue management prices a seat on each leg with a bid price π_l, the opportunity cost of selling it. Solving the deterministic network program for Laurentian's four-leg bank at mean demand gives bid prices of CAD 109 on Boston–Montreal and CAD 490 on Montreal–Paris. The restricted Boston–Paris fare is then exactly marginal: $599 - 109 - 490 = 0$, which is why the program carries only part of that demand. The flexible fare leaves $1{,}349 - 599 = 750$ dollars of network value after paying for both seats. Viewed from one leg, the flexible passenger is worth the fare minus the other leg's bid price: $1{,}349 - 490 = 859$ to Boston–Montreal and $1{,}349 - 109 = 1{,}240$ to Montreal–Paris. Those two numbers add up to 2,099, far more than the fare. They are not an allocation, and they are the numbers that matter for decisions about one leg at a time.

The last observation is general. A prorate always adds up to the fare because that is its definition; marginal values need not add up to anything.

RESULT 12.1 Marginal leg values do not add up

Let $\mathcal{L}$ be the set of legs and $\Pi(A)$ the maximal network contribution achievable with the legs in $A \subseteq \mathcal{L}$, re-optimizing passenger flows. Define the marginal value of leg l as $\Delta_l = \Pi(\mathcal{L}) - \Pi(\mathcal{L} \setminus \{l\})$. If legs are complements, in the sense that Π is supermodular, then

$$\sum_{l \in \mathcal{L}} \Delta_l \;\geq\; \Pi(\mathcal{L}) - \Pi(\emptyset),$$

and if they are substitutes (submodular Π) the inequality is reversed.

Proof. Number the legs $1, \dots, L$ and let $A_i = \{1, \dots, i\}$. Telescoping, $\Pi(\mathcal{L}) - \Pi(\emptyset) = \sum_{i=1}^{L} [\Pi(A_i) - \Pi(A_{i-1})]$. Supermodularity means increasing differences: adding leg i is worth at least as much to a larger set of legs. Since $A_{i-1} \subseteq \mathcal{L} \setminus \{i\}$, each term satisfies $\Pi(A_i) - \Pi(A_{i-1}) \leq \Pi(\mathcal{L}) - \Pi(\mathcal{L} \setminus \{i\}) = \Delta_i$. Summing gives the result; submodularity reverses every inequality. ∎

The simplest case shows the size of the gap. If a single itinerary with fare f uses two legs and nothing else, removing either leg loses

the whole fare, so $\Delta_1 = \Delta_2 = f$ and the marginal values sum to $2f$. If instead two parallel flights each could carry the same passengers, removing one loses little because the other recaptures them, and the marginal values sum to less than the total. Hub networks are full of both relationships: feeders and long-haul flights are complements, parallel frequencies are substitutes. Any rule that splits revenue so that it adds up, whether by distance, local fares or a cooperative-game formula, is a convention for reporting, not a measure of what a leg is worth. This is the formal content of the handbook's warning that revenue allocated to a leg is not the marginal value of that leg (Network Planning Handbook 2026).

12.3 Feed, network contribution and displacement

Belobaba's lecture lists what a leg adds to and takes from the rest of the network (Belobaba 2006i). It adds the revenue earned on other legs by passengers who travel only because this leg exists. It takes the cost of handling its connecting passengers at the hub, the extra passenger costs on the connecting flights, and the opportunity cost of selling seats beyond the hub that travellers from other markets could have used, which he calls **network displacement cost**.

> **DEFINITION 12.3 Network contribution and displacement**
>
> The **network contribution** of a leg is the contribution it generates on other legs: revenue from connecting passengers allocated elsewhere, net of their passenger-related costs. Its **displacement cost** is the contribution lost on other legs because its passengers take seats that other travellers would have bought. Both are measured against a baseline network without the leg.

The handbook turns this into one comparison, computed across every affected flight and journey (Network Planning Handbook 2026). For a scenario s and baseline b, with retained passenger revenue R, other incremental revenue O (cargo, ancillaries) and avoidable operating cost AC,

$$\Delta\mathrm{NC} = \sum_{\text{flights, journeys}} \left[(R+O-\mathrm{AC})_s - (R+O-\mathrm{AC})_b \right] - F_{\text{launch}}, \quad (12.1)$$

after which incremental fixed and launch costs F_{launch} and longer-horizon capital requirements are added as the decision requires.

Two habits keep (12.1) honest. Count every passenger once, at the revenue the network actually keeps: the handbook's Exercise A ends with the warning not to prorate a connecting ticket and also count its full value elsewhere (Network Planning Handbook 2026). And do not add a generic feed uplift to the output of a network simulator that already models connections and recapture; the feed is already in it (Network Planning Handbook 2026).

WORKED EXAMPLE 12.2 What the Seville leg is worth to Laurentian's bank

Return to the four-leg bank of Example 10.2, with Quebec City and Boston feeding Paris and Seville. A deterministic network program, which chooses how many passengers of each fare product to carry at mean demand subject to every leg's capacity (Chapter 20), earns CAD 365,949 per bank with the Seville flight and CAD 245,499 without it. The Seville leg is worth CAD 120,450 of revenue to the bank.

Where does the number come from? The passengers of the Seville flight pay CAD 121,709 in O&D fares: 81,388 on the leg's own fare products and 40,321 from Quebec City and Boston. But their arrival changes who else flies. With Seville in the network, both feeders become full and the program drops 16 restricted Boston–Paris passengers (CAD 9,584), 3 restricted Quebec City–Paris passengers (CAD 1,947) and one Quebec City local (CAD 159). The Paris seats they free are refilled by 19 restricted Montreal–Paris locals who were spilled before (CAD 10,431). The net displacement is CAD 1,259, and $121{,}709 - 1{,}259 = 120{,}450$.

Allocations get close to this number only by accident. Under mileage prorate the Seville leg is credited with CAD 119,421, because it keeps 93 to 96 percent of every connecting fare; under straight-rate prorate it is credited with CAD 112,620. The differences, CAD 1,029 and CAD 7,830, are the network contribution net of displacement that a flight report on each convention leaves out.

Two cautions. The program is deterministic: it treats every product's demand as exactly its mean, so it understates displacement. Example 10.2 found that random demand makes the two feeders turn away about 24 more travellers per bank once Seville is added. And these are fare revenues before costs;

the full evaluation below works with net retained revenue and avoidable costs.

12.4 Costs a route decision can avoid

On the cost side the rule is the one Chapter 3 develops: count the costs that differ between the scenario and its baseline over the horizon of the decision. The handbook's cost table (Network Planning Handbook 2026) classifies them by what changes and when (Table 12.2).

Table 12.2 Cost buckets and their treatment in route decisions (adapted from the handbook).

Cost bucket	Typical drivers	Treatment in a route decision
Fuel	Aircraft, payload, winds, routing	Use inputs for the direction and season
Crew	Duty time, pairings, overnights, bases	Small schedule changes can create a discrete extra crew or hotel cost
Airport and navigation	Movements, weight, passengers	Reconcile every charge, including passenger charges
Maintenance	Hours, cycles, checks	Separate short-run wear from scheduled events
Passenger service and distribution	Passengers, cabin, channel	Consistent with the revenue definition
Station and launch	Staff, contracts, marketing	Separate one-off launch, recurring station cost and conditional incentives
Aircraft ownership	Fleet commitments, leases	Often unavoidable in the short run; the aircraft time still has an opportunity cost
Overhead	Shared functions	Report for accountability, but do not treat as avoidable

The handbook draws the two symmetric conclusions (Network Planning Handbook 2026). An accounting loss is not a reason to cancel a flight if most of its costs are unavoidable and it contributes positively. Positive contribution is not a reason to launch one if a better use of the same aircraft exists or the longer-run costs are not covered.

Laurentian's trip cost model (Chapter 3) is of the avoidable kind: a cost per block hour, a cost per departure and a cost per passenger. For the A321XLR on Montreal–Seville these are CAD 11,000 per block hour, CAD 3,500 per departure and CAD 50 per passenger. The eastbound flight blocks 7 hours and the westbound 7 hours 55

minutes, the times implied by the published schedule when Seville is six hours ahead of Montreal (Network Planning Handbook 2026), so the westbound trip costs CAD 10,083 more:

$$\mathrm{TC}_{\mathrm{E}} = 80{,}500 + 50\,Y, \qquad \mathrm{TC}_{\mathrm{W}} = 90{,}583 + 50\,Y.$$

With CAD 620 of allocated net fare, CAD 30 of ancillary revenue and CAD 50 of passenger cost, each passenger contributes $r = 600$ dollars, and cargo's CAD 3,500 per departure happens to equal the departure cost. A flight therefore covers its avoidable cost at $77{,}000/600 = 128.3$ passengers eastbound and $87{,}083/600 = 145.1$ westbound, load factors of 70.5 and 79.7 percent.

12.5 The opportunity cost of aircraft time

The aircraft a new route uses is rarely idle in the baseline. If the route is approved, something else does not fly. That forgone contribution is the opportunity cost in Belobaba's criterion, and it is often the cost that decides.

DEFINITION 12.4 Shadow price of aircraft time

The **shadow price** ω of a fleet's aircraft time is the increase in the network's optimal contribution from one more hour of that fleet's aircraft time, in CAD per aircraft-hour. In a fleet assignment model it is the dual value of the fleet's availability constraint (Chapter 14).

Where does ω come from? The simplest model is a knapsack. A planner has H hours of a fleet's aircraft time in a week and candidate uses $m = 1, \dots, M$, each with contribution V_m and aircraft time h_m; let $x_m \in [0, 1]$ be the share of use m that is flown.

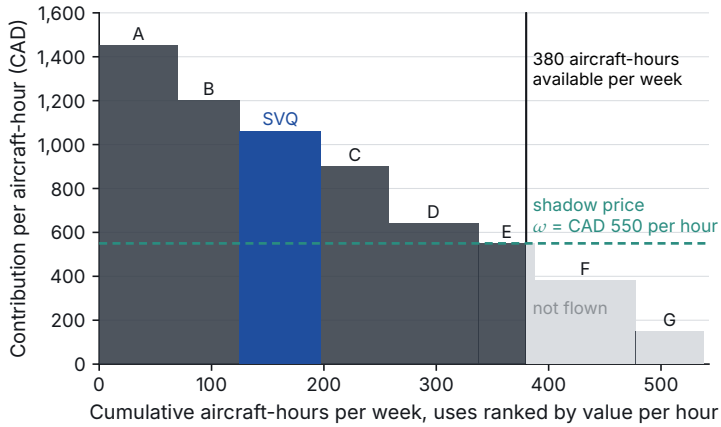


Figure 12.1 A merit order of candidate uses of Laurentian's A321XLR aircraft time in a spring or autumn week, ranked by contribution per aircraft-hour (uses A to G are illustrative). With 380 hours available, use E straddles the limit and sets the shadow price at CAD 550 per hour. Seville, at CAD 1,061 per hour, is comfortably inside the plan.

RESULT 12.2 The ratio rule for scarce aircraft time

The linear program

$$\max_{x} \sum_{m} V_m x_m \quad \text{s.t.} \quad \sum_{m} h_m x_m \le H, \quad 0 \le x_m \le 1$$

is solved by ranking uses by contribution per aircraft-hour, V_m/h_m, and accepting them in that order until the hours run out. The shadow price of aircraft time is the ratio of the use that straddles the limit, $\omega = V_k/h_k$, and use m is flown only if

$$V_m - \omega h_m \ \ge \ 0.$$

Proof. The dual is min $H\omega + \sum_m u_m$ subject to $h_m\omega + u_m \ge V_m$ and $\omega, u_m \ge 0$. Take $\omega = V_k/h_k$ for the straddling use k, and $u_m = [V_m - \omega h_m]^+$. This is dual feasible. The greedy primal solution sets $x_m = 1$ for uses with $V_m/h_m > \omega$, which are exactly those with $u_m > 0$, sets $x_m = 0$ for uses with $V_m/h_m < \omega$, whose dual constraints are slack, and uses up all H hours, which is consistent with $\omega > 0$. Complementary slackness holds, so both solutions are optimal (Appendix B), and the acceptance condition is the reduced cost $u_m \ge 0$. ∎

Figure 12.1 shows the rule as an economist would draw a supply curve. The shadow price is set at the margin by the least valuable use that still flies, not by the average use, and it rises in seasons when more valuable uses compete for the same aircraft. It is also a marginal value. Seville's four weekly rotations use 72.7 of the 380 hours, which is not marginal, and removing them would release hours that go first to the rest of use E at CAD 550 and then to use F at CAD 380. Re-solving the knapsack without Seville gives an exact opportunity cost of CAD 28,917 a week, or CAD 7,229 per round trip, against the screening charge $\omega h = 550 \times 18.17 = 9{,}992$ per round trip. For large blocks of aircraft time, re-solve the fleet model with and without the route; the shadow price is a screening device.

The handbook adds a warning that every planner needs once (Network Planning Handbook 2026). There are two legitimate ways to account for aircraft scarcity: compare explicitly modelled, mutually exclusive alternatives directly, or charge a shadow price for aircraft time in a screening model. Mixing them counts the opportunity cost twice.

> **PITFALL**
>
> Double counting the opportunity cost. In the handbook's Seville case the candidate contributes CAD 8,000 and the best feasible alternative for the same aircraft CAD 11,000, so the alternative is better by CAD 3,000. An analyst who first subtracts the alternative's 11,000 from the candidate as an opportunity cost (leaving minus 3,000) and then also compares the result with the alternative's 11,000 reports a gap of 14,000. One who nets each option against the other reports minus 3,000 against plus 3,000, a gap of 6,000. Both rank the two options correctly here, but both overstate how much is at stake, and either can distort the comparison once there are more than two options. Use one method per analysis and say which.

12.6 The handbook's exercises

Besides the spill exercise solved in Chapter 9, the handbook contains two worked exercises on route economics, both with invented numbers (Network Planning Handbook 2026). They are reproduced here, then extended with the spill model.

WORKED EXAMPLE 12.3 Exercise A: a full-looking route that is not the best use

One departure of a candidate flight carries 110 local passengers paying CAD 220 and 40 connecting passengers paying CAD 450 for their whole itineraries. Cargo and ancillaries contribute CAD 3,000 net. The flight's avoidable operating cost is CAD 32,000. Carrying the connecting passengers on existing feeder flights adds CAD 2,000 of operating cost there and displaces CAD 4,000 of other contribution, and passengers switching to the new flight from the airline's other services take CAD 3,500 of revenue with them. All figures are incremental except ticket revenue, whose baseline losses are deducted explicitly. Then

$$\begin{aligned}\text{gross ticket revenue} &= 110 \times 220 + 40 \times 450 = 42{,}200,\\ \Delta\text{NC} &= 42{,}200 + 3{,}000 - 32{,}000\\ &\quad - 2{,}000 - 4{,}000 - 3{,}500 = 3{,}700.\end{aligned}$$

An independently evaluated use of the same aircraft earns CAD 4,500, so the alternative is better by CAD 800 (Network Planning Handbook 2026).

Three lessons follow. A full flight of 150 passengers does not settle the decision. The next question is whether a different departure time could raise the candidate's contribution without needing another aircraft. And the CAD 450 connecting fares must enter once, in full, in a network comparison like this one, not prorated to the flight and counted again elsewhere. The CAD 3,500 deducted for switching passengers is revenue, not contribution, which is right if the services they leave keep operating unchanged.

The second exercise is the handbook's Montreal–Seville case, which uses the schedule of the publicly announced service and invented economics for one departure in a leg-allocation view, in which connecting revenue is split consistently across legs and later lines capture only changes elsewhere (Network Planning Handbook 2026). Table 12.3 reproduces it.

Table 12.3 The handbook's invented economics for one Montreal–Seville departure (adapted from the handbook). The numbers are illustrative and are not estimates for any airline.

Item	Assumption or calculation	CAD
Passengers	150 on a 182-seat aircraft, 82.4% load factor	
Net passenger revenue allocated to this leg	150×620	93,000
Ancillary revenue	150×30, not already in passenger revenue	4,500
Cargo contribution	Net of its incremental handling costs	3,500
Avoidable flight cost	Includes passenger-related costs at this load	(88,000)
Initial flight contribution	Revenue and contribution above, less flight cost	13,000
Additional contribution on affected other flights	Net of their incremental costs and any displacement	6,000
Lost contribution on own alternative services	Cannibalization not already captured above	(9,000)
Expected incremental disruption cost	Additional to costs already included	(2,000)
Incremental network contribution versus no new flight	Under this accounting convention	8,000

A feasible alternative using the same aircraft and time windows contributes CAD 11,000 against the same baseline, so the alternative is better by CAD 3,000: the candidate's positive contribution and its 82.4 percent load factor do not establish that it should be chosen (Network Planning Handbook 2026). Two further checks complete the exercise. Comparing 150 passengers at CAD 620 with 170 at CAD 530 gives passenger revenue of CAD 93,000 against CAD 90,100, despite the higher load: load factor alone cannot settle the decision. And the assumption that would reverse the ranking is worth CAD 3,000 per departure, which across 150 passengers is CAD 20 more retained fare each, holding demand and every network effect fixed. The handbook treats the CAD 20 as a threshold for discussion, not a pricing recommendation, because a higher fare would change demand and mix (Network Planning Handbook 2026).

The case assumes 150 passengers. A planner should ask where that number comes from, and Chapter 9 answers that it cannot be a forecast of mean demand: a flight whose demand averages 150 car-

ries fewer than 150 on average, because some departures spill. Laurentian's case data give the forecast behind the handbook's number.

WORKED EXAMPLE 12.4 The Seville case with spill-based loads

Laurentian forecasts unconstrained demand per eastbound departure of $\mathcal{N}(150, 40^2)$ in its shoulder season, $\mathcal{N}(175, 42^2)$ at its peak and $\mathcal{N}(110, 35^2)$ in winter, for 182 seats. The names describe demand levels, not months: Laurentian's forecast makes autumn and spring Seville's strongest months, so its calendar (Section 12.7) applies the peak level to autumn and spring and the shoulder level to summer. By the expected-load formula of Chapter 9, expected load is $\mu - \sigma L(z)$ with $z = (182 - \mu)/\sigma$.

Shoulder. $z = 32/40 = 0.80$, $L(0.80) = 0.2897 - 0.80 \times 0.2119 = 0.1202$, so expected spill is $40 \times 0.1202 = 4.8$ and expected load 145.2 (load factor 79.8 percent); 21 percent of departures leave full. **Peak.** $z = 0.167$, $L(0.167) = 0.321$, expected spill 13.5 and expected load 161.5 (88.7 percent); 43 percent of departures leave full. **Winter.** $z = 2.06$, expected spill 0.25 and expected load 109.7 (60.3 percent).

Keep the handbook's fares and cost model, and scale its network effects with the load: CAD 40 of feed contribution and CAD 60 of cannibalization per passenger (6,000 and 9,000 at 150 passengers), plus CAD 2,000 of disruption cost per departure. Each passenger contributes $620 + 30 - 50 = 600$ dollars and cargo offsets the departure cost, so the flight contribution is $600\,\mathbb{E}[Y] - 77{,}000$:

- shoulder: $600 \times 145.2 - 77{,}000 = 10{,}115$; network effects $-20 \times 145.2 - 2{,}000 = -4{,}904$; incremental network contribution CAD 5,211;
- peak: flight contribution 19,907; incremental 14,677;
- winter: flight contribution $-11{,}153$; incremental $-15{,}348$.

Against the CAD 11,000 alternative, the shoulder departure falls short by CAD 5,789 instead of 3,000, the peak departure beats it by CAD 3,677, and the winter departure falls short by CAD 26,348.

Two lessons follow. Treating mean demand as the load overstated the shoulder flight's contribution by 4.8 passengers, or CAD 2,885 per departure: the error of replacing demand by its

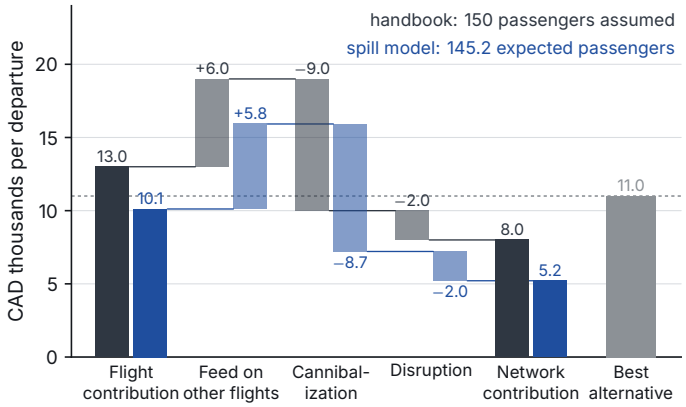


Figure 12.2 The handbook's Montreal–Seville bridge from flight contribution to incremental network contribution for one eastbound departure, at the handbook's assumed 150 passengers and at the spill-based expected load of 145.2 for the same demand forecast. Both fall short of the CAD 11,000 alternative; the spill correction widens the gap from CAD 3,000 to CAD 5,789.

mean, which Chapter 9 warns about, here nearly doubles the margin by which the route loses. And the season matters far more than the spill correction: the same route is a clear winner at its peak and a heavy loser in winter. Figure 12.2 draws the shoulder-season bridge both ways.

12.7 Laurentian YUL–SVQ: a full seasonal evaluation

A launch decision is about a schedule for a year, in both directions, against the best use of the aircraft in each period. This section builds that evaluation for Laurentian, step by step, with every assumption stated.

The real schedule. Air Canada announced Montreal–Seville on 14 October 2026: flight AC958 leaves Montreal at 20:30 and arrives in Seville at 09:30 the next day, AC959 leaves Seville at 12:45 and arrives in Montreal at 14:40, with an A321XLR of 14 business and 168 economy seats. Service starts on 8 September 2027 at three flights a week and rises to four from 1 October 2027. The published operating periods run from 8 September 2027 to 8 January 2028 and from 18 February to 3 June 2028 (Air Canada 2026g): an autumn-to-spring service with a mid-winter break. The release gives no reasons for

that pattern. Laurentian's case borrows the schedule and nothing else.

Laurentian's assumptions. Table 12.4 lists them. The demand, fares, costs and network effects come from the running case and the handbook's case; the demand calendar, the westbound demand, the scaling of network effects and the value of aircraft time are this chapter's illustrative assumptions. Laurentian's forecast puts the market's strongest demand in autumn and spring, a softer summer and a weak winter. Its fleet model values XLR time most in the northern summer, when every long-haul aircraft has strong uses, and least in winter.

Table 12.4 Assumptions of the Laurentian YUL–SVQ evaluation. Westbound demand is assumed equal to eastbound demand in each period.

Input	Value	Source
Aircraft, seats	A321XLR, 182	case
Block time, cycle	7 h 00 east, 7 h 55 west; 18 h 10 from departure to return	public schedule
Demand per departure (each direction)	$\mathcal{N}(175, 42^2)$ autumn and spring; $\mathcal{N}(150, 40^2)$ summer; $\mathcal{N}(110, 35^2)$ winter	case; calendar assumed
Calendar	Autumn 9 weeks (Sep–Oct), winter 21 (Nov–Mar), spring 9 (Apr–May), summer 13 (Jun–Aug)	assumed
Frequency	4 round trips a week	public schedule
Revenue	CAD 620 net fare, CAD 30 ancillary per passenger; CAD 3,500 cargo per departure	case
Avoidable trip cost	CAD 11,000 per block hour, 3,500 per departure, 50 per passenger	case
Network effects	feed contribution CAD 40 and cannibalization loss CAD 60 per passenger; disruption cost CAD 2,000 per departure	handbook, scaled
Value of aircraft time ω	CAD 550 per hour autumn and spring; 1,000 summer; 150 winter	assumed

Two of these inputs tie back to earlier chapters. The mean of 150, the case's shoulder level, which Laurentian's calendar applies to summer, is made up of about 82 local travellers, the nonstop's share of the Montreal–Seville market in Example 4.4, and about 68 connecting travellers from the seven markets behind the gateway (Example 10.2). And the cannibalization loss covers the travellers that Example 4.4 moves from Laurentian's own connection over

Paris, about 11 per departure: what Laurentian loses on them is their contribution on the Paris itinerary less the value of the Paris seats they free, which on an over-demanded flight is the displacement cost of Section 12.3, not zero.

Per round trip. For each period and direction, expected load comes from the spill model, the flight contribution from the trip cost model, and the incremental network contribution from the scaled network effects. Writing r for contribution per passenger, e_{f} and e_{c} for the feed and cannibalization effects per passenger, δ for the disruption cost, K for cargo and TC_d^0 for the trip cost of direction d before passenger costs, the value of one round trip against the best alternative use of the aircraft is

$$V = \sum_{d \in \{\mathrm{E},\mathrm{W}\}} \left[(r + e_{\mathrm{f}} - e_{\mathrm{c}})\, \mathbb{E}[Y_d] + K - \mathrm{TC}_d^0 - \delta \right] - \omega h, \qquad (12.2)$$

where $h = 18.17$ hours is the aircraft time the rotation occupies (the h_m of Result 12.2 for this use). Table 12.5 applies (12.2) period by period.

Table 12.5 Laurentian YUL–SVQ by period and direction, in CAD thousands. Load is expected passengers per departure; East and West are incremental network contributions per departure; ωh is the value of the aircraft time per round trip in its best alternative use; Net deducts it from the round trip. A mature year has 36 round trips in autumn, 84 in winter, 36 in spring and 52 in summer.

Period	Load	East	West	Round trip	ωh	Net	Net per year
Autumn	161.5	14.7	4.6	19.3	10.0	9.3	334
Winter	109.7	(15.3)	(25.4)	(40.8)	2.7	(43.5)	(3,654)
Spring	161.5	14.7	4.6	19.3	10.0	9.3	334
Summer	145.2	5.2	(4.9)	0.3	18.2	(17.8)	(927)

Figure 12.3 reads the table as a picture, and three conclusions stand out.

Direction. With identical demand in both directions, the westbound flight earns CAD 10,083 less than the eastbound one, the cost of 55 extra minutes of block time against the prevailing winds. A route judged on its eastbound departure, as the handbook's case is, looks better than its round trip. Note also that the block times hold only when Seville is six hours ahead: Europe and North America change their clocks on different dates, so from 31 October to 6 November 2027 and from 12 to 25 March 2028 the offset is five hours, and either the local times or the block times change (Chapter 13).

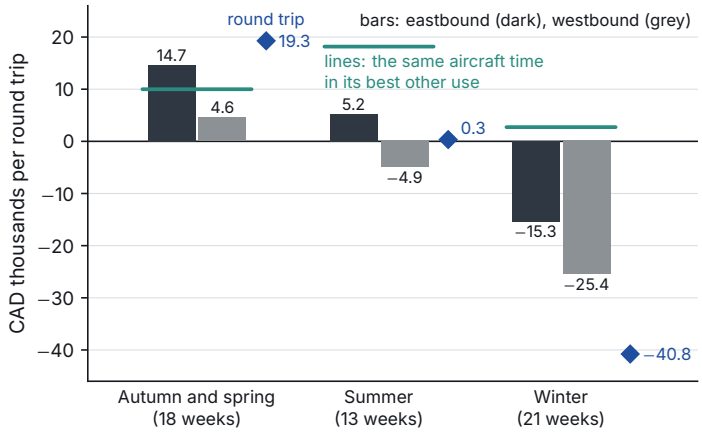


Figure 12.3 Incremental network contribution per round trip of Laurentian's YUL–SVQ by period and direction, against the value of the same aircraft time in its best alternative use (lines). Autumn and spring beat the alternative; summer covers its avoidable costs but not the value of the aircraft; winter loses money even before the aircraft's opportunity cost.

Season. Summer covers its avoidable costs (CAD 339 per round trip) but the aircraft is worth CAD 18,167 elsewhere: the route is profitable and still wrong. Winter loses CAD 40,779 per round trip before any opportunity cost.

Pattern. Totals over the year depend on which periods are flown (Table 12.6). The year-round average round trip loses CAD 9,713 against not flying at all; an attractive average can hide an unprofitable season, and here the average is not even attractive. Flying only in autumn and spring earns CAD 1.39 million against not flying and CAD 668,103 against the best alternative use of the aircraft.

Table 12.6 Annual value of four operating patterns for Laurentian YUL–SVQ, in CAD, in a mature year.

Operating pattern	Round trips	Against not flying	Against the best alternative
Year-round	208	(2,020,279)	(3,913,246)
Autumn, winter and spring (September to May)	156	(2,037,910)	(2,986,210)
Autumn, spring and summer	124	1,405,133	(258,933)
Autumn and spring	72	1,387,503	668,103

The recommended pattern is two seasons a year. Laurentian's winter assumption drives the difference with the real service's longer operating periods: with demand of 110 per departure against

182 seats, every winter week loses money, while an airline with stronger winter demand, holiday traffic or idle aircraft in winter could reasonably fly through most of it. A monthly rather than seasonal evaluation would sharpen the pattern, as the handbook recommends (Network Planning Handbook 2026), and so would the station and restart costs of suspending service twice a year, which the per-departure numbers leave out.

12.8 Scenarios, sensitivity and the threshold that reverses the decision

The handbook gives three rules for presenting uncertainty (Network Planning Handbook 2026). Use a base, an adverse and a favourable case with coherent assumptions, because lower demand tends to come with lower fares and higher fuel prices with other costs, and independently toggling every input produces impossible combinations. Keep structural uncertainty, forecast error and execution risk apart. And report the sensitivity that matters most: a matrix of demand capture against retained fare, with the fuel or block-time threshold that reverses the recommendation.

Coherent scenarios. In Laurentian's adverse case demand is 10 percent below forecast, the retained fare is CAD 600, the cost per block hour is 8 percent higher, and because the same conditions depress the aircraft's other uses, the value of aircraft time is 15 percent lower. In the favourable case demand is 8 percent higher, the fare CAD 640, block-hour cost 5 percent lower and aircraft time 10 percent more valuable. The recommended autumn-and-spring pattern is worth CAD 668,103 a year against the alternative in the base case, minus CAD 1.43 million in the adverse case (minus CAD 0.82 million even against not flying), and CAD 2.14 million in the favourable case, in which summer flying also beats its alternative and the autumn, spring and summer pattern is worth CAD 2.33 million. The downside is larger than the base-case gain: the route is a bet on fares and demand, and the memo must say so.

One input at a time. A tornado chart moves each input across a plausible range while holding the others at base (Figure 12.4). It is a diagnostic of which inputs deserve attention, not a scenario analysis.

Retained fare, the cost per block hour and the demand level dominate, and each can reverse the decision within its range; a 10 percent fare shortfall alone turns CAD 668,103 into a loss of CAD 774,000. The demand bar is asymmetric (minus 1.35 million, plus 0.78 million) because spill caps the upside: a 15 percent shortfall removes 16

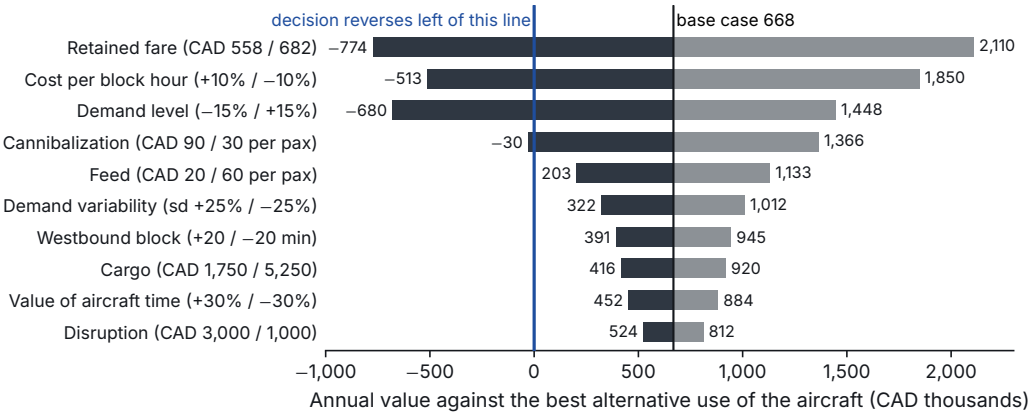


Figure 12.4 Tornado chart for the recommended autumn-and-spring pattern: its annual value against the best alternative use of the aircraft when one input moves across the range shown and the others stay at base. Bars that cross the accent line reverse the recommendation.

expected passengers per departure, while a 15 percent excess adds only 9, because 182 seats cannot carry the extra demand on the busiest days. Cannibalization comes next, and at CAD 90 per passenger it too reverses the decision, if only just: the network effects of Section 12.3 are not a rounding error. The value of aircraft time matters less than intuition suggests in this pattern, because autumn and spring are periods when the aircraft's alternatives are only moderately valuable.

The threshold. Setting (12.2) to zero and solving for the retained fare gives the fare below which the recommendation reverses,

$$f^{*}(\theta) = \frac{\sum_{d}(\mathrm{TC}_{d}^{0} + \delta - K) + \omega h}{\sum_{d} \mathbb{E}[Y_{d} \mid \theta]} - f_{\mathrm{anc}} + c_{\mathrm{pax}} + e_{\mathrm{c}} - e_{\mathrm{f}}, \qquad (12.3)$$

where θ is **demand capture**, the realized share of forecast demand (which scales both μ and σ), f_{anc} is ancillary revenue and c_{pax} the passenger cost. At the forecast ($\theta = 1$) in autumn and spring, the numerator is $171{,}083 + 4{,}000 - 7{,}000 + 9{,}992 = 178{,}075$ and the denominator $2 \times 161.5 = 323.0$, so $f^{*} = 551.3 - 30 + 50 + 60 - 40 = 591$ dollars: CAD 29, or 4.6 percent, below the planned fare. Because the fixed costs are spread over fewer passengers, the threshold rises at an increasing rate as demand falls: CAD 607 at 95 percent capture, 627 at 90 percent and 653 at 85 percent. At the planned fare, demand capture below 91.7 percent reverses the decision.

Figure 12.5 is the handbook's matrix. The same algebra gives the thresholds the handbook asks for on the cost side. Solving (12.2)

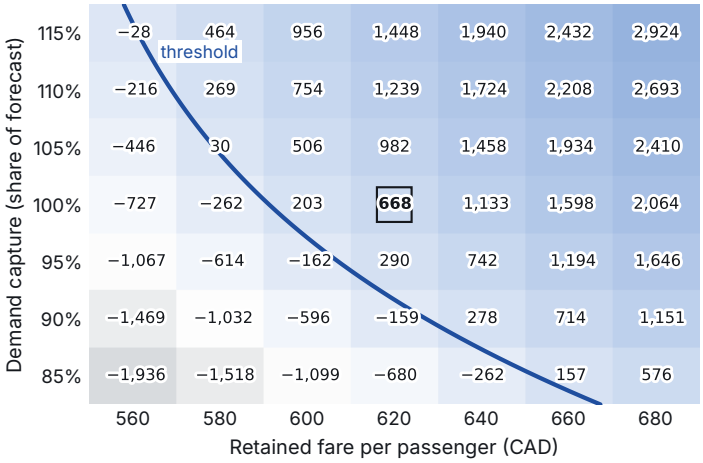


Figure 12.5 Two-way sensitivity of the autumn-and-spring pattern's annual value against the best alternative (CAD thousands) to demand capture and retained fare. The line is the threshold of (12.3); the base case is boxed.

for the block-hour cost gives CAD 11,622, 5.7 percent above plan, the kind of increase a fuel price movement can produce. Lengthening the westbound block, whether by winds or routing, adds both block-hour cost and aircraft time, and reverses the decision at 8 hours 43 minutes, 48 minutes more than planned. And if hot-day or winter-wind performance forces Laurentian to block seats on the westbound flight, the decision reverses at 29 blocked seats. Following the handbook, each of these is a threshold to put in front of the people who own the input (revenue management, fuel and Finance, flight operations), not a forecast.

12.9 Ramp-up after launch

The evaluation so far describes a mature year. A new route does not start mature: awareness, distribution, corporate contracts and partner feed take time to build, and the handbook asks for the ramp-up to be explicit, with airport incentives shown with their duration, conditions and post-expiry economics rather than embedded indefinitely in cost (Network Planning Handbook 2026).

A simple model multiplies forecast demand t weeks after launch by

$$\theta(t) = 1 - (1 - \theta_0)\, e^{-t/\tau},$$

starting at θ_0 of maturity and closing the gap with time constant τ. Laurentian assumes, for illustration, $\theta_0 = 0.80$ and $\tau = 12$ weeks, so

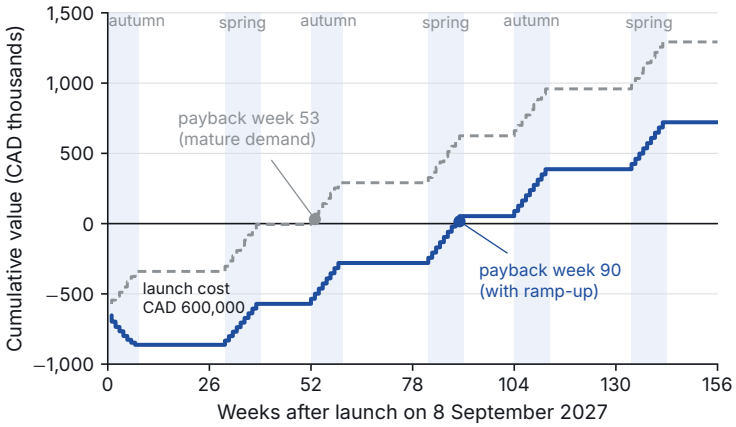


Figure 12.6 Cumulative value of the autumn-and-spring pattern against the best alternative use of the aircraft from launch on 8 September 2027, with mature demand from the first day (dashed) and with demand ramping up from 80 percent of the forecast (solid). Shaded weeks are the operating periods.

demand reaches 90 percent of the forecast after about two months and 98 percent by the first spring. It also assumes a one-off launch cost of CAD 600,000 for station set-up, marketing and training.

WORKED EXAMPLE 12.5 Paying back the launch

Laurentian's first season is short: from 8 September 2027 at three flights a week and four from October, 28 round trips before the autumn period ends. With mature demand those round trips would be worth CAD 259,818 against the alternative. With demand at 80 to 89 percent of the forecast, the first autumn instead loses CAD 262,697, because expected load falls with demand (by 14 percent at 80 percent of the forecast) while the costs and the value of the aircraft's alternatives do not. Spring 2028 starts at 98 percent of maturity. Including the launch cost, the mature path pays back in the 53rd week (September 2028); the ramp-up path pays back in the 90th week (late May 2029), and after three years it has accumulated CAD 0.72 million against CAD 1.29 million with mature demand (Figure 12.6).

The planner learns two things. The first season of a new route is a poor guide to its long-run economics, in both directions: a weak first autumn is expected, and a strong one may be stim-

ulation that fades. And the launch decision should be judged over a horizon long enough to include several mature seasons, which is the capital view of Definition 12.1.

The post-launch review closes the loop. The handbook recommends comparing outcomes with the forecast vintage that was approved, not a later revision, and bridging the difference into volume, fare, cabin and point-of-sale mix, cargo, cost, completion and disruption, separating launch ramp-up from a persistent forecast bias (Network Planning Handbook 2026).

12.10 The one-page decision memo

The output of all this work is a short document that a decision-maker can read in five minutes. The handbook's template for a route-decision memo lists its contents: the decision requested and its deadline; the baseline and alternatives; proposed dates, frequencies, fleet and timings; market and itinerary evidence; retained revenue by cabin and flow; cost and contribution definitions; network effects; aircraft, crew and airport feasibility; downside cases; the recommendation; and outstanding conditions with accountable owners (Network Planning Handbook 2026). It adds that a recommendation should state what would make it wrong, and that the right conclusion can be that the evidence is insufficient pending named inputs. Table 12.7 fills in the template for Laurentian.

Table 12.7 A one-page decision memo for Laurentian YUL–SVQ, following the handbook's template. All figures are illustrative.

Memo element	Laurentian YUL–SVQ
Decision and deadline	Approve YUL–SVQ for 2027–28 with the A321XLR before the schedule is published
Baseline and alternatives	Baseline: no Seville service, XLR time used by the fleet plan. Alternatives: year-round; autumn, spring and summer; autumn and spring
Schedule	YUL 20:30, SVQ 09:30 next day; SVQ 12:45, YUL 14:40. Block 7 h 00 east, 7 h 55 west, rotation 18 h 10. Three a week from 8 September 2027, four from October
Market evidence	Mean demand per departure 175 (autumn, spring), 150 (summer), 110 (winter); of the 150, about 82 local and 68 connecting from seven markets behind the gateway
Revenue and costs	CAD 620 net fare and CAD 30 ancillary per passenger, CAD 3,500 cargo per departure; avoidable cost CAD 11,000 per block hour, 3,500 per departure, 50 per passenger
Network effects	Feed adds CAD 40 and cannibalization costs CAD 60 per passenger; disruption costs CAD 2,000 per departure; feeder displacement to be confirmed with revenue management
Result	Autumn and spring: CAD 9,279 per round trip better than the alternative, CAD 668,103 a year. Summer: CAD 17,828 worse per round trip. Winter: CAD 43,504 worse
Downside	Adverse case CAD 1.43 million a year worse than the alternative; first season CAD 262,697 worse with ramp-up; payback in the 90th week
Recommendation	Launch for autumn and spring only; do not schedule summer or winter
What would make it wrong	Retained fare below CAD 591; demand capture below 91.7 percent; block-hour cost above CAD 11,622; 29 or more westbound seats blocked
Conditions and owners	Fare and mix (revenue management); westbound performance (flight operations); crew pairing and station costs (crew planning, airports); contribution definitions (Finance)

IN PRACTICE **Route evaluation tools in 2026**

The logic of this chapter is built into commercial planning systems. Lufthansa Systems describes its NetLine/Plan as producing O&D-based flight and network profitability by chaining demand and market-share models, spill and recapture models, and cost and revenue models, and claims full scenario evaluations in under 30 seconds for up to 250,000 markets at once (Lufthansa Systems 2026); these are vendor claims. Such a tool computes something between views 1 and 2 of Definition 12.1, depending on how its revenue allocation, recapture and cost boundaries are configured, and the analyst's job is to know which. The aircraft at the centre of this chapter's case is also changing the set of routes evaluated: Cirium's October 2026 review counts 50 A321XLRs delivered worldwide and describes them as opening long-haul markets too thin for widebodies, at the price of reduced cargo capacity (Cirium 2026b). Every such market is a candidate for exactly the evaluation of Section 12.7.

PITFALL

Judging a route by its load factor. The handbook's case runs at 82.4 percent and is not the best use of the aircraft; carrying 170 passengers at CAD 530 earns less passenger revenue than 150 at CAD 620; and Laurentian's summer flights, at 79.8 percent, cover their costs while losing CAD 17,828 per round trip against the aircraft's alternative. Load factor is a symptom of demand, price and gauge, not a measure of value (Network Planning Handbook 2026).

12.11 Summary

- Route economics has three views: flight contribution under a stated convention, incremental network contribution against a baseline, and the capital and strategic case. A route can pass one and fail another, and the decision horizon determines which costs and which view apply.
- Prorates split connecting fares by distance or local fares so that allocations add up; marginal leg values do not add up, because feeders and long-haul legs are complements. Allocated leg revenue is a reporting convention, not a marginal value.

- The value of a leg to the network is the change in the network's optimal contribution when the leg is added, net of the passengers it displaces; in Laurentian's bank, Seville is worth CAD 120,450 of fares per bank after CAD 1,259 of displacement at mean demand.
- Count only costs that change with the decision; an accounting loss is not a reason to cancel and a positive contribution is not a reason to launch.
- Aircraft time has a shadow price set at the margin of the fleet plan. Charge it in a screening model or compare explicit alternatives, never both; for large blocks of hours, re-solve.
- Spill-based loads lower the handbook's Seville flight contribution by CAD 2,885 per departure; seasons matter more, turning the same route from a CAD 3,677 winner at its peak (autumn and spring) into a CAD 26,348 loser in winter against the alternative.
- Laurentian's YUL–SVQ is worth flying in autumn and spring only (CAD 668,103 a year against the alternative), with a westbound penalty of CAD 10,083 per trip; a fare 4.6 percent below plan or demand capture below 91.7 percent reverses the decision.
- Make ramp-up, launch costs and incentives explicit; the first season is a poor guide to the route's long-run value.
- The memo states the recommendation, the evidence, the downside and the thresholds that would make it wrong, with an owner for each open input.

Table 12.8 Key formulas of the chapter.

Quantity	Formula
Mileage prorate	$f_{jl} = f_j\, d_l / \sum_{k \in \ell(j)} d_k$
Straight-rate prorate	$f_{jl} = f_j f_l^{\text{loc}} / \sum_{k \in \ell(j)} f_k^{\text{loc}}$
Network value of a product	$f_j - \sum_{l \in \ell(j)} \pi_l$
Marginal value of a leg	$\Delta_l = \Pi(\mathcal{L}) - \Pi(\mathcal{L} \setminus \{l\})$
Incremental network contribution	$\Delta\text{NC} = \sum [(R+O-\text{AC})_s - (R+O-\text{AC})_b] - F_{\text{launch}}$
Screening with aircraft time	approve if $V_m - \omega h_m \geq 0$, $\omega = V_k / h_k$ at the margin
Expected load	$\mathbb{E}[Y] = \mu - \sigma L((C-\mu)/\sigma)$
Round-trip value	(12.2); fare threshold (12.3)

12.12 Further reading

Belobaba's lecture on route planning and evaluation (Belobaba 2006i) is eighteen slides long and should be read in full; its sample network and its table of measures by decision are the basis of

Section 12.1. The handbook's sections on route economics, its two exercises and its Montreal–Seville case (Network Planning Handbook 2026) are the source of Sections 12.6 and 12.10, and its warnings about double counting and feed uplifts are worth rereading before every business case. Barnhart, Kniker and Lohatepanont (2002) show why itinerary-level passenger flows and recapture belong inside fleet decisions, and Lohatepanont and Barnhart (2004) extend the idea to choosing flights and fleets together. Talluri and van Ryzin (2004b) give the theory of bid prices and displacement used in Example 12.1, and The Global Airline Industry (Belobaba, Odoni, and Barnhart 2015) covers the planning process and route evaluation at textbook length.

CHAPTER

13

Frequency and Schedule Design

IN THIS CHAPTER

- Why frequency is part of the product: schedule delay, total trip time, the frequency elasticity of demand and the S-curve.
- The central trade-off between frequency and aircraft size, priced with the spill model of Chapter 9.
- How departure times are chosen: time-of-day demand, time zones and the constraints that push flights off the peak.
- Hub banks: connection windows, minimum connecting time, how many connections a bank creates and what its peaks cost.
- Rotation arithmetic in UTC and the fleet a timetable needs, on Laurentian's Montreal–Seville pattern.
- Seasons, SSIM, airport slots under the WASG and airport operating restrictions.
- Schedule design as an optimization problem, and how to decide whether a 20-minute retiming is worth it.

DRAWS ON Belobaba, Airline schedule development and Airline economics review (MIT 16.75J, 2006); Barnhart, The schedule design problem (MIT 1.206J, 2003); IATA, ACI and WWACG, Worldwide Airport Slot Guidelines, edition 4 (2025); handbook; Lohatepanont and Barnhart (2004)

A route is not yet a product. What a customer buys is an itinerary on a date, at particular times, with a particular frequency, aircraft, cabin and connection; the same city pair can perform very differently at 07:00 and at 14:00 because customer schedules, feeder arrivals, competitor departures and aircraft rotations differ (Network Planning Handbook 2026). Much of a network planning analyst's week is spent on these attributes: adding a second daily frequency, swapping aircraft between routes, moving a departure 20 minutes to catch a bank, deciding whether a seasonal route runs three or four

times a week. Each change looks small on the departure board and ripples through connections, rotations, crews, slots and the demand of other flights.

This chapter supplies the economics and the mechanics behind those decisions. It starts with the economics of frequency (how waiting for a departure enters trip time and therefore demand, Chapter 4) and prices the frequency versus gauge trade-off with the spill arithmetic of Chapter 9. It then turns to departure times, connecting banks, aircraft rotations computed with clocks that change on different dates on each side of the Atlantic, and the seasonal slot system, and closes with the optimization view of schedule design and a worked retiming decision. Route profitability is the subject of Chapter 12, fleet assignment of Chapter 14, aircraft and crew routing of Chapter 15 and schedule slack of Chapter 16.

13.1 From a route to a timetable

Given the routes an airline has decided to operate and the fleet it owns, Belobaba splits **schedule development** into four linked decisions: **frequency planning** (how often), **timetable development** (at what times), **fleet assignment** (with which aircraft type) and **aircraft rotation planning** (which sequence of flights each aircraft flies, so that the network balances) (Belobaba 2006f). The process starts a year or more before departure. Frequency plans come first; timetables and rotations are set two to six months ahead; revisions continue until the flight operates.

Four supply terms recur. A **flight leg** is a nonstop operation of an aircraft from one airport to another with a departure and an arrival time. A **flight** is one or more legs flown consecutively, usually by one aircraft under one flight number. A **route** is the sequence of links served under one flight number. A passenger's **itinerary** (or path) is the combination of legs used to complete a journey in an origin-destination market (Belobaba 2006f). Chapter 2 explains why legs and journeys must never be mixed in counting.

The decomposition is a convenience, not a property of the problem. A departure time fixes which connections exist, which aircraft and crews can fly the flight and which slot it needs, so the subproblems must be iterated across their boundaries (Barnhart and Cohn 2004). The size of the problem also makes most timetable changes **incremental**: one change of departure time can move downline times, connections and rotations, and even the number of aircraft the schedule needs (Belobaba 2006f). The analyst's daily work is therefore what-if evaluation against a baseline. The handbook's discipline for

keeping it auditable is a **schedule card** for every proposed service: both directions, operating dates and days, local and UTC times and dates, block times, minimum and planned turns, aircraft and seats by cabin, feasible connections at both ends, the implied rotation, and the owner of every unverified assumption (Network Planning Handbook 2026). Table 13.2 builds one for Laurentian's Seville service.

13.2 Frequency, schedule delay and total trip time

A traveller has a time at which she would like to leave. Unless a flight departs then, she must leave earlier or later, and the gap is a cost even though it is spent on the ground. Frequency shrinks it.

> **DEFINITION 13.1 Schedule delay**
>
> Let t^{*} be a traveller's preferred departure time and d the departure she takes. Her **schedule delay** (or schedule displacement) is $|t^{*} - d|$. The **expected schedule delay** $\mathrm{SD}(f)$ of a timetable with f daily departures averages it over preferred times, each traveller taking the departure closest to her preferred time. If that flight is full and she must move further, the extra displacement is **stochastic delay** (Douglas and Miller 1974).

To see how SD depends on f, take an operating day of length H hours (for example 06:00 to 22:00, so $H = 16$) and suppose preferred times are spread uniformly over it.

RESULT 13.1 Expected schedule delay with f departures

If preferred departure times are uniform on $[0, H]$ and travellers take the nearest departure, then for any f departure times

$$\mathrm{SD}(f) \;\geq\; \frac{H}{4f}, \tag{13.1}$$

with equality exactly when the departures are equally spaced at $d_j = (2j-1)H/(2f)$, $j = 1, \dots, f$. If travellers can only take a departure at or after their preferred time, the minimum expected wait is $H/(2f)$, reached at $d_j = jH/f$, the end of each sub-period. This needs the last departure at H, since otherwise travellers who want to leave after it have no flight that day, or a schedule that wraps around the day (Chapter 4).

Proof. The nearest-departure rule splits $[0, H]$ into catchment intervals $I_1, \dots, I_f$ with lengths ℓ_j summing to H, departure d_j lying in I_j. The integral of $|t - d|$ over an interval of length ℓ is smallest when d is at the midpoint, where it equals $\ell^2/4$. Hence

$$H\,\mathrm{SD}(f) = \sum_{j=1}^{f} \int_{I_j} \left|t - d_j\right| \mathrm{d}t \;\geq\; \sum_{j=1}^{f} \frac{\ell_j^2}{4}$$

$$\geq\; \frac{1}{4f}\Big(\sum_{j=1}^{f} \ell_j\Big)^2 = \frac{H^2}{4f}$$

by the Cauchy-Schwarz inequality, with equality exactly when every $\ell_j = H/f$ and every d_j is a midpoint. In the one-sided case with $d_f = H$, each catchment is $(d_{j-1}, d_j]$ with $d_0 = 0$ (with wrap-around, $d_0 = d_f - H$), the integral of $d_j - t$ over it is $\ell_j^2/2$, and the same argument gives $H/(2f)$, with equality when every $\ell_j = H/f$. ∎

Belobaba writes total trip time as fixed time (access, egress and airport processing) plus flight time plus a schedule displacement of K/f hours (Belobaba 2006d); Result 13.1 identifies $K = H/4$ for travellers who can move either way:

$$T(f) = t_0 + t_b + \frac{K}{f}, \qquad K = \frac{H}{4}, \tag{13.2}$$

with t_0 the fixed time and t_b the block time. Going from f to $f+1$ departures saves $K/(f(f+1))$ hours: with $H = 16$, the second daily frequency saves two hours of expected displacement, the third 40 minutes, the fourth 20 and the eighth 4.3.

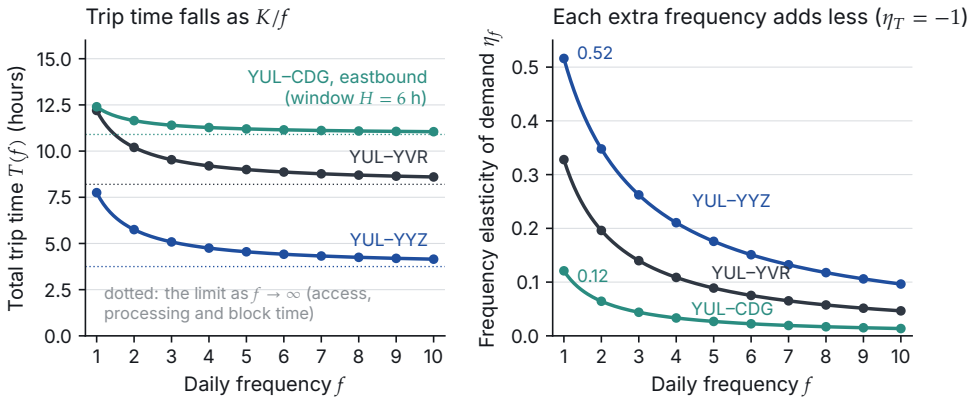


Figure 13.1 Left: total trip time as daily frequency grows for three Laurentian routes, with the schedule delay of Result 13.1; the dotted lines are the limits as frequency grows without bound. Right: the frequency elasticity of demand (13.3) with a time elasticity of -1. Access, processing and block times are illustrative; the transatlantic route uses a six-hour eastbound departure window instead of a full day.

Trip time matters because it enters demand. With the constant-elasticity market demand of Chapter 4, $D = M P^{\eta_P} T^{\eta_T}$, where P is the average fare, η_P the price elasticity and $\eta_T < 0$ the time elasticity, frequency acts on demand only through T.

RESULT 13.2 The frequency elasticity of demand

Under (13.2) and constant-elasticity demand, the elasticity of market demand with respect to daily frequency is

$$\eta_f = \frac{\partial D}{\partial f}\frac{f}{D} = |\eta_T|\,\frac{K/f}{T(f)}, \tag{13.3}$$

the time elasticity times the share of schedule delay in total trip time. It lies between 0 and $|\eta_T|$ and falls towards zero as f grows. Adding one frequency multiplies demand by $\left(T(f+1)/T(f)\right)^{\eta_T}$.

The proof is the chain rule, with $\partial T/\partial f = -K/f^2$. The result explains three pieces of airline lore. Frequency matters more in short-haul markets, where schedule delay is a large share of T (Belobaba 2006d). Its value declines with each addition, so each market has a **saturation frequency** beyond which departures add almost no demand. And it matters more for business travellers, whose time elasticity is larger in absolute value.

WORKED EXAMPLE 13.1 How much demand does a second frequency create?

Take three illustrative Laurentian routes with $\eta_T = -1$: YUL–YYZ ($t_0 = 2.5$ hours of access and processing, $t_b = 1.25$), YUL–YVR ($t_0 = 3.0$, $t_b = 5.2$) and eastbound YUL–CDG ($t_0 = 4.0$, $t_b = 6.9$). The domestic routes have $H = 16$ and $K = 4$; eastbound transatlantic departures fit an evening window of about six hours (Section 13.4), so YUL–CDG has $H = 6$ and $K = 1.5$.

With one daily flight, total trip times are $T(1) = 2.5 + 1.25 + 4 = 7.75$ hours to Toronto, 12.2 hours to Vancouver and $4.0 + 6.9 + 1.5 = 12.4$ hours to Paris. The frequency elasticities at $f = 1$ are $4/7.75 = 0.52$, $4/12.2 = 0.33$ and $1.5/12.4 = 0.12$. A second daily flight cuts trip time to 5.75, 10.2 and 11.65 hours, which raises market demand by $7.75/5.75 - 1 = 34.8$ percent, 19.6 percent and 6.4 percent. A third adds 13.1, 7.0 and 2.2 percent. Figure 13.1 shows the full curves.

On short haul the first few frequencies are a demand decision, not only a capacity decision: they create travel. On long haul, where the flight dominates trip time and time zones squeeze the window, a second daily frequency must be justified by capacity, connections or competition rather than by stimulated demand.

These numbers describe the whole market. When several airlines serve it, frequency also decides how it is shared, and more than proportionally: the airline with more frequency alone serves the travellers who want to leave when only it flies, and shares the times when both fly (Belobaba 2006d). Belobaba's **S-curve** writes airline i's share as

$$\mathrm{MS}_i = \frac{f_i^{\alpha}}{\sum_j f_j^{\alpha}}, \qquad \alpha > 1, \tag{13.4}$$

with historical estimates of α between 1.3 and 1.7 (Belobaba 2006d); Chapter 4 derives and estimates it. Against a competitor with three daily nonstops and $\alpha = 1.5$, Laurentian's share is 16.1 percent with one daily flight, 35.2 with two, 50.0 with three and 60.6 with four: the second frequency adds 19.1 points against 15.0 under proportional shares. For two carriers $\partial \mathrm{MS}_1 / \partial f_1 = \alpha \, \mathrm{MS}_1 (1 - \mathrm{MS}_1) / f_1$, so at parity the S-curve multiplies the proportional effect by α.

That amplification drives frequency competition. If two airlines split a market worth R a day by (13.4) and pay c per daily frequency, the first-order condition $R\alpha\,\mathrm{MS}_1(1-\mathrm{MS}_1)/f_1 = c$ gives a symmetric equilibrium of $f^* = \alpha R/(4c)$ frequencies each (an equilibrium for $0 < \alpha < 2$, as Chapter 4 proves), so the two spend $\alpha R/2$ on frequency: half the market's revenue with proportional shares, three quarters with $\alpha = 1.5$. The model is stylized (fixed fares, symmetric costs, no capacity limits), but its message is robust: steep share effects push competitors towards many frequencies with small aircraft. That is the setting of the gauge decision.

13.3 Frequency versus gauge

Suppose a route needs about 280 seats a day, offered as one large aircraft or two smaller ones. The large aircraft wins on cost per seat and pools demand into one cabin; the smaller ones win on frequency, which creates demand and share. Costs are in dollars per trip and benefits in passengers per day, so neither argument settles the question; the contribution arithmetic of Chapter 9, applied to each option with the demand that option generates, does.

Let a plan fly f daily departures of an aircraft with C seats and trip cost TC, each passenger contributing r. Write μ_1 for mean daily demand with one departure and $u(f)$ for the multiplier frequency applies to it (Result 13.2 where the airline serves the market alone, the S-curve where it competes). If departure j receives a share w_j of demand, with k-factor k, expected daily contribution is

$$\pi(f, C) = r \sum_{j=1}^{f} \Big[\mu_j - \sigma_j L\Big(\frac{C-\mu_j}{\sigma_j}\Big)\Big] - f\,\mathrm{TC}, \qquad \mu_j = w_j\,u(f)\,\mu_1, \qquad \sigma_j = k\,\mu_j, \tag{13.5}$$

where L is the standard normal loss function of Result 9.1.

WORKED EXAMPLE 13.2 Two A220s or one 787-9 to Vancouver?

Return to the YUL–YVR service of Example 9.5, flown by two daily A220-300s with 137 seats each. Demand is $\mathcal{N}(160, 48^2)$ for the evening flight and $\mathcal{N}(100, 30^2)$ for the morning flight. The alternative is one evening 787-9 (298 seats). On the 5.2-hour sector two A220 trips cost CAD 87,200 (CAD 318 per seat) and one 787-9 trip CAD 85,600 (CAD 287 per seat): 24 more seats for CAD 1,600 less.

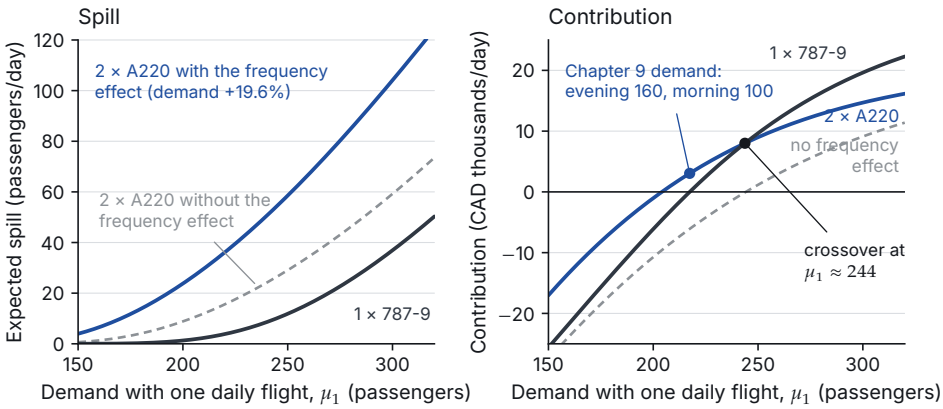


Figure 13.2 Two daily A220-300s against one daily 787-9 on YUL–YVR as Laurentian's one-frequency demand μ_1 grows: expected spill (left) and contribution (right) per day. The blue curve applies the 19.6 percent demand gain of a second frequency (Example 13.1); the dashed grey curve omits it. Laurentian trip costs, $r = 400$, $k = 0.30$.

Demand. If Laurentian is the only nonstop carrier, frequency acts through trip time alone, and Example 13.1 gives $u = 12.2/10.2 = 1.196$: the 260 daily passengers of the two A220s correspond to $\mu_1 = 217.4$ with one daily flight.

Two A220s. The evening flight spills $48\,L((137-160)/48) = 32.8$ passengers a day and the morning flight $30\,L(1.233) = 1.6$, so 225.62 are carried and, at CAD 400 a passenger, contribution is CAD 3,049 a day, or CAD 6,209 with the 7.9 daily recaptures simulated in Example 9.5.

One 787-9. With $\sigma = 65.2$, $z = (298-217.4)/65.2 = 1.24$, spill is 3.4, 214.0 are carried and contribution is $400\times214.0-85{,}600 \approx 0$.

The A220s win by about CAD 3,000 a day, CAD 1.1 million a year, while spilling ten times as many passengers at a higher seat cost. Without the frequency effect ($u = 1$) they would carry 202.8 passengers and lose CAD 6,088 a day: the frequency effect is worth about CAD 9,100 a day. If the two flights' demands fluctuate independently, as Example 9.5 assumes, consolidation lowers the k-factor to $\sqrt{48^2+30^2}/260 = 0.22$ and the 787-9 earns CAD 1,008, still short of the A220s.

Figure 13.2 varies μ_1. The 787-9 overtakes at $\mu_1 \approx 244$, where the A220s face 292 passengers and spill more than 50 a day; by then the better plan is usually to keep both frequencies and upgauge them. In a competitive market the frequency effect is

far larger: against a rival with three daily nonstops and $\alpha = 1.5$, the second frequency multiplies Laurentian's demand by $0.352/0.161 \times 1.022 = 2.23$ (share times market stimulation): the 260 passengers of the two A220s would shrink to about 116 on a single 787-9, which would fly 61 percent empty.

PITFALL

Choosing the aircraft with the lowest cost per seat. Seat cost compares capacity, not plans: it ignores the demand frequency creates, the share it wins and the spill capacity causes. Compare the contributions of complete plans.

A hub adds a further argument for frequency: two daily flights can sit in two banks, each feeding and fed by different flights (Section 13.5). Hence hub carriers prize small aircraft with competitive seat costs on thin spokes, which buy frequency and connectivity without a large trip cost. As a route grows, airlines usually add frequencies until the useful times are filled, then upgauge; at slot-constrained airports (Section 13.7) the extra frequency may not be available at all.

Frequency is a demand decision and gauge is a cost decision. Price both with the same contribution arithmetic, letting demand respond to frequency before spill is computed.

13.4 Choosing departure times

Uniform preferences are a benchmark, not a description. Business demand in North America peaks early morning and late afternoon, around 09:00 and 17:00 by Belobaba's rule of thumb (Belobaba 2006f); leisure demand is flatter. The PODS simulator gives each traveller a decision window, from an earliest acceptable departure to a latest acceptable arrival, drawn from a market time-of-day profile (Chapter 6) (Belobaba 2006g). For a general preferred-time density g, Result 13.1 generalizes with medians in place of midpoints.

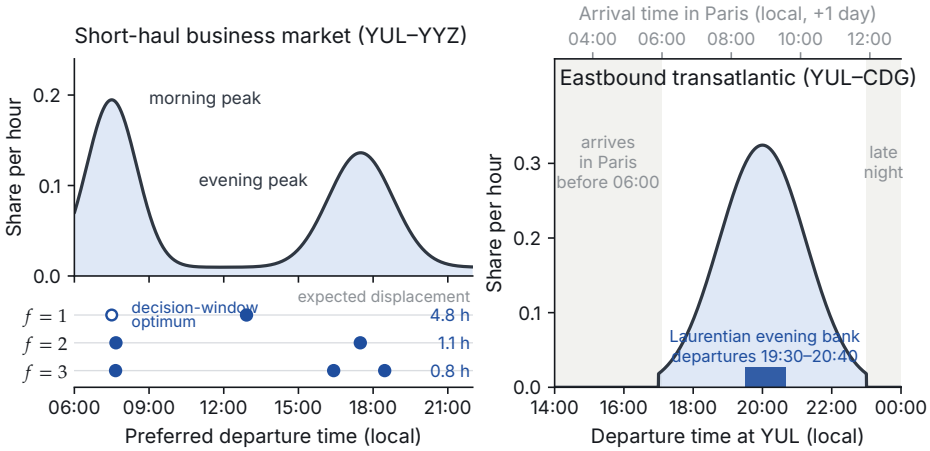


Figure 13.3 Time-of-day demand profiles (illustrative). Left: a business-heavy short-haul market; the rows show the departure times minimizing expected schedule delay for one, two and three daily flights, and the hollow marker the single flight maximizing demand within one hour. Right: the eastbound transatlantic window; departures before about 17:00 reach Paris before 06:00.

RESULT 13.3 Where to place f departures

Let preferred times have density g on $[0, H]$ and let travellers take the nearest departure. A set of departure times $d_1 < \cdots < d_f$ minimizing expected schedule delay satisfies two conditions: catchment boundaries lie at the midpoints $(d_j + d_{j+1})/2$, and each departure d_j is a median of g restricted to its own catchment.

Proof. For fixed departures, assigning each traveller to the nearest departure minimizes her delay, which places the boundaries at midpoints. For a fixed catchment I_j, the derivative of $\int_{I_j} |t - d|\, g(t)\, \mathrm{d}t$ with respect to d is $G_j(d) - (G_j(I_j) - G_j(d))$, where $G_j(d)$ is the demand in I_j below d; it is zero when d splits the catchment's demand in half. ∎

The conditions suggest an algorithm: assign catchments, move each departure to its catchment median, repeat, and restart from several timetables because the objective is not convex. They also expose a modelling choice. Linear displacement cost makes leaving five hours early five times as bad as leaving one hour early, whereas real travellers outside their decision window choose another airline, another day or no trip; with displacement cost capped at a tolerance τ, the best departure maximizes the demand within τ of it.

Figure 13.3 applies the result to an illustrative YUL–YYZ profile with 45 percent of travellers in a morning peak, 40 percent in a broader evening peak and the rest spread out. A single flight minimizing average displacement lands at 12:55, the median: expected delay is 4.8 hours (worse than a uniform market's 4.0) and only 2.1 percent of travellers are within an hour of it; the decision-window criterion moves it to 07:30, within an hour of 18.1 percent. Two flights go to the peaks, 07:40 and 17:29, cutting expected delay to 1.1 hours (2.0 in a uniform market) with 58 percent of travellers within an hour. A third splits the evening peak (07:39, 16:25, 18:28) and lowers delay only to 0.84 hours: in a peaked market the first two frequencies are worth more, and the third much less, than the uniform model suggests.

Time zones compress the window. Flights from the United States to Europe do not leave before about 17:00, because travellers do not want to arrive before 06:00 (Belobaba 2006f), and the same logic holds at YUL: with 6.9 block hours and six hours of time difference, a 17:06 departure from YUL reaches Paris at 06:00, and late departures meet YUL's night restrictions (Section 13.7). The usable window is about six hours, hence $H = 6$ in Example 13.1, and transatlantic departure times are set less by local preferences than by the bank they connect with.

Belobaba lists the constraints that push departures off the peaks (Belobaba 2006f). An aircraft leaving A at 09:00 and arriving at B at 11:00 can leave B at 12:00, off-peak; holding it for the 17:00 peak costs utilization. Most airlines therefore keep turns short and fly some off-peak departures, to protect frequency share and position aircraft for peaks elsewhere, leaving little buffer for maintenance and weather. Banks require spoke arrivals inside a window; time zones, slots and noise curfews restrict feasible times; crews and maintenance add requirements. Competitors matter too: departing ten minutes before a rival captures the travellers who prefer earlier times, a Hotelling-style logic.

13.5 Hub banks and connections

A hub multiplies markets by timing: if flights from many cities arrive together and flights to many cities leave together shortly afterwards, every arriving passenger can reach every departing city with one stop.

DEFINITION 13.2 Bank and connection window

A **bank** (or wave) is a group of arrivals at a hub followed by a group of departures, timed so that passengers can connect (Network Planning Handbook 2026). Its **connection window** $[\mathrm{MCT}, \bar{\tau}]$ runs from the **minimum connecting time** (MCT), the shortest interval the applicable airline, airport, terminal and connection-type rules allow for transferring a passenger and baggage, to $\bar{\tau}$, the longest connection the analyst treats as sellable.

DEFINITION 13.3 Feasible connection

An arrival i at time a_i and a departure j at time d_j form a **feasible connection** when

$$\mathrm{viable}(i,j) = 1 \iff \mathrm{MCT}(i,j) \le d_j - a_i \le \bar{\tau}, \tag{13.6}$$

with both times as full timestamps (date, time and zone) rather than clock faces (Network Planning Handbook 2026).

WORKED EXAMPLE 13.3 The handbook's connection screen

Three inbound flights arrive at 15:10, 15:35 and 16:05; an outbound flight departs at 17:00 (Network Planning Handbook 2026). The connecting times are 110, 85 and 55 minutes.

With an MCT of 60 minutes the first two connect and the third does not. Retiming the outbound to 17:15 gives 125, 100 and 70 minutes, so all three connect if $\bar{\tau} \ge 125$. With an MCT of 90 minutes only the first connects at 17:00, and the first two at 17:15: the retiming again gains one connection, but a different one. With a commercial limit of $\bar{\tau} = 120$ minutes and an MCT of 60, the 17:15 departure gains the 16:05 arrival but loses the 15:10 arrival, and the count does not change.

The retiming is worth making only if the itineraries it adds are worth more than the demand lost elsewhere, the extra cost and the effects on the return rotation, and that value changes with

the MCT. None of these numbers is an actual YUL rule (Network Planning Handbook 2026).

Laurentian's evening transatlantic bank shows the screen at hub scale (Table 13.1 and Figure 13.4; times illustrative). Of the 56 pairs formed by eight North American arrivals between 17:55 and 18:45 and seven European departures between 19:30 and 20:40, 53 are feasible with a 60-minute MCT and 40 with a 90-minute MCT; the 53 average 108 minutes and the longest, Vancouver to Rome, takes 165. Connections are lost first in the corner where the last arrivals meet the first departures.

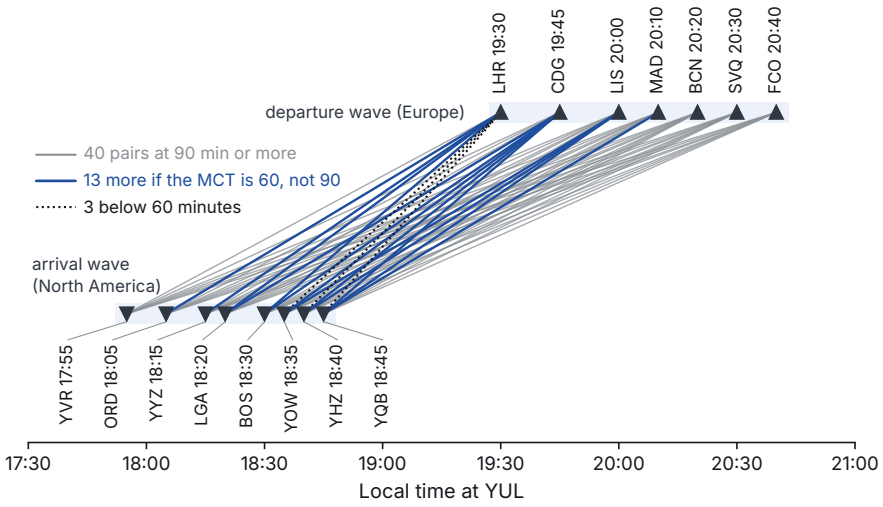


Figure 13.4 Laurentian's illustrative evening bank at YUL. Each line joins an arrival to a departure; grey lines are connections that survive a 90-minute MCT, blue lines are the 13 additional connections available with a 60-minute MCT, and dotted lines the three pairs below 60 minutes.

Table 13.1 Connecting times in minutes in Laurentian's illustrative evening bank at YUL. Values in parentheses fail a 60-minute MCT; values marked † pass a 60-minute but fail a 90-minute MCT.

Arrival	LHR 19:30	CDG 19:45	LIS 20:00	MAD 20:10	BCN 20:20	SVQ 20:30	FCO 20:40
YVR 17:55	95	110	125	135	145	155	165
ORD 18:05	85†	100	115	125	135	145	155
YYZ 18:15	75†	90	105	115	125	135	145
LGA 18:20	70†	85†	100	110	120	130	140
BOS 18:30	60†	75†	90	100	110	120	130
YOW 18:35	(55)	70†	85†	95	105	115	125
YHZ 18:40	(50)	65†	80†	90	100	110	120
YQB 18:45	(45)	60†	75†	85†	95	105	115

How many connections a hub creates follows from counting.

RESULT 13.4 Connections in banked and rolling hubs

A hub has n arrivals and n departures a day, each serving a different spoke, over an operating period of length H, with minimum and maximum connecting times MCT and $\bar{\tau}$ in the same units.

1. If a bank's last arrival precedes its first departure by at least MCT and its first arrival precedes its last departure by at most $\bar{\tau}$, a bank with n_b flights each way connects every arrival to every departure, n_b^2 pairs, of which $n_b(n_b - 1)$ are useful when each spoke is served in both directions (the rest return the passenger to her origin).
2. A day of B such banks of n/B flights each way, more than $\bar{\tau}$ apart, offers $n^2/B - n$ connections.
3. If arrivals and departures are instead spread evenly, n/H per unit of time (a **rolling** hub), each arrival finds about $(n/H)(\bar{\tau} - \text{MCT})$ departures in its window, for about $n^2(\bar{\tau} - \text{MCT})/H$ connections in all.

Banking therefore multiplies connectivity by about $H/(B(\bar{\tau} - \text{MCT}))$.

The proof is direct counting: in (1) every pair satisfies (13.6); (2) multiplies $n_b(n_b - 1)$ by B with $n_b = n/B$; (3) multiplies the arrivals by the expected number of departures in a window of length $\bar{\tau} - \text{MCT}$. Fewer, larger banks create more connections; a rolling hub still connects, but less often and with longer waits.

Figure 13.5 prices the trade-off in an illustrative hub with four banks, each of twelve arrivals within 40 minutes followed an hour later by twelve departures within 40 minutes. With an MCT of 60 minutes and $\bar{\tau} = 180$, the banked day offers $48^2/4 - 48 = 528$ connections, with a median wait of 100 minutes. Spreading the arrivals evenly from 07:00 to 20:15, each aircraft turning in 75 minutes, leaves 315 with a median of 126, a ratio of 1.68 against the $795/(4 \times 120) = 1.66$ of Result 13.4. The bank's price is its peak: five movements per 15 minutes against two, twelve aircraft on the ground at once against five, and 100 minutes of ground time per aircraft against 75, utilization spent waiting for the bank.

Those peaks load everything a hub shares: gates, ramp crews, security, customs, baggage and the runway. A historical profile of a U.S. banked hub in the MIT course shows operations per 15 minutes exceeding instrument-weather capacity at every bank (Barn-

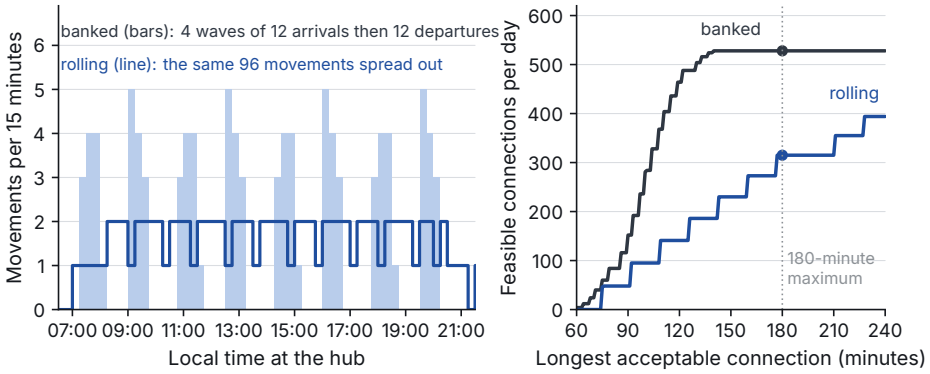


Figure 13.5 The price of connectivity in an illustrative hub with 48 arrivals and 48 departures a day. Left: runway movements per 15 minutes in a day of four banks (bars) and in a rolling version of the same flights (line). Right: feasible connections as the longest acceptable connection grows, with a 60-minute MCT; the dots mark a 180-minute limit.

hart 2003), and banks make delays contagious, since a late inbound flight either holds its connecting departures or strands its passengers (Chapter 16). A rolling (de-peaked) hub spreads flights to flatten the peaks, at the cost of fewer and longer connections; Jiang (2006) develops models for de-peaking a banked schedule that keep as many weighted connecting itineraries as possible without losing profitability, tested on data from a major U.S. airline. In practice hubs bank the flows that matter most (for YUL, North America to Europe in the evening and Europe to North America in the afternoon) and spread the rest.

IN PRACTICE **Banks at Air Canada's hubs in 2026**

Air Canada develops Toronto Pearson, Montreal-Trudeau and Vancouver as its international hubs (Air Canada 2025). At an investor conference in September 2026, its chief financial officer described the banks as synchronized so that U.S. inbound traffic meets international departures "within a couple of hours", called the A220 (about 130 seats at the seat cost of a Boeing 737 or Airbus A320) the key enabler for serving smaller U.S. cities, and noted the appeal of arriving in Europe around 07:00 to 08:00 (Investing.com 2026). These are this chapter's levers: bank timing, small-gauge feeders and time-zone windows. The source is a published conference transcript; treat its figures as management statements.

PITFALL

Inflating connectivity. A feasible pair is not a market: each candidate must still be scored by demand, share, cabin availability, elapsed time and contribution, without counting mutually exclusive itineraries as separate passengers. And treating the MCT as a target produces connections that are valid but misconnect whenever the inbound is late: MCT compliance makes a connection schedule-valid, not reliable (Network Planning Handbook 2026).

13.6 Rotations, clocks and the fleet a timetable needs

A timetable is operable only when every departure has an aircraft. The unit is the **rotation**, the sequence of flights an aircraft flies before returning to its start. For an out-and-back route the handbook's cycle is outbound block plus away turn plus return block plus base turn, $7.0 + 1.5 + 8.0 + 2.0 = 18.5$ hours in its example, which by itself does not show that one aircraft can fly the route every day (Network Planning Handbook 2026). Aircraft **utilization**, block hours per aircraft per day, falls whenever aircraft wait for peaks or banks or routes run less than daily.

RESULT 13.5 How many aircraft a timetable needs

A timetable repeats with period Π (a day or a week) and is flown by one fleet. A flight occupies an aircraft from departure until it is ready again (arrival plus minimum turn). Let $A(t)$ be the number of flights occupying aircraft at time t, BH the block hours per period, and h the block hours one aircraft can sustain per period after maintenance and other unavailability.

1. Any operation needs at least BH/h aircraft (pooling bound) and at least $\max_t A(t)$ aircraft (simultaneity bound).
2. Ignoring maintenance, and with as many ready events as departures at every station over a period, the minimum number of aircraft is

$$N^* = A(t_0) + \sum_s \max_{t_0 \le t < t_0 + \Pi} [-N_s(t)], \tag{13.7}$$

where $N_s(t)$ counts ready events minus departures at station s between a reference time t_0 and t.

Proof. For the bounds: each aircraft supplies at most h block hours per period, and two flights occupying aircraft at the same instant need two aircraft. For the count: at t_0 every aircraft is either occupying a flight or on the ground at some station, so the fleet is $A(t_0) + \sum_s g_s(t_0)$, where g_s is the ground inventory. Inventory evolves as $g_s(t) = g_s(t_0) + N_s(t)$ and can never be negative, so $g_s(t_0) \ge \max_t [-N_s(t)]$; this is nonnegative because $N_s(t_0) = 0$. Setting $g_s(t_0)$ equal to the bound is feasible (assign each departure any aircraft on the ground) and periodic, because balance makes $N_s(t_0 + \Pi) = 0$. Finally, the fleet at any time t is $A(t) + \sum_s g_s(t) \ge A(t)$, so $N^* \ge \max_t A(t)$. ∎

Equation (13.7) is the count-line logic of the aircraft-count constraint in fleet assignment (Chapter 14) (Hane et al. 1995). The pooling bound is the handbook's fleet lower bound, whose denominator must reflect realistic availability rather than 24 hours a day (Network Planning Handbook 2026). Flown daily, the handbook's cycle needs one aircraft by the count; if an aircraft can sustain 14 block hours a day once maintenance is allowed for (an illustrative figure), its 15 daily block hours need $15/14 = 1.07$ aircraft, so such a route is flown from a pool rather than by one tail.

Rotations across time zones add clocks. Schedules are sold in local time but computed in UTC, and offsets change on different dates: Quebec observes daylight saving time from the second Sunday of March to the first Sunday of November (Québec 2007), the European Union from the last Sunday of March to the last Sunday of October (European Parliament and Council of the European Union 2001), and IATA seasons start on the European dates (IATA, ACI, and WWACG 2025). These rules can change: Quebec consulted on ending clock changes in 2024 (Ministère de la Justice du Québec 2024) and British Columbia stopped them in 2026 (Government of British Columbia 2026), so the dates below assume the law in force in October 2026. Around both season boundaries Montreal and Seville are five hours apart instead of six: from 31 October to 6 November 2027 and from 12 to 25 March 2028.

Laurentian's Seville service mirrors a publicly announced pattern, subject to change: YUL 20:30, Seville 09:30 the next day; Seville 12:45, YUL 14:40; A321XLR; three weekly from 8 September 2027 and four weekly from 1 October, with no departures from 9 January to 17 February 2028 and a last departure on 3 June 2028 (Network Planning Handbook 2026; Air Canada 2026g). Its days, and its operation through the whole winter, are illustrative: Wednesday, Friday and Sunday, adding Monday from October.

WORKED EXAMPLE 13.4 The Seville rotation in UTC

Summer time (YUL at UTC−4, Seville at UTC+2). The 20:30 departure is 00:30 UTC the next day and the 09:30 arrival is 07:30 UTC, a 7-hour block. The 12:45 departure is 10:45 UTC, a turn of 3 hours 15 minutes, and the 14:40 arrival is 18:40 UTC, a block of 7 hours 55 minutes. The aircraft is back 18 hours 10 minutes after leaving, 5 hours 50 minutes before another 20:30 departure, as in the handbook (Network Planning Handbook 2026).

Five-hour weeks (UTC−4 and UTC+1). These weeks belong to the winter season. Keeping Montreal times and block times keeps the summer UTC times, an hour earlier than the winter pattern below, and moves Seville's local times an hour earlier: arrival 08:30, departure 11:45. Keeping the published 12:45 departure from Seville instead would stretch the turn to 4 hours 15 minutes and bring the aircraft home at 15:40. At a coordinated airport such one-hour shifts at a season's edge are filed as

daylight-saving adjustments (Section 13.7). In the four-weekly pattern, 12 of the winter season's 84 departures fall in these weeks.

Winter (UTC−5 and UTC+1). The six-hour difference and the local pattern return, but every UTC time is an hour later than in summer: 01:30, 08:30, 11:45 and 19:40.

Dates. A Sunday 20:30 departure is a Monday flight in UTC, so the commercial Wednesday, Friday and Sunday flights are Thursday, Saturday and Monday operations in Seville. Weekly patterns need an explicit local departure-date convention (Network Planning Handbook 2026).

Aircraft time. A round trip flies 14.92 block hours: 44.75 a week at three weekly, with the aircraft away 32 percent of the week, and 59.67 at four weekly, away 43 percent (Figure 13.6). At 14 sustainable block hours a day, Seville uses 0.46 and 0.61 of an aircraft-week; the rest must be sold to other routes for the aircraft to earn its ownership cost. The October pattern leaves three Montreal windows of 29 hours 50 minutes and one turn of 5 hours 50 minutes, between the Sunday flight's return and the Monday departure, which must exceed the planned minimum turn at YUL.

Table 13.2 Schedule card for Laurentian's YUL–SVQ rotation (A321XLR, 182 seats), holding Montreal times and block times fixed. Summer time: 8 September to 30 October 2027 and from 26 March 2028; five-hour weeks: 31 October to 6 November 2027 and 12 to 25 March 2028; winter: 7 November 2027 to 11 March 2028. Seville times are on the day after the Montreal departure.

Item	Summer time	Five-hour weeks	Winter
Clocks, YUL and SVQ	UTC−4, UTC+2	UTC−4, UTC+1	UTC−5, UTC+1
Depart YUL, local (UTC)	20:30 (00:30)	20:30 (00:30)	20:30 (01:30)
Arrive SVQ, local (UTC)	09:30 (07:30)	08:30 (07:30)	09:30 (08:30)
Depart SVQ, local (UTC)	12:45 (10:45)	11:45 (10:45)	12:45 (11:45)
Arrive YUL, local (UTC)	14:40 (18:40)	14:40 (18:40)	14:40 (19:40)
Block east, west	7:00, 7:55	7:00, 7:55	7:00, 7:55
SVQ turn; time away	3:15; 18:10	3:15; 18:10	3:15; 18:10

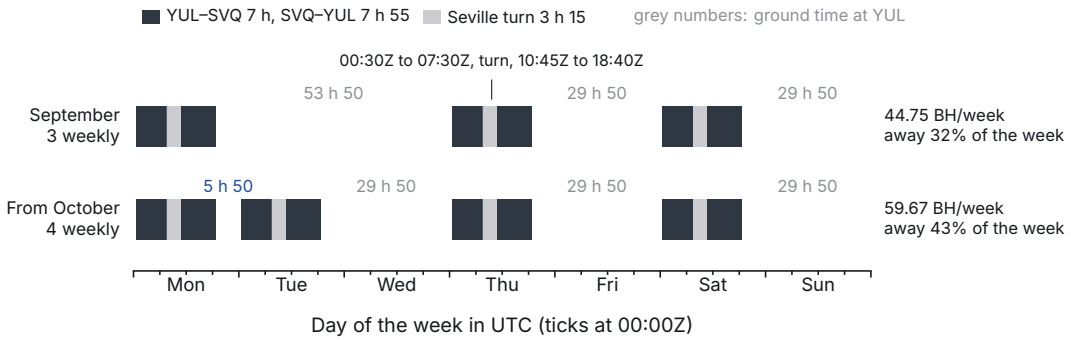


Figure 13.6 One week of Laurentian's YUL–SVQ rotation in UTC (summer-time offsets), with three weekly flights in September and four from October. Numbers above each row are the aircraft's ground time at Montreal between rotations; the blue figure is the 5-hour-50-minute turn between the Sunday and Monday departures.

PITFALL

Clock-face arithmetic. Subtracting local times across time zones, across midnight or across a daylight-saving change gives wrong block times, turns and connection windows. Store a UTC timestamp for computation and the local timestamp for commercial meaning, keep time-zone identifiers and date offsets, and never hard-code a single year-round UTC offset (Network Planning Handbook 2026).

13.7 Seasons, SSIM and slots

Schedules are planned by **IATA scheduling season**: summer begins on the last Sunday of March and winter on the last Sunday of October (IATA, ACI, and WWACG 2025), so S27 runs from 28 March to 30 October 2027, W27 from 31 October 2027 to 25 March 2028 and S28 from 26 March to 28 October 2028. Schedule data move in the formats of IATA's Standard Schedules Information Manual (SSIM), which also covers coordination messages, MCT data and time zones; edition 36 is effective from 1 March 2026 to 28 February 2027 (Network Planning Handbook 2026). Where airport capacity is short, schedules must also fit the Worldwide Airport Slot Guidelines (WASG) of Airports Council International, IATA and the Worldwide Airport Coordinators Group, the industry standard wherever national rules do not override it; slot communications use SSIM formats and UTC (IATA, ACI, and WWACG 2025).

DEFINITION 13.4 Coordination levels, slots and historic precedence

Under the WASG (IATA, ACI, and WWACG 2025):

- a **Level 1** airport has infrastructure generally adequate for demand at all times;
- a **Level 2** airport may be congested at some periods, resolved by schedule adjustments agreed between airlines and a **facilitator**; no slots are allocated;
- at a **Level 3** airport a **coordinator** allocates slots, and an airline must hold one before operating;
- an **airport slot** is a coordinator's permission for a planned operation to use the airport infrastructure needed to arrive or depart at a Level 3 airport on a specific date and time, based on planned on-block and off-block times;
- a **series of slots** is at least five slots at the same or about the same time on the same day of the week, distributed regularly in a season; from Summer 2027 a series requires at least seven slots in a Northern Summer season and still five in a Northern Winter season;
- **historic precedence** entitles an airline to keep a series in the next equivalent season if it operated the series at least 80 percent of the time.

Use it or lose it. Precedence requires 80 percent use of the series held on the **Historics Baseline Date**, 23:59 UTC on 31 January for summer and 31 August for winter (IATA, ACI, and WWACG 2025). Later cancellations count as non-use unless justified by events outside the airline's control (airport or airspace closure, severe weather, strikes); earlier ones shorten or split the counted period only if they cover five or more consecutive weeks or exceed 20 percent of it. Each day of the week is a separate series, daylight-saving time changes do not affect eligibility, and where daylight-saving dates differ by less than five weeks at a season boundary, as between Canada and Europe, airline and coordinator agree in advance how to file those weeks, whose time changes are treated as ad hoc; the historic time is the one held for most of the season (IATA, ACI, and WWACG 2025).

Allocation and calendar. Unchanged historic slots are allocated first; the remaining **pool** is split half to new entrants, a **new entrant** being an airline that would hold fewer than seven slots at the airport that day (three daily rotations) (IATA, ACI, and WWACG

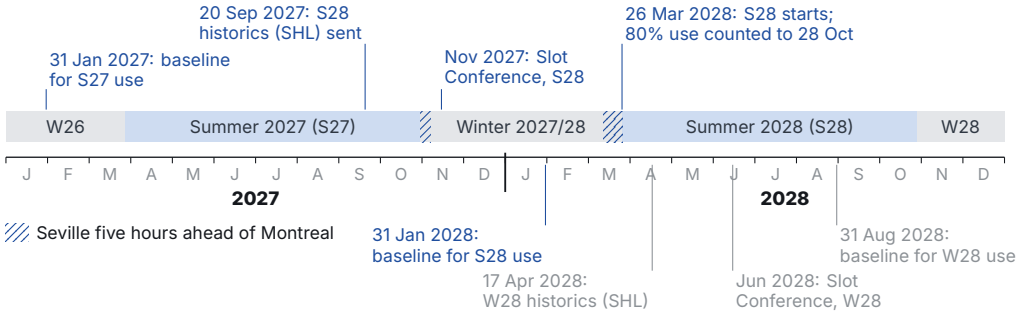


Figure 13.7 The coordination calendar that decides an airline's slots for Summer 2028 (blue) and Winter 2028 (grey), under WASG edition 4. Agreed-historics, submission, allocation and series-return deadlines fall between these dates in IATA's calendar. Hatched weeks: Seville is five hours ahead of Montreal instead of six.

2025). Historic slots cannot be withdrawn to make room for new entrants, year-round extensions have priority, and remaining conflicts are settled by criteria including connectivity, competition and curfews at the other end of the route. Coordinators send historic slots by the third Monday of September (for the next summer) or of April (for the next winter), airlines dispute them up to seven days before the Initial Submission Deadline, and the Slot Conference meets in November for the next summer and in June for the next winter, with discussion of fares or market entry prohibited (Figure 13.7) (IATA, ACI, and WWACG 2025).

WORKED EXAMPLE 13.5 Counting slots

The 80 percent test. S28 has 31 weeks, so a weekly series held all season has 31 slots and must operate at least $\lceil 0.8 \times 31 \rceil = 25$ times, tolerating six unjustified non-uses; a W27 series of 21 slots needs 17.

Seville at a Level 3 airport. Suppose, for illustration, that Seville were coordinated. Its slots fall on the UTC day after the Montreal departure, so the Wednesday flights form a Thursday series: eight Thursdays between the 8 September launch and the end of S27, enough even under the new seven-slot summer minimum. The Monday flights added in October would be Tuesday slots, only four before the season ends: not a series, so no historic precedence. A six-week peak-summer program, a series under the old five-slot rule, no longer qualifies from S27.

New entrant. One rotation a day uses an arrival and a departure slot, so an airline new to the airport remains a new entrant up to three daily rotations.

Three distinctions keep slot analysis honest. A slot is not a traffic right: allocation is independent of the rights granted under bilateral air service agreements, and slots may be requested provisionally before rights are obtained (IATA, ACI, and WWACG 2025); Chapter 10 covers the freedoms of the air. A slot is not a parking stand or an air traffic flow restriction either (Network Planning Handbook 2026). And coordination is the exception: IATA counted 395 airports under coordination in 2026, 216 of them fully slot-coordinated, and estimated that about 43 percent of the world's passengers depart from slot-coordinated airports (IATA 2026a). In 2024 the Canadian Airports Council told a House of Commons committee that only Toronto Pearson, Vancouver and Billy Bishop Toronto City were Level 3 in Canada, citing Montreal-Trudeau, Calgary and Quebec City as examples of Level 2 airports (Canadian Airports Council 2024); designations change, so check the current list.

Slots are one line of the constraint sheet the handbook recommends for every airport, alongside runway and aircraft performance, gates, customs, baggage throughput, handling hours, deicing, noise, construction and alternates, each validated against current approved information (Network Planning Handbook 2026). At home, it warns against saying that YUL closes at night: Aéroports de Montréal describes 24-hour operation with restrictions and exceptions by aircraft category, so the planner needs the exact category and permission (Network Planning Handbook 2026). A summary last updated by the airport in March 2011 bars Chapter 3 jets above 45,000 kg maximum take-off weight from departing between 00:00 and 07:00 or arriving between 01:00 and 07:00 without an exemption (Boeing 2011), and in December 2025 Aéroports de Montréal proposed barring take-offs and landings from 01:00 to 06:00 (CBC News 2025). The sources differ in date and status: obtain the rule in force for the aircraft and season planned.

13.8 Schedule design as an optimization problem

Given origin-destination demands and fares, the fleet, and each type's operating characteristics and costs, **schedule design** chooses the flights and aircraft types that maximize contribution (Barnhart 2003). Belobaba's 2006 view was that no model could find an op-

timal timetable among so many combinations of times, demand estimates and constraints (Belobaba 2006f); the research of the early 2000s took a middle route and optimized changes to an existing schedule. **Fleet assignment with time windows** keeps the flights but lets each departure move by a few minutes (plus or minus 10 in the lecture) when that improves utilization or the match of capacity to demand (Rexing et al. 2000; Barnhart 2003). **Incremental schedule design** goes further: the base schedule is split into mandatory flights and deletion candidates, planners add candidate new flights and alternative times, and the model selects legs from this **master flight list** while assigning fleet types to them (Barnhart 2003; Lohatepanont and Barnhart 2004).

The difficulty is demand. Removing a flight does not simply delete its passengers: some move to the airline's other flights, some to competitors, in amounts that depend on what else is removed. In Barnhart's example (Table 13.3), a market served by four flights carries 450 passengers; removing one optional flight leaves 410, removing both leaves 300, while adding up the effects of each removal predicts 330. Demand responds nonlinearly to the set of flights offered, as the S-curve predicts when frequency share falls.

Table 13.3 Demand and supply interaction in one market (adapted from Barnhart, 2003). The first-order estimate adds the effect of each removal on its own: 40 of the removed flight's passengers move to the second flight each time. The true outcome needs a second-order correction of -30.

Flights offered	Loads by flight	Total	Change
Base: two mandatory, two optional	100, 150, 100, 100	450	
One optional flight removed	100, 190, 120	410	-40
Both optional flights removed	100, 200	300	-150
Both removed, first-order estimate	100, 230	330	-120

Lohatepanont and Barnhart (2004) build two models on itinerary-based fleet assignment (Barnhart, Kniker, and Lohatepanont 2002), the subject of Chapter 14. Let $\mathcal{L} = \mathcal{L}^M \cup \mathcal{L}^O$ be mandatory and optional legs; $\mathcal{K}$ the fleet types, with N_k aircraft of C_k seats; $x_{ik} = 1$ if leg i is flown by type k at cost c_{ik}; and $\mathcal{P}$ the itineraries, with unconstrained demand D_p, fare fare_p and $\delta_i^p = 1$ if itinerary p uses leg i, so that leg i faces unconstrained demand $Q_i = \sum_p \delta_i^p D_p$. Of the t_p^r passengers redirected from itinerary p to r, a fraction b_p^r (the recapture rate) accept; r may be a null itinerary for passengers lost. The **constant market share** model (ISD-FAM) is

$$
\begin{aligned}
\min\ & \textstyle\sum_i \sum_k c_{ik}\, x_{ik} + \sum_p \sum_r (\text{fare}_p - b_p^r\, \text{fare}_r)\, t_p^r \\
\text{s.t.}\ & \textstyle\sum_k x_{ik} = 1 \ \ (i \in \mathcal{L}^M), \qquad \sum_k x_{ik} \le 1 \ \ (i \in \mathcal{L}^O), \\
& y_{k,v^-} + \textstyle\sum_{i \in \text{In}(k,v)} x_{ik} = y_{k,v^+} + \sum_{i \in \text{Out}(k,v)} x_{ik} \ \ \text{for all } k, v, \\
& \textstyle\sum_v y_{k,v}^{\text{cl}} + \sum_{i \in \text{CL}(k)} x_{ik} \le N_k \ \ \text{for all } k, \\
& \textstyle\sum_k C_k x_{ik} + \sum_p \sum_r \delta_i^p\, t_p^r - \sum_p \sum_r \delta_i^p\, b_r^p\, t_r^p \ \ge \ Q_i \ \ (i \in \mathcal{L}), \\
& \textstyle\sum_r t_p^r \le D_p \ \ (p \in \mathcal{P}), \qquad x_{ik} \in \{0,1\},\ t_p^r \ge 0,\ y \ge 0.
\end{aligned}
\tag{13.8}
$$

The first constraint row is **flight selection**. The next two are fleet assignment: flow balance at every node v of each fleet's time-space network, with ground arcs y_{k,v^-} entering and y_{k,v^+} leaving the node and legs $\text{In}(k,v)$ arriving at and $\text{Out}(k,v)$ departing from it, and an aircraft count across the count line, adding the ground arcs $y_{k,v}^{\text{cl}}$ and legs $\text{CL}(k)$ that cross it, which is (13.7) as a constraint. The last rows are the passenger mix model: seats plus passengers redirected away from a leg, minus passengers recaptured onto it, must cover its unconstrained demand. Recaptured passengers fly on the receiving itinerary, so the recapture term uses δ_i^p for passengers moving from r to p. A deleted leg has no seats, so its itineraries are redirected in full and deletion works through recapture while total demand stays fixed. Since total unconstrained revenue is constant, the objective maximizes contribution.

The **variable market share** model (ESD-FAM) lets demand respond. For each optional itinerary $q \in \mathcal{P}^O$, a binary z_q equals 1 exactly when all n_q of its legs are flown:

$$
z_q \le \textstyle\sum_k x_{ik} \ \ (i \in \mathcal{L}(q)), \qquad z_q \ge \sum_{i \in \mathcal{L}(q)} \sum_k x_{ik} - (n_q - 1). \tag{13.9}
$$

The demand corrections ΔD_q^p record the change in the demand of itinerary p when q disappears, with $\Delta D_q^q = -D_q$. The model replaces each D_p, in the demand rows and in Q_i, by

$$
\tilde{D}_p = D_p + \sum_{q \in \mathcal{P}^O} \Delta D_q^p\, (1 - z_q), \tag{13.10}
$$

and adds $-\sum_q \sum_p \text{fare}_p\, \Delta D_q^p\, (1 - z_q)$ to the objective, the revenue lost when q goes net of the revenue that moves to other itineraries. The corrections are data, so the model stays linear, but they are first-order: as Table 13.3 shows, removing two flights is not the sum of

removing each. These equations adapt the lecture's, with the recapture index and the sign convention of the corrections made consistent (Barnhart 2003).

Because neither demand model is exact, the method iterates: solve, recompute demand for the chosen schedule with a richer model, evaluate revenue with the passenger mix model, adjust recapture rates or corrections where the two disagree, and solve again (Barnhart 2003). The reported results are historical: optimized schedules flew fewer flights with fewer aircraft, for claimed gains of roughly USD 100 to 350 million a year against planners' schedules, and the variable-share model took 12 hours on an 800-leg schedule where the constant-share model took 40 minutes (Barnhart 2003). The lecture's caveats are a checklist for any optimized schedule: plans are disrupted, competitors respond, and the models assume deterministic demand, optimal revenue management control, accurate forecasts and known recapture rates.

IN PRACTICE **Schedule tools in 2026**

Commercial planning systems evaluate schedules by simulating origin-destination demand. Lufthansa Systems describes NetLine/Plan as chaining demand and market-share, spill and recapture, and cost and revenue models into flight and network profitability, with scenario evaluation in under 30 seconds (a vendor claim) and modules for scheduling, hub design, fleet assignment and slots (Lufthansa Systems 2026); its confirmation that Air Canada uses NetLine/Ops ++, an operations product, says nothing about planning tools (Lufthansa Systems 2025a; Network Planning Handbook 2026). The 2026 AGIFORS Schedule and Strategic Planning program includes cabin-level origin-destination fleet assignment, integrated planning with bank structures, robust scheduling and, from Air Canada, aircraft routing under uncertainty (AGIFORS 2026). The models of this section are the logic inside such tools; the planner's job is to know which demand and recapture assumptions a tool uses and to test its answers against this chapter's arithmetic.

13.9 When is a schedule change worth it?

A schedule change is worth making when it raises the network's expected contribution against the best feasible alternative, after the

cost of the disruption it invites. The handbook's first-pass framework is

$$\mathbb{E}[\text{net contribution}] = \text{planned contribution} - \mathbb{E}[\text{incremental disruption cost}], \tag{13.11}$$

where planned contribution is the incremental network contribution of Chapter 12 across all affected flights and journeys, and disruption cost covers delays, missed connections, cancellations, reaccommodation and rotation recovery not already in the cost forecast (Network Planning Handbook 2026). For a retiming the handbook's checklist is connections gained and lost, demand shifts, departure competitiveness, arrival restrictions, aircraft return timing and crew feasibility, under low, base and high delay cases, ending with the binding constraint and the assumption most likely to reverse the decision (Network Planning Handbook 2026). Maintenance windows and slots belong on the list too: a 20-minute change can alter a connection, a duty period and a maintenance window at once.

WORKED EXAMPLE 13.6 18:00 or 18:20 to Vancouver?

Laurentian's evening YUL–YVR flight, a 737-8 with 169 seats after the upgauge of Example 9.4, departs at 18:00; the proposal is 18:20. All inputs are illustrative.

Connections. Feeders arrive from Quebec City at 17:05, Boston at 17:10, Halifax at 17:15 and Ottawa at 17:20, with MCTs of 45 minutes for domestic connections and 60 for the transborder one from Boston. At 18:00 only Quebec City (55 minutes) and Halifax (45, exactly the minimum) connect; at 18:20 all four do (75, 70, 65 and 60 minutes), and none is lost.

Demand. Local demand falls from 140 to 136 a day, because a competitor leaves at 18:30. The existing connections carry 12 a day (7 from Quebec City, 5 from Halifax) and the new ones add 6 from Boston and 2 from Ottawa, so mean demand rises from 152 to 156.

Spill. With $k = 0.30$, expected spill rises from 10.94 to 12.89 and carried passengers from 141.06 to 143.11: four more passengers of demand become only 2.055 more carried, because the flight is often full (differences here are computed before rounding). At CAD 400 each, the gain is CAD 822 a day (400×2.055).

Reliability. If a feeder is late with probability 0.30 and its lateness is exponential with mean 20 minutes, a connection with s minutes of slack above the MCT misconnects with probability $0.30\,e^{-s/20}$. Expected misconnections are $7 \times 0.30\,e^{-10/20} + 5 \times 0.30 = 2.77$ a day at 18:00, mostly on the Halifax connection scheduled at its minimum, and 2.40 at 18:20, because the existing connections gain slack. The retiming avoids 0.378 misconnections a day, worth CAD 95 at CAD 250 per misconnected passenger.

The bank peak. At 18:20 the departure falls inside the arrival wave of the evening bank (17:55 to 18:45, Figure 13.4), competing for ramp and taxiways. Operations estimates 2, 4 or 7 extra minutes of taxi-out in the low, base and high cases, at the 737-8's CAD 150 per block minute.

Aircraft and crew. The aircraft reaches Vancouver at 20:32 instead of 20:12 local time, with summer clocks. British Columbia has kept UTC−7 all year since 2026, so in winter it is two hours behind Montreal, not three, and for the same Montreal times every Vancouver time here is an hour later (Government of British Columbia 2026). Its next departure, a 23:55 overnight flight reaching Montreal at about 07:30 (after YUL's restricted early-morning window), keeps a turn of 3 hours 23 minutes instead of 3 hours 43. The crew's layover shortens by 20 minutes and, by assumption, crew planning confirms it remains legal.

Table 13.4 Evaluating the 18:20 retiming of Example 13.6 against 18:00. The base case is worth about CAD 116,000 a year.

Daily effect (CAD)	Low delay	Base	High delay
Extra passengers carried, at CAD 400	822	822	822
Fewer misconnections, at CAD 250	95	95	95
Taxi-out in the bank peak (2, 4, 7 min at 150)	−300	−600	−1,050
Net change	617	317	−133

Table 13.4 gives the recommendation: retime to 18:20 unless operations expects the bank peak to add more than about 6.1 minutes of taxi-out, the break-even. The binding constraint is not the aircraft or the crew but the evening bank at YUL. The assumption most likely to reverse the decision is the local share lost to the 18:30 competitor: in the base case, losing 5.6 local passengers a day instead of 4 wipes

out the gain. A good memo states both thresholds, so the forecasting and operations teams know which of their numbers the decision rests on. None of the effects is visible in the flight's own statement: the retiming changed demand on the flight, the value of four feeders, the reliability of existing connections and the congestion of a bank shared with other flights.

> **PITFALL**
>
> Evaluating a retiming on its own flight. The flight's revenue may rise while the network's falls: new connections can displace higher-value local passengers on a full flight, a later departure can strand passengers on the day's last connection, and a departure moved into a bank peak can delay flights outside the proposal (Network Planning Handbook 2026).

13.10 Summary

- Schedule development sets frequencies, departure times, aircraft types and rotations, iteratively and mostly incrementally; the product is an itinerary at a time, not a route.
- With f departures at the centres of equal sub-periods of an H-hour day, expected schedule delay is $H/(4f)$; the frequency elasticity of demand is the time elasticity times the share of schedule delay in trip time.
- The S-curve gives the frequency leader a disproportionate share; symmetric competitors each fly $\alpha R/(4c)$ frequencies and spend a fraction $\alpha/2$ of market revenue on them.
- Compare frequency and gauge by contribution, letting demand respond to frequency before computing spill: two small aircraft often beat one large aircraft with a lower seat cost.
- Optimal departure times are catchment medians; peaked demand, decision windows and time zones (a six-hour eastbound transatlantic window) matter more than the uniform model suggests.
- A bank of n_b flights each way creates about n_b^2 connections; banking multiplies connectivity over a rolling hub by about $H/(B(\bar{\tau} - \mathrm{MCT}))$, at the price of peaks and of aircraft waiting for the bank.
- A timetable needs at least BH/h and $\max_t A(t)$ aircraft; the exact minimum counts aircraft in the air and on the ground at one instant. Compute in UTC: Montreal and Seville are five hours apart for a week in autumn and two in spring.

- WASG slots are held in UTC by series, kept with 80 percent use and separate from traffic rights; from Summer 2027 a summer series needs seven slots.
- Schedule design models select flights and fleets together, adjusting demand by recapture or explicit corrections; a retiming is worth making when the network's expected contribution, net of disruption cost, beats the alternative.

Table 13.5 Key formulas of the chapter.

Quantity	Formula
Expected schedule delay, f departures centred in equal sub-periods	$\mathrm{SD} = H/(4f)$
Total trip time	$T(f) = t_0 + t_b + K/f$, $K = H/4$
Frequency elasticity of demand	$\eta_f = \lvert\eta_T\rvert\ (K/f)/T(f)$
S-curve market share	$\mathrm{MS}_i = f_i^{\alpha} / \sum_j f_j^{\alpha}$
Symmetric frequency equilibrium	$f^* = \alpha R/(4c)$
Feasible connection	$\mathrm{MCT} \le d_j - a_i \le \bar{\tau}$
Connections, B banks against rolling	$n^2/B - n$ against $n^2(\bar{\tau} - \mathrm{MCT})/H$
Fleet bounds	$N \ge \mathrm{BH}/h$ and $N \ge \max_t A(t)$
Minimum fleet, one type	$N^* = A(t_0) + \sum_s \max_t[-N_s(t)]$
Historic precedence	operated slots $\ge 0.8\times$ slots in the series

13.11 Further reading

Belobaba's schedule development lecture (Belobaba 2006f) is the short introduction to frequency planning, timetables and their constraints, and his economics review (Belobaba 2006d) supplies total trip time and the S-curve. Barnhart's schedule design lecture (Barnhart 2003) develops the incremental models of Section 13.8, published by Lohatepanont and Barnhart (2004); Barnhart and Cohn (2004) map the whole schedule planning problem, and Rexing and colleagues (2000) introduce time windows into fleet assignment. The handbook's schedules chapter (Network Planning Handbook 2026) contains the screening rule, rotation arithmetic, the Seville case and the retiming exercise. The Worldwide Airport Slot Guidelines (IATA, ACI, and WWACG 2025) are short enough to read in full before a first slot conference. Douglas and Miller (1974) introduced frequency and stochastic delay into airline economics, Jiang (2006) treats de-peaking, and Belobaba, Odoni and Barnhart (2015) cover schedule optimization at textbook length.

CHAPTER

14

Fleet Assignment

IN THIS CHAPTER

- Where fleet assignment sits in planning, and why choosing an aircraft type is a network decision rather than a flight-by-flight one.
- Time-space networks: flight, ground and wrap-around arcs, and the count line that turns aircraft flows into aircraft.
- The fleet assignment model (FAM), every constraint derived, with spill costs from Chapter 9.
- Why the model solves quickly, and what node consolidation, islands and branch and bound do.
- A sixteen-flight Laurentian bank through Montreal, solved exactly and compared with a planner's hand-built lines of flying.
- Network effects: why leg-based spill misprices connecting passengers, and the itinerary-based model built on the passenger mix model with recapture.
- Re-fleeting close to departure (demand driven dispatch), and the links to schedule design, maintenance and crews.

DRAWS ON Barnhart, Airline fleet assignment (MIT 16.75J, 2006); Barnhart, Airline schedule planning, lectures 1 to 6 and problem set 2 (MIT 1.206J, 2003); Belobaba, Airline schedule development (2006); Boeing, Demand driven dispatch; handbook; Hane et al. (1995); Barnhart, Kniker and Lohatepanont (2002)

By the time fleet assignment runs, the timetable exists. Schedule design (Chapter 13) has fixed which flights operate and when: the 09:00 to Vancouver, the 16:20 back from Toronto, the late departures that put aircraft to bed in Quebec City and Boston. What remains open is which aircraft type flies each departure. On one flight the question is the gauge choice of Chapter 9: a 137-seat A220-300 is cheaper to fly, a 169-seat 737-8 spills fewer passengers. On a schedule the question changes character. Every aircraft that lands must take off again, each type exists in a fixed number, and an aircraft placed on the morning Vancouver flight is unavailable for the morn-

ing Toronto peak. The best type for each flight in isolation is almost never a feasible plan.

Fleet assignment is where the spill economics of Chapter 9, the fleet of Chapter 11 and the timetable of Chapter 13 meet. This chapter builds the standard model from first principles, solves a small Laurentian network exactly, and then examines the model's main weakness: it values passengers flight by flight, while passengers buy itineraries. The chapter closes with re-fleeting close to departure and with the links to aircraft routing and crews (Chapter 15). The linear and integer programming used here is collected in Appendix B.

14.1 The problem and its place in planning

Airline schedule planning is usually split into four steps solved in sequence: schedule design selects the flights, fleet assignment chooses an aircraft type for each flight so that contribution is maximized, aircraft routing chains the flights into maintenance-feasible sequences for individual aircraft, and crew scheduling builds legal work schedules (Barnhart 2006, 2003). In Belobaba's description of schedule development, timetables and aircraft rotations are defined two to six months before departure, after frequencies, and revisions continue until the day of operation (Belobaba 2006f).

DEFINITION 14.1 Fleet assignment

Given a daily schedule of flights $i \in \mathcal{F}$, a set of fleet types $k \in \mathcal{K}$ with N_k aircraft of type k, minimum turn times by type and station, the cost of operating each flight with each type, and demand and fare data, the **fleet assignment problem** chooses one type for every flight so as to maximize total contribution (revenue minus the costs that depend on the choice), subject to two physical requirements: aircraft of each type must flow in balance through every station, and no type may be used more than it exists.

Three features of the definition matter for the analyst. First, the decision is a type, not a tail: fleet assignment says that the 09:00 to Vancouver is a 737-8, not which 737-8 (Network Planning Handbook 2026). Assigning individual aircraft, with their maintenance due dates, is aircraft routing (Chapter 15). Second, the objective counts only costs and revenues that change with the type: block-hour and departure costs, and the revenue lost to spill. Ownership

costs and overheads are sunk at this horizon, because the fleet is given. Third, a "type" is any group of interchangeable aircraft: two cabin configurations of one airframe are two fleets if they sell different products.

The objective can be written as a cost or as a contribution. Barnhart's single-flight teaching example has demand of 100 passengers at USD 100 and four types with 80, 100, 120 and 150 seats, costing USD 5,000 to 8,000 per trip (historical teaching numbers) (Barnhart 2006). Operating cost plus lost revenue is USD 7,000, 6,000, 7,000 and 8,000; contribution is USD 3,000, 4,000, 3,000 and 2,000. The two lists rank the types identically because they add to the same constant, the revenue of all demand. In general, for flight i and type k,

$$\text{contribution}_{ik} = \underbrace{r_i\,\mu_i}_{\text{constant}} - \underbrace{(\text{operating cost}_{ik} + \text{spill cost}_{ik})}_{c_{ik}},$$

so minimizing $\sum c_{ik}$ and maximizing contribution are the same problem. We use whichever reads better and call c_{ik} the **assignment cost**.

14.2 Time-space networks

An aircraft is a unit of flow that moves through stations and time. The natural picture is a network whose nodes are events at a station and whose arcs are either flights or periods on the ground.

> **DEFINITION 14.2 Time-space network**
>
> For each fleet type k, the **time-space** (or time-line) **network** has one horizontal line per station. A **departure node** sits at the departure time of each flight. An **arrival node** sits at the flight's **ready time**, its arrival time plus the minimum turn time of type k at that station. A **flight arc** joins the departure node to the arrival node. **Ground arcs** join consecutive nodes on each station line and carry aircraft that wait. For a schedule that repeats every day, a **wrap-around arc** joins the last node of each station to its first node, so aircraft on the ground at night start the next day. A vertical **count line** at a chosen time crosses some ground arcs and some flight arcs; the aircraft on them are all the aircraft of that type in service.

Ending the flight arc at the ready time, rather than the arrival time, builds the turn into the network: an aircraft can take any departure

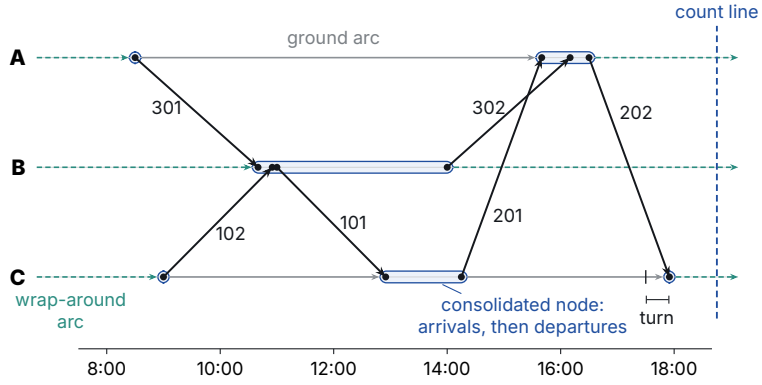


Figure 14.1 The time-space network of the Boeing 737-500 in Cheapo Airlines, the six-flight problem set of MIT 1.206J (Barnhart, 2003). Flight arcs end at the ready time, arrival plus the 25-minute turn. Shaded boxes are consolidated nodes: blocks of arrivals followed by departures. The wrap-around arcs carry the aircraft that stay overnight; the count line crosses them.

at or after its ready time, so every path through the network is an operable sequence for that type (Barnhart 2003). Because turn times differ by type, so do the networks. Without wrap-around arcs every arc points forward in time and the network is acyclic, which is what makes shortest paths on it cheap (Appendix B).

An older representation, used by Abara (1989) at American Airlines, joins each arriving flight directly to every departing flight it can feed, one **connection arc** per feasible turn. Connection arcs can forbid particular turns, but their number grows with the square of the departures at a station; ground arcs grow linearly, which is why large models use the time line (Barnhart 2003).

WORKED EXAMPLE 14.1 Building the network for Cheapo Airlines

Problem set 2 of the MIT schedule planning course gives a fictitious carrier with six daily flights between airports A, B and C, and two types: the Boeing 737-500 with 108 seats and a 25-minute turn, and the Airbus A320 with 144 seats and a 35-minute turn, at most two of each (Barnhart 2003). Flight 301 leaves A at 08:30 and arrives at B at 10:15; flight 102 leaves C at 09:00 and arrives at B at 10:30; flight 101 leaves B at 11:00 for C; flight 302 leaves B at 14:00 for A; flight 201 leaves C at 14:15 for A; flight 202 leaves A at 16:30 for C.

Each flight creates one departure node and one arrival node, so each type has 12 nodes, 12 ground arcs (one after each node, the last one wrapping around) and 6 flight arcs. The full model has $2 \times 6 = 12$ assignment variables and 24 ground variables, 36 in all.

At B the 737-500 sees, in order, the ready times of 301 (10:40) and 102 (10:55), then the departures of 101 (11:00) and 302 (14:00): arrivals followed by departures. Any aircraft in that block can take either departure, so the four events collapse into one node (Figure 14.1). The A320 sees 301 ready at 10:50, 101 leaving at 11:00, 102 ready at 11:05 and 302 leaving at 14:00. The aircraft from 102 cannot take 101, so B keeps two nodes. Consolidating every such block, without merging across the count line, leaves 6 nodes for the 737-500 and 7 for the A320: 13 ground arcs and 25 variables instead of 36. If the count line is ignored, the evening arrivals can also be merged with the next morning's departures, which gives 9 nodes and 21 variables; the extra nodes are the price of keeping a ground arc on the count line, where the model counts aircraft.

The planner's lesson is that the network, not the timetable, decides which aircraft can feed which flight, and the network differs by type. Five minutes of turn time change what the A320 can do at B.

The count line deserves a moment, because it is how a flow model counts aircraft. Let $T_k(t)$ be the number of type-k aircraft crossing a vertical line at time t: on ground arcs that span t plus on flight arcs whose departure-to-ready interval contains t.

RESULT 14.1 The count line can be drawn anywhere

If the flows of type k satisfy balance at every node, $T_k(t)$ is the same for every t. The number of aircraft a solution uses is therefore the flow across any single count line.

Proof. Take two times $t < t'$ and the nodes with times in $(t, t']$. Every arc crossing t but not t' ends at one of these nodes; every arc crossing t' but not t starts at one of them; arcs crossing both, or neither, are irrelevant. Summing the balance equation (inflow equals outflow) over these nodes, the inflow consists of the arcs crossing t but not t' plus arcs starting and ending inside the window, and the outflow of the arcs crossing t' but not t plus the same internal arcs. Hence the flow crossing t equals the flow crossing t'. With wrap-around arcs the same argument holds on the cyclic day. ∎

A second result tells us how many aircraft any given assignment needs. Fix the flights given to type k and a station s. Order the events from the count time and let $A_s(t)$ and $D_s(t)$ count the ready arrivals and the departures of type k at s up to time t. If g_s aircraft sit at s at the count time, the number on the ground at time t is $g_s + A_s(t) - D_s(t)$.

RESULT 14.2 Aircraft needed by an assignment

An assignment is balanced for type k if $A_s = D_s$ over the whole day at every station. It then needs

$$n_k = \sum_s \max_t [D_s(t) - A_s(t)] + |\mathcal{F}_k^0|$$

aircraft of type k, where $\mathcal{F}_k^0$ is the set of its flights whose arcs cross the count line. The maximum includes the count time itself, where both counts are zero, so it is never negative.

Proof. The ground count must be non-negative at all times, so $g_s \geq \max_t [D_s(t) - A_s(t)]$; the smallest feasible value is this maximum, and any larger value parks idle aircraft. Balance over the day makes the count at the end of the day equal to g_s, so the pattern repeats. Adding the aircraft on flight arcs crossing the count line gives the flow across it, which by Result 14.1 is the fleet used. ∎

The maximum in Result 14.2 is the deepest deficit of departures over arrivals: the moment in the day when the station has sent out the most aircraft it has not yet received back. A spreadsheet that

tracks aircraft on the ground, station by station, computes exactly this quantity.

14.3 The fleet assignment model

The model in its standard form is due to Abara (1989) and Hane et al. (1995) and is the one taught in both MIT courses (Barnhart 2003, 2006). Its variables are $x_{ik} = 1$ if flight i is flown by type k (0 otherwise), and $y_g \geq 0$, the number of type-k aircraft on ground arc g of the network of type k. For a node v of that network, $I(v)$ and $O(v)$ are the flights arriving at and departing from v, and y_{v^-} and y_{v^+} are the ground arcs entering and leaving v. Write $\mathcal{G}_k^0$ for the ground arcs and $\mathcal{F}_k^0$ for the flights of type k that cross the count line.

$$
\begin{aligned}
\min \ & \sum_{k\in\mathcal{K}} \sum_{i\in\mathcal{F}} c_{ik}\, x_{ik} \\
\text{s.t. } & \textstyle\sum_{k\in\mathcal{K}} x_{ik} = 1, && i \in \mathcal{F}, \\
& y_{v^-} + \textstyle\sum_{i\in I(v)} x_{ik} - y_{v^+} - \sum_{i\in O(v)} x_{ik} = 0, && v \in \mathcal{V}_k,\ k \in \mathcal{K}, \\
& \textstyle\sum_{g\in\mathcal{G}_k^0} y_g + \sum_{i\in\mathcal{F}_k^0} x_{ik} \leq N_k, && k \in \mathcal{K}, \\
& x_{ik} \in \{0,1\}, \quad y_g \geq 0.
\end{aligned}
\tag{14.1}
$$

Cover constraints. Every flight in the timetable is flown by exactly one type. The timetable is an input; a flight that should perhaps not operate is a question for schedule design. Combinations that are not allowed (a type without the range, an airport that cannot take a widebody, a noise rule) are removed from the model, which is the same as giving them an infinite cost (Barnhart 2003).

Balance constraints. At every node of every type's network, aircraft that arrive on the ground arc from the previous node or on flights equal aircraft that leave on the ground arc to the next node or on flights. This is conservation of flow: aircraft are neither created nor destroyed at a station. It is what makes the model a network model. Each type's variables form a circulation in its own network, and the cover constraints tie the types together.

Count constraints. By Result 14.1, the flow across the count line is the number of aircraft of the type in use, which may not exceed the number available. If the count constraint binds, its dual price measures the value to the schedule of one more aircraft of that type, within the limits that Appendix B sets out for dual prices.

Integrality. Only the x_{ik} need to be integer. Once they are, the ground flows can always be chosen integral: by Result 14.2 the minimal ground counts are differences of counts of flights.

Feasibility of (14.1) implies that aircraft rotations exist: each type's integer circulation splits into cycles, and each cycle is a line of flying for one or more aircraft. It does not imply that those rotations are maintenance-feasible, which is why routing comes next (Barnhart 2006).

Side constraints. Real models add rules that the basic form ignores, each of which is a linear constraint (Barnhart 2003; Clarke et al. 1996):

- a **required through flight**, two flights sold under one flight number with no change of aircraft, needs $x_{ik} = x_{jk}$ for every k;
- **maintenance**: at least m_k aircraft of type k must spend the night at stations equipped for their checks, a lower bound on the wrap-around ground arcs of those stations;
- **station rules**: type k may not serve station s (delete the variables), or at most a share α of the departures from s may use noisy types, $\sum_{i \in O(s)} \sum_{k \in \mathcal{K}_{\text{noisy}}} x_{ik} \le \alpha \, |O(s)|$;
- **crew hours**: the block hours flown by a crew group must fit its available hours, $\sum_i \sum_{k \in \text{group}} \text{BH}_i \, x_{ik} \le H$.

Side constraints are cheap to write and expensive to solve with, because they break the network structure that keeps the model easy (Section 14.5).

Fleet assignment prices the gauge of every flight in the currency of the whole schedule: a type is worth having on a flight only if the network can bring an aircraft of that type there and take it away again.

14.4 Spill costs by flight and type

The operating cost OC_{ik} of flying flight i with type k is the straightforward part: block-hour cost times block hours, plus a cost per departure. The revenue side is the hard part, because the model must say, for every flight and every type, how much revenue the type would lose to spill.

The lecture version uses deterministic demand: spill cost equals average fare times $\max(0, D_i - C_k)$ (Barnhart 2003). Chapter 9

showed why this understates spill: demand varies from day to day, and a flight whose mean demand fits in the aircraft still spills on its busy days. The natural coefficient uses the expected spill of normal demand with mean μ_i and standard deviation σ_i, net of recapture at rate b_i:

$$c_{ik} = \mathrm{OC}_{ik} + (1 - b_i)\, r_i\, \sigma_i\, L\!\left(\frac{C_k - \mu_i}{\sigma_i}\right), \tag{14.2}$$

where C_k is the seat count of type k, r_i the contribution per passenger (fare net of passenger-variable cost), and L the standard normal loss function. The contribution of the pair is then $r_i\mu_i - c_{ik}$.

Two inputs need care. The **recapture rate** b_i is the share of spilled passengers who travel on another of the airline's flights. A classical way to set it is from the quality of service index (QSI) of the alternative: a passenger turned away from itinerary p and offered another itinerary r (here r names an itinerary, not a contribution) is assumed to accept with probability $\mathrm{QSI}_r/(\mathrm{QSI}_r + m_c)$, where m_c is the competitors' share of the market (Barnhart 2003). If Laurentian's other same-day Toronto flight has a QSI of 0.18 and competitors hold 42 percent of the market, $b = 0.18/(0.18{+}0.42) = 0.30$. Modern practice estimates such diversion rates with the choice models of Chapters 5 and 6.

The **average fare** of a flight is harder than it looks, because revenue belongs to itineraries and a connecting passenger pays one fare for two flights. Any leg fare is an allocation. The MIT lecture lists three rules: **full** (credit the whole itinerary fare to each leg), **partial** (divide it by the number of legs) and **shared** (divide it among the legs that are short of seats); and two ways to price the spilled passengers: a single **representative fare**, or **integration**, which spills the lowest allocated fares first (Barnhart 2003). On a three-fleet, 226-flight test problem, the best representative-fare rule produced an assignment USD 2,600 per day worse than the optimum, while shared fares with integration came within USD 8 per day (historical results) (Barnhart 2003). We return to why no rule is right in general in Section 14.7.

WORKED EXAMPLE 14.2 Pricing the Vancouver departure

Laurentian's 09:00 YUL–YVR (flight 6 of the instance in Section 14.6) has mean demand $\mu = 172$ with $k = 0.30$, so $\sigma = 51.6$. The demand includes 48 connecting passengers from Quebec City, Boston and New York. Prorating their fares by distance and averaging with the local fare of CAD 430 gives a leg fare of CAD 455.65; net of the CAD 20 passenger cost, $r = 435.65$. Spill from this flight is assumed lost ($b = 0$).

For the A220-300, $z = (137 - 172)/51.6 = -0.678$ and $L(z) = 0.8265$, so expected spill is $51.6 \times 0.8265 = 42.647$ passengers and the spill cost is \$435.65 ×42.647 = \$ CAD 18,579. The operating cost for 5.2 block hours is \$8{,}000 ×5.2 + 2{,}000 = \$ CAD 43,600, as in Chapter 3. With unconstrained contribution \$435.65 ×172 = \$ CAD 74,932, the A220 contributes \$74{,}932 - 43{,}600 - 18{,}579 = \$ CAD 12,753.

For the 737-8, $z = -0.058$, $L = 0.4287$, expected spill 22.12 and spill cost CAD 9,637; the operating cost is \$9{,}000 ×5.2 + 2{,}500 = \$ CAD 49,300, and the contribution CAD 15,995. The 737-8 saves CAD 8,943 of spill for CAD 5,700 of extra operating cost: it is worth CAD 3,243 more on this departure.

Recapture changes such numbers where alternatives exist. On the 16:20 YYZ–YUL (flight 12), 144 of the 178 passengers are locals who could take the 10:50 instead; with a 30 percent rate for them, $b_{12} = 0.30 \times 144/178 = 0.243$, and the A220's spill cost falls from CAD 8,520 to CAD 6,452.

The planner learns two things. A flight that looks almost right-sized at its mean (172 passengers for 169 seats, a deterministic spill of 3) still spills 22 passengers a day on average. And the coefficient is only as good as the leg fare and recapture rate behind it, both of which are modelling choices rather than data.

Figure 14.2 shows the coefficients for the whole instance. The 737-8 earns its extra seats on the Vancouver round trip and, more modestly, on the two Toronto peaks. Everywhere else its extra operating cost exceeds the spill it saves. The chart already hints at the answer of Section 14.6, and at why it is not the obvious one: the 737-8 must spend the night somewhere, and every overnight station costs it money.

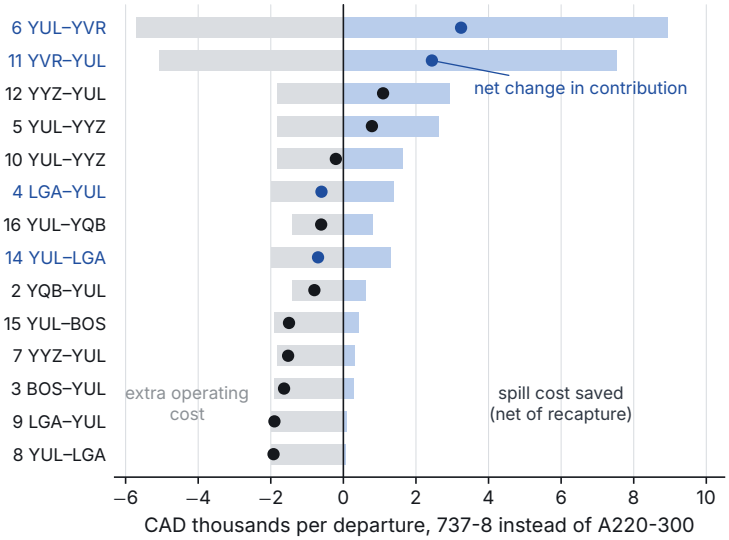


Figure 14.2 Where the larger aircraft pays. For each flight either narrowbody can fly, the spill cost (net of recapture) the 737-8 saves over the A220-300 and its extra operating cost; the dot is the net change in contribution. Blue labels mark the flights that the optimal assignment of Section 14.6 gives to the single 737-8.

14.5 Solving the model

Model (14.1) is an integer program with one binary variable per eligible flight-type pair. For an airline with 2,000 flights and ten types that can mean twenty thousand binaries, yet the model solves quickly. Three properties explain why.

A tight linear relaxation. Relaxing $x_{ik} \in \{0,1\}$ to $0 \le x_{ik} \le 1$ gives a linear program whose optimum bounds the integer optimum (Appendix B). Without the cover and count constraints, each type's problem would be a pure network flow problem, whose linear relaxation has integral optimal solutions. The cover constraints couple the types and the count constraints cap them; both allow fractional solutions, but in practice the relaxation is usually very close to integral. On Barnhart's computational sample of 2,044 legs and nine fleet types (18,487 columns, 7,827 rows), the root linear relaxation was USD 21,401,658 per day and the best integer solution USD 21,401,622: a gap of USD 36, solved in 974 seconds on the hardware of the time (Barnhart 2006). Hane et al. (1995) combined an interior point method, the dual steepest-edge simplex method, cost perturbation against degeneracy, model aggregation and branching on whole cover constraints, and reported optimality gaps of at

most 0.02 percent, more than two orders of magnitude faster than a default branch-and-bound code (historical).

Node consolidation. The second property is that much of the network is redundant.

RESULT 14.3 Node consolidation is exact

At each station, with events ordered from the count time, merge every maximal block of consecutive arrival nodes followed by consecutive departure nodes into a single node. The merged network has exactly the same feasible assignments as the original one.

Proof. Inside a block, the ground arcs carry the aircraft present before the block plus the arrivals so far minus the departures so far. Because every arrival in the block precedes every departure, this count rises and then falls, so it never drops below the smaller of the counts on the ground arcs entering and leaving the block. Requiring those two arcs to be non-negative therefore implies that all the internal arcs are non-negative, and the internal arcs can be deleted without changing which flight assignments are feasible. Because no block straddles the count time, the arcs that cross the count line, and with them the count constraints, are unchanged. ∎

Islands. The third property is that, at stations that are not maintenance bases, the minimum number of aircraft on the ground at some moment of the day can be taken to be zero: aircraft are too expensive to leave at a spoke all day (Barnhart 2003). The total number of aircraft on the ground at such a station, all types together, is then fixed by the schedule (up to differences in turn times) and is zero at the moments when departures most exceed arrivals. Any ground arc spanning such a moment carries no aircraft of any type and can be deleted, which cuts the station's time line into separate **islands**. An island holding one arrival and one departure turns the two flights into a through pair: the same aircraft must fly both. Unlike node consolidation, islands rest on an assumption. Two types can take turns on the ground so that a station is never empty although no aircraft idles all day, and islands exclude such plans; they need extra aircraft and rarely pay, but they can be optimal (Example 14.4).

Branch and bound. When the relaxation is fractional, branch and bound fixes variables and re-solves (Appendix B). Fixing $x_{ik} = 1$ is a big decision and fixing $x_{ik} = 0$ a small one, so the branches

are unbalanced. Branching on special ordered sets, setting a group of variables of one flight to sum to zero or to one, gives balanced branches and fewer linear programs, and the search can stop once the gap between the best solution and the bound is small enough (Barnhart 2003).

WORKED EXAMPLE 14.3 Through flights make the relaxation fractional

Take the Laurentian bank of Section 14.6. Without side constraints, its linear relaxation is integral and the solver stops at the root node. Now suppose sales want two one-stop services sold under single flight numbers: Boston–Vancouver in the morning (flights 3 and 6) and Vancouver–New York in the evening (flights 11 and 14). Each through pair adds the constraints $x_{3k} = x_{6k}$ and $x_{11,k} = x_{14,k}$ for every type.

The relaxation becomes CAD 210,122.0, with twenty variables at exactly one half on ten flights. Half a 737-8 spends the night in Boston and half in New York; in the morning the Boston half takes the 08:45 to Toronto and the New York half the 09:00 to Vancouver. Each through constraint holds as an equation between halves, although no aircraft actually flies from Boston to Vancouver.

Branch on $x_{6,737\text{-}8}$. With $x_{6,737\text{-}8} = 1$ the relaxation is infeasible: the single 737-8 would have to spend the night in both Boston and New York. With $x_{6,737\text{-}8} = 0$ the relaxation is integral at CAD 208,162.7, with the 737-8 on the Quebec City line. All branches are resolved, so this is optimal after three linear programs. The integrality gap was CAD 1,959, or 0.94 percent.

Two lessons follow. The two through flights together cost CAD 5,640 per day, while the Boston–Vancouver through alone costs CAD 1,831 and the evening one alone nothing: side constraints interact. And with several aircraft of a type, a through constraint on types does not by itself put the same aircraft on both flights; routing must enforce that.

14.6 A Laurentian bank solved

We now solve a complete small instance. All its data are illustrative; the seat counts, unit costs and passenger costs are those of the running case.

Laurentian operates 16 daily flights through Montreal (Table 14.1), all times in Montreal time. Narrowbodies serve Quebec City, Boston and New York LaGuardia, spending the night there, plus Toronto twice a day and Vancouver once. A daily Paris rotation uses the 787-9, which leaves Montreal at 19:00 and returns at 12:25. Three types are available for this sub-network: two A220-300s (137 seats, 35-minute turn), one 737-8 (169 seats, 40 minutes) and one 787-9 (298 seats, 75 minutes). Paris is reserved for the 787-9, for its cabin product; the 787-9 may also fly the Toronto trunk during its afternoon at Montreal. Demand comes from 24 itineraries, one local itinerary per flight and 8 connections through Montreal such as Quebec City–Vancouver and Toronto–Paris, with $k = 0.30$. Leg fares prorate connecting fares by distance. Local passengers on Toronto and New York flights are recaptured on the other same-day frequency at rate 0.30. Costs follow the running case's trip cost model.

Table 14.1 The Laurentian instance (illustrative). Times are Montreal time; μ is mean leg demand and Fare the prorated average fare (CAD). The last three columns are expected contributions $r_i \mu_i - c_{ik}$ per departure (CAD); a blank cell means the type may not fly the flight.

No.	Flight	Times	μ	Fare	A220	737-8	787-9
1	CDG–YUL	04:30–12:25	268	655			20,807
2	YQB–YUL	06:15–07:10	112	181	8,006	7,207	
3	BOS–YUL	06:30–07:55	100	223	5,702	4,068	
4	LGA–YUL	06:40–08:10	128	255	12,890	12,287	
5	YUL–YYZ	08:45–10:05	160	230	16,111	16,897	7,929
6	YUL–YVR	09:00–14:12	172	456	12,753	15,995	
7	YYZ–YUL	10:50–12:10	104	230	8,831	7,309	−3,827
8	YUL–LGA	11:00–12:30	90	280	8,422	6,498	
9	LGA–YUL	13:15–14:45	92	280	8,896	6,996	
10	YUL–YYZ	13:45–15:05	140	194	9,310	9,103	−1,251
11	YVR–YUL	15:15–19:50	158	454	16,394	18,833	
12	YYZ–YUL	16:20–17:40	178	198	12,613	13,707	6,034

No.	Flight	Times	μ	Fare	A220	737-8	787-9
13	YUL–CDG	19:00–01:55	280	656			40,052
14	YUL–LGA	21:15–22:45	126	257	12,954	12,256	
15	YUL–BOS	21:30–22:55	104	228	6,805	5,307	
16	YUL–YQB	22:00–22:55	116	185	8,877	8,265	

WORKED EXAMPLE 14.4 **Sixteen flights, four aircraft**

The model. The node-consolidated networks of the three types have 33 nodes and 33 ground arcs; with the 34 eligible assignment variables the model has 67 variables and 52 constraints. Islands delete 17 of the ground arcs, all at spokes, leaving 50 variables; with this fleet they do not change the optimum. (With four A220s and two 737-8s they would: they exclude the plan of the upper bound below, in which a 737-8 waits at Toronto by day and an A220 spends the night there.) Solved with the HiGHS solver through `scipy.optimize.milp` (Huangfu and Hall 2018), the linear relaxation is already integral, so branch and bound stops at the root node, after about ten milliseconds.

The optimum is CAD 213,803 per day (Figure 14.3). The 787-9 flies only Paris: the Toronto round trip that fits its afternoon at Montreal (flights 10 and 12) would contribute CAD 4,783 with the 787-9 against CAD 21,923 with A220s. The single 737-8 spends the night in New York and flies LaGuardia–Montreal, the Vancouver round trip, and Montreal–LaGuardia (flights 4, 6, 11 and 14). The two A220s do everything else. The count line at 01:00 crosses four aircraft: two A220s on the ground at Quebec City and Boston, the 737-8 at LaGuardia, and the 787-9 in the air to Paris, exactly the fleet available.

A naive plan. Before running the model, the planner built the schedule as four lines of flying, one per aircraft: a Quebec City line (2, 5, 7, 10, 12, 16), a Boston line (3, 6, 11, 15), a New York line (4, 8, 9, 14) and the Paris line. Choosing the best type for each whole line, the 737-8 goes on the Boston line, worth CAD 2,550 per day more than an A220 there (on the Quebec City line

it would lose CAD 1,259, on the New York line CAD 5,125). This plan earns CAD 211,972.

Why the model does better. The model is not bound to the planner's lines. In the time-space network the Boston and New York aircraft are both on the ground at Montreal on the morning of each day, and again in the evening, so they can exchange their work (Figure 14.4). The 737-8 still flies to Vancouver, but it spends the night in New York instead of Boston. On the Boston flights it would lose CAD 1,634 and CAD 1,498 against an A220; on the LaGuardia flights only CAD 603 and CAD 698, because New York demand (128 and 126) is closer to the 737-8's size. The swap gains CAD 1,831 per day, about CAD 668,000 per year, on four aircraft.

An upper bound. Giving every flight its best type regardless of balance and fleet size earns CAD 216,985, putting the 737-8 on flights 5, 6, 11 and 12. That plan needs four A220s and two 737-8s, six narrowbodies instead of three. The network constraints cost CAD 3,182 per day, the price of flying real aircraft.

The value of an aircraft. Re-solving with one more aircraft shows what each is worth to this timetable: a third A220 raises contribution by CAD 1,301 per day and a second 737-8 by CAD 887, while a second 787-9 adds nothing. The solver's dual prices of the count constraints are CAD 887 for the 737-8 and zero for the 787-9, but CAD 1,410 for the A220. The difference comes from degeneracy: with one A220 fewer the timetable cannot be flown at all, so the dual price of the A220 count is not unique, and any value from CAD 1,301 up is a valid dual; the solver reports one of them (Appendix B.3). Without the 737-8 or the 787-9 the model is infeasible: the timetable needs four aircraft, and changing that is a schedule design question (Chapter 13).

The example makes a general point about where optimization earns its keep. Each individual choice in the optimum is easy to defend; what a planner misses is the combination, because it requires breaking lines of flying that were built by hand and look natural.

Reported gains from implementations are of a similar relative size to the 0.9 percent found here, although each airline measured them against its own previous practice rather than in a controlled comparison (all historical). Abara reported, for one application at American Airlines, operating costs 0.4 percent lower and an operating margin 1.4 percent higher (Abara 1989). Delta's Coldstart model, covering

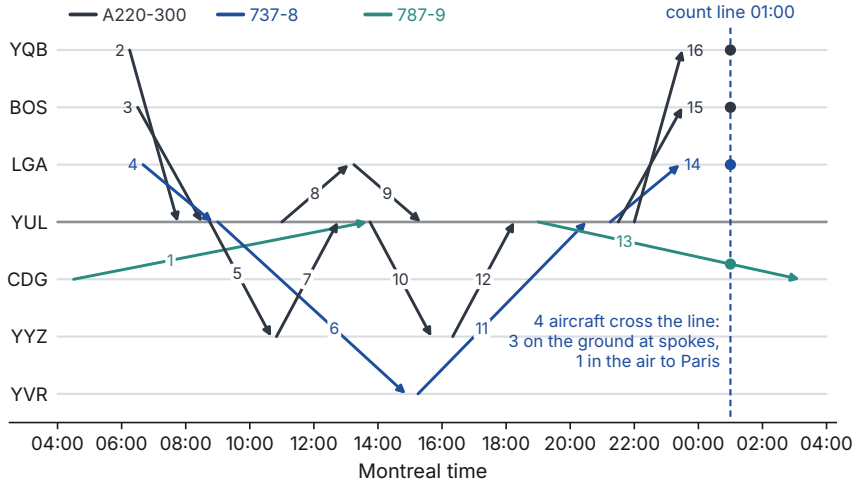


Figure 14.3 The optimal fleet assignment of the Laurentian bank in its time-space network. Arcs run from departure to ready time and are coloured by type; numbers are flights (Table 14.1). At the 01:00 count line the model counts four aircraft: three on the ground at spokes and the 787-9 on its way to Paris.

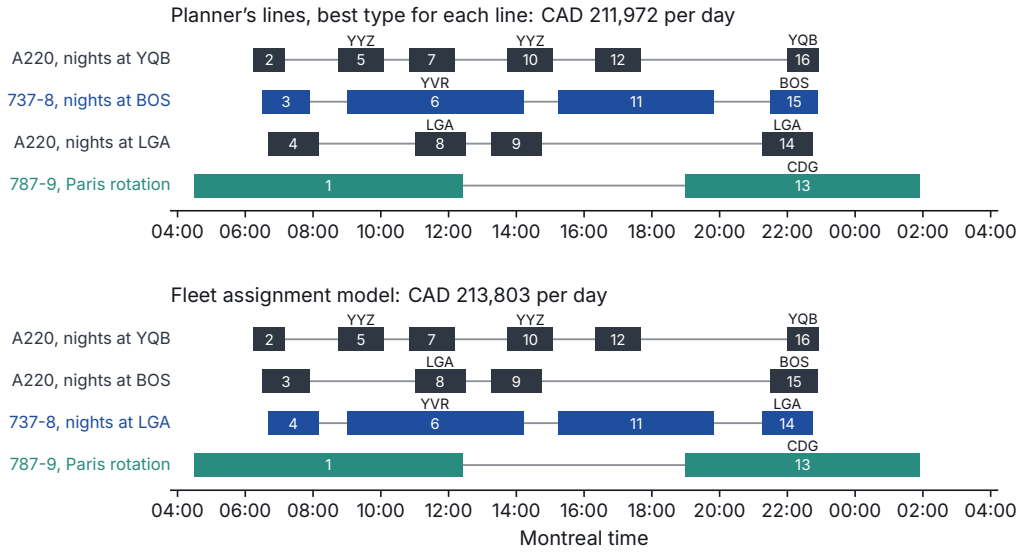


Figure 14.4 The planner's lines of flying with the best type for each line (top) and the rotations implied by the optimal assignment (bottom). The model exchanges the daytime work of the Boston and New York aircraft at Montreal, so that the 737-8 spends its nights where its extra seats are worth most while still flying to Vancouver.

about 2,500 daily flights flown by some 450 aircraft in 10 fleets, was expected to save USD 300 million over three years (Subramanian et al. 1994). USAir reported an annual benefit of at least USD 15 million from its model (Rushmeier and Kontogiorgis 1997).

> **PITFALL**
>
> Reading the fleet-size duals as aircraft values for fleet planning. Even the CAD 1,301 per day that a third A220 adds is its value to this timetable, which was built for four aircraft. A real fleet decision changes the schedule too, and its value comes from the flights the aircraft would enable (Chapters 11 and 12).

14.7 Network effects: itinerary-based fleet assignment

The model of Section 14.3 values each flight as if its passengers flew only that flight. Passengers buy itineraries, and a connecting passenger needs a seat on every leg. This creates two errors that no choice of leg fare can remove (Barnhart 2003, 2006). With **network effects**, a passenger spilled from one leg also frees a seat on the other legs of the itinerary, so the spill cost of a leg depends on the capacity of the others; the leg-based model implicitly allows a passenger to be spilled from some legs of the trip but not others. With **recapture**, a spilled passenger may take another flight of the airline, but only if it has a seat.

Barnhart's illustration has two flights, X–Y and Y–Z, and a fleet of 100-seat (A) and 200-seat (B) aircraft. Demand is 75 X–Y passengers at USD 200, 150 Y–Z at USD 225 and 75 X–Z connecting at USD 300 (historical teaching numbers). Table 14.2 compares two leg-based spill estimates, both spilling the lowest allocated fares first, with the true optimum computed on the whole network.

Table 14.2 Spill cost (USD per day) of four assignments of two flights, by three methods (adapted from Barnhart, 2003; recomputed). Partial allocation credits the X–Z fare half to each leg, full allocation credits all of it to each leg.

Types	Partial fare	Full fare	Network optimum	Passengers spilled
A–A	30,000	38,125	31,875	50 X–Z, 75 Y–Z
A–B	11,250	15,625	12,500	25 X–Z, 25 X–Y
B–A	22,500	28,125	28,125	125 Y–Z
B–B	3,750	5,625	5,625	25 Y–Z

Partial allocation understates the spill of all four assignments; full allocation overstates it on two and happens to be exact on the other two. Neither can be right everywhere, and the reason is structural.

RESULT 14.4 Network spill is not leg-additive

There are networks in which no costs $s_1(k_1)$ and $s_2(k_2)$, one per leg and type, reproduce the optimal network spill $S(k_1, k_2)$ of every assignment.

Proof. Take Barnhart's network-effects example: legs A–B and B–C, with 80 local A–B passengers at USD 200, 90 local B–C at USD 250 and 50 A–C connecting at USD 400 (Barnhart 2006). With 80 seats on both legs, the best plan carries the locals and spills all connections and 10 B–C locals: $S(80, 80) = 22{,}500$. With 150 seats on both, nothing is spilled: $S(150, 150) = 0$. With 80 seats on A–B only, the best plan carries all 50 connections and 30 A–B locals: $S(80, 150) = 10{,}000$. With 80 seats on B–C only, it carries the connections and 30 B–C locals: $S(150, 80) = 15{,}000$. Leg-additive costs would require $S(80, 80) + S(150, 150) = S(80, 150) + S(150, 80)$, but $22{,}500 \neq 25{,}000$. ∎

The remedy is to let the model choose which passengers to carry. Its revenue engine is the **passenger mix model** (PMM) of Barnhart, Kniker, and Lohatepanont (2002), developed in Kniker's thesis.

DEFINITION 14.3 Passenger mix model

Given the seats on every flight, the unconstrained mean demand D_p and fare f_p of every itinerary $p \in \mathcal{P}$, and recapture rates b_{pr} (the probability that a passenger who wanted p accepts r when offered it), the **passenger mix model** chooses how many passengers of each itinerary to carry or redirect so as to maximize revenue.

In this section p and r both index itineraries; r is not the contribution per passenger r_i of Section 14.4, which the itinerary model replaces by the fares f_p.

Write $a_{ip} = 1$ if itinerary p uses flight i (the incidence a_{lj} of the book's notation), CAP_i for the seats on flight i, and x_{pr} for the passengers who wanted p and fly r, with $b_{pp} = 1$. The basic model is (Barnhart 2003)

$$\begin{aligned} \max \ & \textstyle\sum_p \sum_r f_r \, x_{pr} \\ \text{s.t.} \ & \textstyle\sum_p \sum_r a_{ir} \, x_{pr} \le \text{CAP}_i, \quad i \in \mathcal{F}, \\ & \textstyle\sum_r x_{pr} / b_{pr} \le D_p, \qquad p \in \mathcal{P}, \qquad x_{pr} \ge 0. \end{aligned}$$

The second constraint counts attempts: to get x_{pr} passengers onto r, the airline must turn x_{pr}/b_{pr} of them away from p. Most passengers fly what they asked for, so it pays to model only the exceptions. Let t_{pr} be the number of passengers who wanted p and whom the airline redirects to r, including a null itinerary $r = \emptyset$ with no legs and no fare, meaning lost. Let $Q_i = \sum_p a_{ip} D_p$ be the unconstrained demand on flight i.

RESULT 14.5 The keypath form of the passenger mix model

With the change of variables $x_{pp} = D_p - \sum_{r \neq p} t_{pr}$ and $x_{pr} = b_{pr}\, t_{pr}$, the passenger mix model is equivalent to

$$\begin{aligned} \min\ & \textstyle\sum_p \sum_r (f_p - b_{pr} f_r)\, t_{pr} \\ \text{s.t.}\ & \textstyle\sum_p \sum_r (a_{ip} - a_{ir}\, b_{pr})\, t_{pr} \geq Q_i - \mathrm{CAP}_i, \quad i \in \mathcal{F}, \\ & \textstyle\sum_r t_{pr} \leq D_p, \quad p \in \mathcal{D}, \qquad t_{pr} \geq 0. \end{aligned} \tag{14.3}$$

Proof. Revenue is

$$\begin{aligned} \textstyle\sum_p f_p x_{pp} + \sum_p \sum_{r \neq p} f_r b_{pr} t_{pr} = {} & \textstyle\sum_p f_p D_p \\ & \textstyle - \sum_p \sum_r (f_p - b_{pr} f_r)\, t_{pr}, \end{aligned}$$

and the first term is a constant, so maximizing revenue is minimizing the net cost of redirections. On flight i, the passengers carried are Q_i, minus those redirected away from itineraries using i, plus those recaptured onto itineraries using i; requiring this to be at most CAP_i gives the capacity row. The demand row says that no more passengers can be redirected away from p than wanted it, which is $x_{pp} \geq 0$ in the new variables; the original demand row then holds automatically, with $t_{p\emptyset}$ in the role of its slack. ∎

In the capacity row, the recaptured passengers $b_{pr} t_{pr}$ occupy seats on the legs of r, the itinerary they move to; some printed versions of the lecture slides place them on the legs of p instead, which is an error. The objective coefficient $f_p - b_{pr} f_r$ is the net revenue lost per passenger redirected, small when the alternative is good and likely to be accepted, which is why most variables stay at zero. Barnhart reports solving the model for 60,215 itineraries in 15,678 markets on 2,037 flights in about 100 seconds by generating columns and rows only as needed (historical, Appendix B) (Barnhart 2003).

The **itinerary-based fleet assignment model** (IFAM) of Barnhart, Kniker, and Lohatepanont (2002) makes the seats a decision. It keeps the cover, balance and count constraints of (14.1) with $c_{ik} = \mathrm{OC}_{ik}$, operating costs only, adds the redirection variables and their cost to the objective, and replaces the fixed capacity in (14.3) by the seats of the assigned type:

$$\textstyle\sum_k C_k\, x_{ik} + \sum_p \sum_r (a_{ip} - a_{ir}\, b_{pr})\, t_{pr} \geq Q_i, \qquad i \in \mathcal{F}.$$

Spill and recapture are then computed consistently with the fleet choice, itinerary by itinerary. The price is size and fractionality: on a

schedule of about 2,000 legs, 76,000 itineraries, 21,000 markets and nine types, the linear relaxation took over 1.5 hours and the integer program over 4 hours, even with row and column generation saving 88 and more than 95 percent of the rows and columns (historical) (Barnhart 2003). Barnhart estimated annual improvements over the basic model of about USD 30 million from network effects and about USD 70 million from recapture for the airline studied, and stressed that these are upper bounds (historical) (Barnhart 2003, 2006). A subnetwork-based model with composite variables, which decides the types of tightly linked legs jointly, matched the itinerary-based model's solutions in the lecture's tests with run times close to the basic model's (Barnhart 2003).

WORKED EXAMPLE 14.5 Cheapo Airlines: legs versus itineraries

Return to the six flights of Example 14.1. Flights 301 and 101 carry 40 A–C passengers connecting at B for USD 130, and flights 102 and 302 carry 50 C–A passengers at USD 130; all other demand is local, at USD 150 on the B flights and USD 100 on flights 201 and 202. The leg data of the problem set credit the full connecting fare to each leg, so flight 301, for example, has 130 passengers at an average of USD 143.85. Operating costs are USD 6,000 to 10,500 per flight for the 737-500 and USD 9,000 to 13,500 for the A320 (historical teaching numbers) (Barnhart 2003).

Leg-based model. With 108 seats the 737-500 would spill 22 to 32 passengers on each of the four B flights at about USD 143 each, which outweighs its lower operating cost. The model puts A320s on all four B flights and the 737-500 on flights 201 and 202, at a cost of USD 66,200 per day.

Itinerary-based model. The cheapest passengers to spill are the connecting ones: each pays USD 130 for a seat on two flights. Flying 737-500s on all six flights and spilling 32 A–C and 32 C–A connecting passengers relieves both legs of each pair at once; only flight 201 must also spill 12 locals. Total spill is USD 9,520 and operating cost USD 51,000, a total of USD 60,520 per day. The leg-based estimate of spill for that same plan is USD 16,668: it spills each leg separately, 54 passengers on each pair of connecting flights where 32 connecting passengers relieve both, and prices them at leg average fares above the USD 130 actually lost. The itinerary-based plan costs 8.6 percent less

than the leg-based one. If 20 percent of spilled A–C passengers would take flight 202, which has eight empty seats on the 737-500, the itinerary-based cost falls to USD 59,880.

The planner's lesson is that a fleet decision can turn on which passengers are spilled, and that leg data quietly decide this with an allocation rule nobody chose deliberately.

On the Laurentian bank, an IFAM with the same 24 itineraries, mean demands, fares net of passenger costs and recapture rates moves the 737-8's nights from New York to Quebec City. The reason is not network effects but averages: at their means of 128 and 126 passengers, the New York flights fit in an A220, so a model that uses mean demand sees no spill there and parks the larger aircraft on the shortest spoke, where its extra cost is lowest. Yet the New York flights exceed 137 passengers on about 40 percent of days. To compare the plans fairly, we simulated 1,000 days of random itinerary demand and allocated passengers optimally on each day, the evaluation method of Barnhart's lectures (Barnhart 2003). The leg-based assignment beats the planner's lines by CAD 1,967 per day (standard error 100), the deterministic IFAM by CAD 1,540 (standard error 71): the itinerary model is CAD 427 per day worse. Barnhart's simulations of hub-and-spoke problems found the opposite ranking, with network effects dominating demand variability (historical) (Barnhart 2003). Which effect dominates is an empirical question about a particular network, not a property of the models.

PITFALL

Treating a deterministic network model's revenue as a forecast. Valued at mean demand, the passenger mix model credits Laurentian's optimal plan with CAD 261,296 per day, while the simulated average is CAD 225,641: demand variability alone removes 14 percent. Network models fix the allocation of revenue to legs, not the variability of demand; use expected spill or simulation for the level, and the deterministic model only to compare plans.

14.8 Re-fleeting close to departure: demand driven dispatch

Fleet assignment is solved months before departure with forecasts of average demand. By the time a flight is half booked, the airline knows much more about that particular day. **Demand driven dispatch**, or D^3, is Boeing's name for reassigning aircraft of different sizes close to departure as bookings reveal demand (Boeing Commercial Airplane Group, n.d.; Berge and Hopperstad 1993).

DEFINITION 14.4 Demand driven dispatch

Demand driven dispatch re-solves the fleet assignment of a published schedule during the booking period, using updated demand forecasts for each departure, and swaps aircraft between flights where aircraft of different capacities, flown by the same crews, are on the ground at the same station at the same time.

The brochure's motivating example (historical) is a 148-seat aircraft in markets where its average load factor is about 65 percent and capacity is rarely exceeded. A 128-seat aircraft could carry the demand on about 80 percent of days, and a 108-seat aircraft on about 60 percent (Boeing Commercial Airplane Group, n.d.). Its chart of 1989 United States domestic trip costs on a 500-nautical-mile sector shows the 128-seat and 108-seat aircraft about 6 and 11 percent cheaper than the 148-seat one. The brochure warns against the obvious conclusion: permanently assigning the smaller aircraft would spill heavily on the busy days. The value lies in using the small aircraft only on the days that do not need the large one.

WORKED EXAMPLE 14.6 The value of a swap

If demand is normal, the brochure's two percentages pin it down: 108 seats suffice 60 percent of the time and 128 seats 80 percent, so $108 = \mu + 0.253\,\sigma$ and $128 = \mu + 0.842\,\sigma$, giving $\sigma = 34.0$ and $\mu = 99.4$. The 148-seat aircraft then has an expected load factor of 66.4 percent, close to the brochure's 65, and is overfull on 7.6 percent of days.

Two departures of this kind leave the same hub bank, one with the 148-seat aircraft and one with the 108-seat aircraft. With a fixed assignment they spill $1.17 + 9.69 = 10.86$ passengers a day between them. Suppose that, when about half the passengers have booked, the forecast of each departure reveals a share ρ of the variance of its demand, and the airline then gives the larger aircraft to the departure with the larger forecast. Costs do not change, since the same two aircraft fly. With independent demands and $\rho = 0.5$, expected spill falls to 6.41 passengers, a saving of 4.45 a day; with perfect information the saving would be 6.34 (Figure 14.5). At an illustrative CAD 150 of contribution per passenger, the swap is worth CAD 667 a day, about CAD 244,000 a year for the pair, or CAD 122,000 per aircraft, which happens to fall inside the brochure's claimed range of USD 50,000 to 200,000 per aircraft per year (historical).

If the two demands are positively correlated, because both respond to the same day-of-week or economic swings, the value shrinks: with a correlation of 0.6 even perfect information saves only 3.96 passengers a day. Swaps pay when demands move independently and when aircraft of different sizes meet on the ground, which is why the brochure looks for them at the connecting banks of a hub.

The constraints are what make D³ hard, and the brochure lists them: dispatch and ground handling, catering, maintenance planning, the reservation system, cabin crews, and above all flight crews (Boeing Commercial Airplane Group, n.d.). A swap is attractive only between types that the same pilots can fly; otherwise every swap breaks a crew pairing. This is why the concept became practical with families of aircraft under one type rating, such as the 737 series of the time. Maintenance routing must survive the swap, since the aircraft due for a check must still reach the base. Flow balance must hold: a swap at a hub exchanges the rest of two aircraft's days until their paths cross again. And the reservation system must

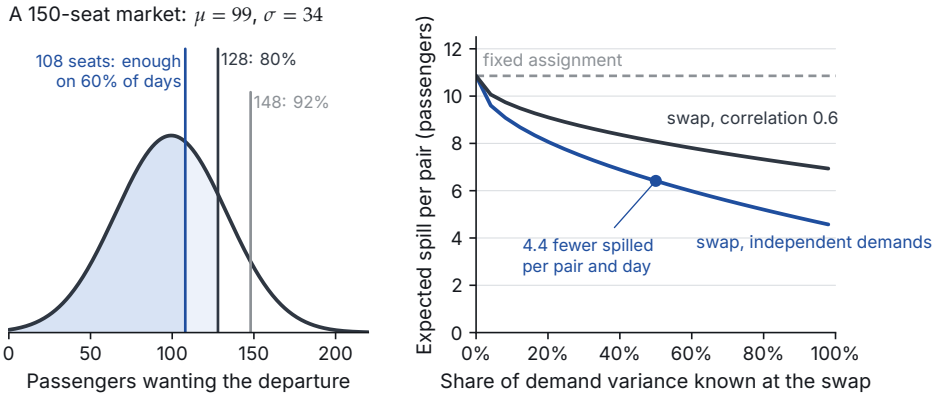


Figure 14.5 Demand driven dispatch. Left: the brochure's "150-seat market", normal demand implied by its statement that 108 seats suffice on 60 percent of days and 128 seats on 80 percent (historical). Right: expected spill of two departures served by one 148-seat and one 108-seat aircraft, with a fixed assignment and with a swap made after a forecast that reveals a share of the demand variance.

re-inventory the flights; an aircraft cannot be downsized below the bookings already taken.

Timing is a trade-off: early in the booking period the forecast is poor; late, the spill has already happened. The brochure places the highest payoff when 40 to 60 percent of passengers have booked; closer to departure, swaps mainly shift flying to smaller aircraft when either would carry the load, and they continue until a day or two before departure, when the dispatch schedule is fixed (Boeing Commercial Airplane Group, n.d.).

The reported benefits are historical and come from the seller of the aircraft. In the brochure's case study of a 40-aircraft fleet serving 22 cities with 244 daily flights, D³ with the same fleet raised revenue by 0.7 percent at almost unchanged operating cost, cut spill from 1,026 to 893 passengers a day and raised operating contribution by 2.9 percent, about USD 4.7 million a year. With a downsized fleet mix (20 smaller, 10 medium and 10 large aircraft instead of 12, 14 and 14, about 5 percent fewer seats), contribution still rose by 2.7 percent with USD 40 million less aircraft investment (Boeing Commercial Airplane Group, n.d.). Berge and Hopperstad (1993) report studies of airline systems showing a net improvement of 1 to 5 percent in operating profit, and formulate the reassignment as a multicommodity network flow problem solved repeatedly by heuristics. The brochure states that its heuristic assigner reached more than 99.9 percent of the optimum on problems small enough for linear programming.

D^3 is fleet assignment re-solved: the same time-space network, restricted to swaps within a crew-compatible family, with the coefficients (14.2) recomputed from booking-based forecasts. Laurentian's A220-300 and 737-8 belong to different families, so D^3 there would need two sizes from one family.

14.9 Fleet assignment in the planning chain

Fleet assignment sits between steps that each assume the others are fixed. Three links matter to the network planner.

Upstream: schedule design. Retiming flights by a few minutes can improve the assignment and free aircraft. The **fleet assignment model with time windows** replaces each flight arc by a cluster of copies at slightly different times and lets the cover constraint choose one copy, keeping the bank structure so that connections still work (Rexing et al. 2000; Barnhart 2003). In the lecture's tests (historical), windows of plus or minus 10 minutes reduced the aircraft required from 365 to 363 and from 428 to 426 on two schedules, and Barnhart cites USD 50 million of operating cost savings alone for one United States airline (Barnhart 2006). The next step adds optional flights that the model may cancel, and new flights it may add, inside an itinerary-based model: integrated schedule design and fleet assignment (Lohatepanont and Barnhart 2004). Because removing a flight redistributes demand among the remaining ones, this model must handle demand that responds to supply; the lecture's rough estimate of its value for a large airline is USD 100 million to 350 million a year, sensitive to the quality of demand, spill and recapture data (historical) (Barnhart 2006). Chapter 13 develops schedule design itself.

Downstream: aircraft routing and maintenance. A feasible fleet assignment guarantees that each type's flights can be strung into rotations, but not that the rotations take every aircraft to a maintenance base in time. Routing models (Chapter 15) assign tails to those strings; when they fail, the fleet assignment needs maintenance constraints, such as minimum overnights at bases, or a joint model of fleeting and routing (Clarke et al. 1996; Barnhart, Boland, et al. 1998).

Downstream: crews. Pilots are qualified by type or family, so the fleet assignment decides which crew group flies each flight, and therefore the crew pairings that are possible. In the Laurentian bank, the exchange that gains CAD 1,831 a day moves four flights between

the two pilot groups; if that breaks efficient pairings, the crew cost can exceed the gain. This is why integrated crew and fleet models exist and why D³ needs common type ratings (Chapter 15).

For the planner, the practical consequence is that a timetable proposal should be checked against all three: does it fleet (enough aircraft of the right types, Result 14.2), does it route (maintenance), and does it crew. Delay propagation and robustness add a fourth test (Chapter 16).

IN PRACTICE **Fleet assignment in 2026**

At 30 June 2026 Air Canada's operating fleet counted 357 aircraft: 211 mainline aircraft (205 passenger aircraft of ten types and six 767 freighters), 51 at Air Canada Rouge and 95 operated by regional partners as Air Canada Express (Air Canada 2026e). Each subfleet with its own cabin, crew group or operator is a separate commodity in a fleet assignment model, and the commodities change every season as 787-10s and A321XLRs arrive and 737 MAX 8s move to Rouge (Air Canada 2026e). Commercial planning suites include fleet assignment optimizers; Lufthansa Systems, for example, describes a module that reassigns aircraft types to maximize profit using segment or origin-destination passenger forecasts, for strategic and tactical horizons, with rules for through flights, maintenance, ground times and allowed fleets (Lufthansa Systems 2025b). Its schedule optimization material adds that the module can run on revenue management forecasts or actual booking data and supports fleet utilization closer to the day of operations, the descendant of demand driven dispatch (Lufthansa Systems 2025d). Air Canada's use of the company's operations product, NetLine/Ops ++, is public (Lufthansa Systems 2025a); which planning modules it uses, and how they are calibrated, is not (Network Planning Handbook 2026). Open-source solvers such as HiGHS solve models like Example 14.4 in milliseconds (Huangfu and Hall 2018); the hard part is now the inputs: itinerary demand, recapture rates, and spill that reflects how revenue management sells the seats (Chapter 22).

14.10 Summary

- Fleet assignment chooses an aircraft type for every flight of a fixed timetable to maximize contribution, subject to aircraft flow balance and fleet sizes. It assigns types, not tails.

- The time-space network turns aircraft into flows: flight arcs end at ready times, ground and wrap-around arcs hold waiting aircraft, and the flow across any count line is the number of aircraft used.
- The model has cover, balance and count constraints; through flights, maintenance overnights, station and crew rules are linear side constraints that make it harder to solve.
- Assignment costs combine operating cost with expected spill net of recapture, $c_{ik} = \mathrm{OC}_{ik} + (1 - b_i) r_i \sigma_i L(z_{ik})$, using the formula of Chapter 9.
- The model solves fast because its linear relaxation is nearly integral; node consolidation shrinks it without changing its solutions, and islands shrink it further under a mild assumption; side constraints create fractional solutions that branch and bound must resolve.
- On the Laurentian bank the optimum beats the planner's best line-based plan by CAD 1,831 per day by letting two aircraft exchange their work at the hub.
- Leg-based spill misprices connecting passengers; no leg-by-leg cost can represent network spill. The passenger mix model, embedded in IFAM, values passengers by itinerary with capacity-consistent recapture, at a large computational cost.
- Deterministic network models ignore demand variability and overstate revenue; compare plans by simulation when both effects matter.
- Demand driven dispatch re-solves the assignment close to departure within crew-compatible families, gaining most when forecasts are good and demands independent.

Table 14.3 Key formulas of the chapter.

Quantity	Formula
Assignment cost	$c_{ik} = \mathrm{OC}_{ik} + (1 - b_i)\, r_i\, \sigma_i\, L((C_k - \mu_i)/\sigma_i)$
Cover	$\sum_k x_{ik} = 1$
Balance at node v	$y_{v^-} + \sum_{I(v)} x_{ik} = y_{v^+} + \sum_{O(v)} x_{ik}$
Fleet count	$\sum_{\mathcal{G}_k^0} y_g + \sum_{\mathcal{F}_k^0} x_{ik} \le N_k$
Aircraft needed	$n_k = \sum_s \max_t [D_s(t) - A_s(t)]^+ + \lvert\mathcal{F}_k^0\rvert$
Recapture from QSI	$b_{pr} = \mathrm{QSI}_r / (\mathrm{QSI}_r + m_c)$
Redirection cost	$f_p - b_{pr} f_r$ per passenger redirected from p to r

14.11 Further reading

Barnhart's guest lecture in the airline management course (Barnhart 2006) is the shortest complete account of the model, its class exercise and its extensions to schedule design; her schedule planning course (Barnhart 2003) adds the network flow background, the passenger mix model and problem set 2, the source of Examples 14.1 and 14.5. Abara (1989) describes an early implementation at American Airlines, and Hane et al. (1995) the techniques that made large instances solvable; Clarke et al. (1996) add maintenance and crew considerations. Barnhart, Kniker, and Lohatepanont (2002) introduce the itinerary-based model and Lohatepanont and Barnhart (2004) extend it to schedule design; Rexing et al. (2000) add time windows. Sherali, Bish, and Zhu (2006) survey fleet assignment models and algorithms. For demand driven dispatch, read the Boeing brochure (Boeing Commercial Airplane Group, n.d.) and Berge and Hopperstad (1993).

CHAPTER

15

Aircraft Routing and Crew Scheduling

IN THIS CHAPTER

- Where tail routing and crew scheduling sit after fleet assignment, and why airlines solve them in sequence.
- Maintenance routing as a choice of strings, and how an overnight pattern can make every routing infeasible.
- Duties, pairings, legality rules and the max-of-three pay function that turns a timetable into crew cost.
- Crew pairing as set partitioning with an astronomical number of columns, solved by column generation and branch-and-price.
- A Laurentian crew problem built and solved end to end with an integer programming solver.
- Rostering, the coupling of routing and crew through short connects, and Canada's flight crew fatigue rules.
- What a twenty-minute schedule change can break, and how to find out before it is published.

DRAWS ON Barnhart, Airline schedule planning: crew scheduling, aircraft maintenance routing, and crew pairing with maintenance routing (MIT 1.206J, 2003), with Problem Sets 2 and 3; handbook; Transport Canada, AC 700-047

Fleet assignment (Chapter 14) gives every flight an aircraft type. It does not say which aircraft will fly it, nor who will be in the cockpit. Two more planning steps turn a fleeted timetable into something an airline can operate. Aircraft routing chains flights into sequences for individual aircraft, so that every aircraft reaches a maintenance base often enough. Crew scheduling chains the same flights into legal sequences of work for pilots and flight attendants, at the lowest cost the work rules and collective agreements allow. Both steps belong to other teams, maintenance planning and crew planning, yet the network analyst meets them every week. In Barnhart's course

on airline schedule planning, crew cost is presented as the second-largest operating expense after fuel (Barnhart 2003), a 2003 ranking that moves with the price of fuel, and the handbook warns that a sequence that fits the aircraft turns may still create an expensive overnight, require another crew, or fail the duty and rest rules (Network Planning Handbook 2026).

This chapter builds both problems from first principles on one small, fully specified example: a day of 32 Laurentian A220 flights flown by four aircraft out of Montreal. We route the aircraft through maintenance, crew the flights, solve the crew problem with column generation, and then move one flight by twenty minutes to see what breaks. The methods are the ones that solve problems with thousands of flights; the small size lets you check every number. The next chapter (Chapter 16) takes the same day into operations, where weather and mechanical problems test the plan.

15.1 From fleet types to tails and crews

The planning sequence of Chapter 1 has four optimization steps after the network is chosen: schedule design, fleet assignment, aircraft routing and crew scheduling (Barnhart 2003). Each step takes the previous one's output as fixed. The order is not arbitrary. Fleet assignment decides how many seats each flight has, which drives revenue, and it needs only aggregate aircraft counts. Routing needs the fleet decision because aircraft of different types cannot replace one another. Crew scheduling needs both, because pilots are qualified on one fleet type or family, and because the routing decides where crews can stay with their aircraft. Table 15.1 summarizes the four problems that follow fleet assignment.

Table 15.1 The planning problems downstream of fleet assignment (adapted from Barnhart, 2003).

Problem	Decides	Horizon	Objective	Main constraints
Aircraft (maintenance) routing	which aircraft flies each flight	days to weeks	feasibility, through value	maintenance, turns, fleet size
Crew pairing	multi-day work sequences from a crew base	a season, by day of week	crew cost	duty, rest and pay rules
Rostering or bidlines	which person flies which pairings	a month	preferences, fairness	days off, monthly limits, training
Recovery	repairs to all of the above	hours	passenger and cost impact	everything above, in real time

The handbook's warning deserves to be read literally: fleet assignment is not tail assignment (Network Planning Handbook 2026). A feasible fleet assignment guarantees that some aircraft rotations exist for each type, because aircraft flows balance at every station; it does not guarantee that any of them brings each aircraft to a maintenance station in time (Barnhart 2006, 2003). The same timetable can pass the fleet test and fail the next two.

The running example Laurentian's A220 sub-fleet in this chapter flies a daily schedule of 32 flights between Montreal and five eastern stations that share Montreal's time zone: Quebec City (YQB), Ottawa (YOW), Toronto (YYZ), Boston (BOS) and New York LaGuardia (LGA). Table 15.2 shows the timetable as the schedulers built it, as four lines of flying, one per aircraft. Every night one aircraft sleeps at Quebec City, one at Boston and two at Montreal. The minimum aircraft turn is 35 minutes; the planned turns are 35 to 50 minutes.

Table 15.2 Laurentian's illustrative A220 day, as built (Montreal time). Each line starts and ends where its aircraft spends the night. Block times to or from Montreal: YQB 55 minutes, YOW 45, YYZ 1 h 15, BOS 1 h 25, LGA 1 h 30. The 32 flights total 37.5 block hours, 9.4 per aircraft.

Line	Starts at	Flights: number, destination, departure	Ends at
A	YQB	101 YUL 06:00, 102 YYZ 07:45, 103 YUL 09:40, 104 LGA 11:40, 105 YUL 13:50, 106 YYZ 16:00, 107 YUL 17:55, 108 YQB 20:00	YQB
B	BOS	201 YUL 06:15, 202 BOS 08:20, 203 YUL 10:25, 204 YQB 12:30, 205 YUL 14:05, 206 LGA 15:45, 207 YUL 17:55, 208 YOW 20:10, 209 YUL 21:30	YUL
W	YUL	301 LGA 07:00, 302 YUL 09:10, 303 YYZ 11:20, 304 YUL 13:15, 305 YOW 15:05, 306 YUL 16:25, 307 BOS 17:45	BOS
X	YUL	401 YOW 07:30, 402 YUL 08:50, 403 YQB 10:15, 404 YUL 11:50, 405 BOS 13:30, 406 YUL 15:35, 407 YYZ 17:50, 408 YUL 19:45	YUL

15.2 Maintenance routing

Aircraft must visit a maintenance base regularly. The lecture material describes the United States rules and practice of 2003: an A check was required every 60 flying hours, and airlines performed it every 40 to 45 flying hours with no more than three or four calendar days between checks; heavier checks came monthly (B), yearly for narrowbodies (C) or every 15 to 18 months for widebodies (Barnhart 2003). These numbers are historical. Today's intervals come from each operator's approved maintenance program, and the handbook's advice is to ask maintenance planning whether a rotation reaches the right station within the applicable interval, whether the ground window is long enough, and whether hangar capacity exists that night (Network Planning Handbook 2026). For a network planner the operational content is simple: a plane that sleeps at an outstation every night never gets its overnight check.

Why fleet assignment cannot see the problem The fleet assignment model of Chapter 14 can carry an approximate maintenance constraint: it adds maintenance arcs at the maintenance stations of each fleet and requires at least a minimum number of aircraft on them overnight. That guarantees enough maintenance opportunities in aggregate, but one aircraft could be serviced every night and another never (Barnhart 2003). For hub-and-spoke carriers the

approximation usually works, because the hub offers many chances to swap aircraft between rotations; for point-to-point and long-haul networks it often fails, because the flying between visits to a maintenance station is too long. An exact model has to follow individual aircraft.

> **DEFINITION 15.1 String**
>
> A **string** is a sequence of flights that an aircraft can fly in succession, beginning with a departure from a maintenance station and ending with an arrival at a maintenance station, followed by maintenance. Its start time is the first departure; its end time is the last arrival plus the maintenance time (Barnhart 2003; Barnhart, Boland, et al. 1998).

Because every string ends with maintenance, any routing assembled from strings is maintenance feasible by construction. The model only has to choose strings that cover each flight exactly once and that fit the fleet. Let S be the set of legal strings, F the set of flights and N the number of aircraft. Let $x_s = 1$ if string s is chosen, $a_{is} = 1$ if flight i belongs to string s, and c_s the cost of string s (negative for through value, the revenue earned when passengers can stay on the same aircraft). At each maintenance station, list the events where strings start or end in time order; v^- and v^+ denote the ground arcs entering and leaving event v, and $y_g \geq 0$ counts the aircraft waiting on ground arc g. The **count time** is a moment, such as 03:00, at which aircraft are counted: r_s is the number of times string s spans it and $p_g \in \{0,1\}$ whether ground arc g does.

$$
\begin{aligned}
\min \ & \sum_{s \in S} c_s x_s \\
\text{s.t.} \ & \sum_{s \in S} a_{is}\, x_s = 1 && \forall i \in F \\
& y_{v^-} + \sum_{s \text{ ends at } v} x_s = y_{v^+} + \sum_{s \text{ starts at } v} x_s && \forall v \\
& \sum_{s \in S} r_s x_s + \sum_{g} p_g\, y_g \leq N, \quad x_s \in \{0,1\},\ y_g \geq 0.
\end{aligned}
\tag{15.1}
$$

The first row covers every flight once. The second balances aircraft at each maintenance event, which is the only place where balance needs to be written; elsewhere it is implied because each string is a connected path. The third counts aircraft in use at the count time.

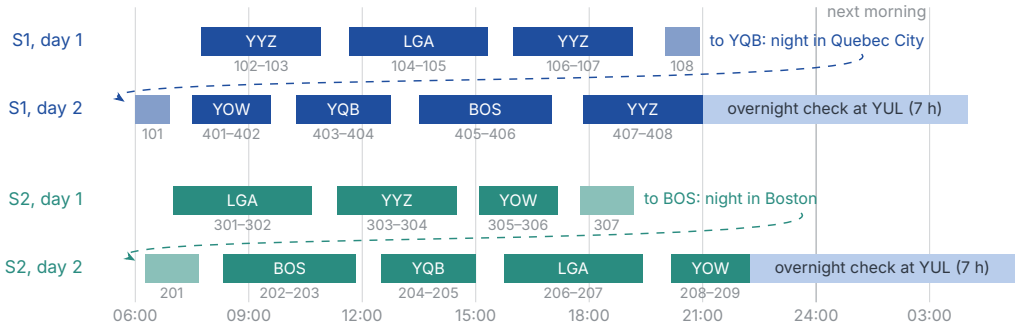


Figure 15.1 The two strings that route Laurentian's four A220s under an every-second-night maintenance rule. Each string covers two consecutive days and ends with an overnight check at Montreal; boxes are round trips from Montreal, labelled with the outstation, and the pale boxes are the one-way legs to and from an outstation night. String S1 needs the aircraft from Quebec City to take the 07:30 Ottawa turn (401) instead of the 07:45 Toronto flight it was built for.

Everything awkward, such as maintenance rules, nonlinear costs and through revenues, hides inside the definition of a legal string (Barnhart 2003). The price is the number of strings, which grows explosively with the network, so the model is solved by column generation, the method of Section 15.5.

Before any model, a counting argument tells the planner whether the overnight pattern can work at all.

RESULT 15.1 Base overnights limit maintenance frequency

Suppose a schedule repeats daily and leaves B aircraft overnight at the maintenance base every night, and suppose each of the N aircraft must spend a night at the base at least once in every k consecutive nights. A maintenance-feasible routing can exist only if

$$B \geq \frac{N}{k}.$$

Proof. Over T consecutive nights the schedule provides BT aircraft-nights at the base. Each aircraft needs at least $\lfloor T/k \rfloor$ of them, so $BT \geq N\lfloor T/k \rfloor \geq N(T/k - 1)$. Divide by T and let T grow: $B \geq N/k$. ∎

The condition is necessary, not sufficient: the aircraft must also be able to rotate through the base nights, which depends on the connections the timetable offers. That second test is what the string model performs.

WORKED EXAMPLE 15.1 Routing four A220s through Montreal

Take the schedule of Table 15.2 and two maintenance rules: every aircraft must sleep at Montreal at least every second night (the tight rule, $k = 2$), or at least every third night ($k = 3$). Both are illustrative stand-ins for a real maintenance program.

The lines as built fail. Line A starts and ends at Quebec City. If the aircraft that flies it today flies it tomorrow, it never sees Montreal at night. Lines of flying are a scheduling convenience, not a routing.

The counting test passes. With $N = 4$ and $B = 2$, Result 15.1 allows any $k \geq 2$. Under $k = 2$ the bound is tight: the two base nights must go to the two aircraft that slept away the night before, every night.

The swap points decide. An aircraft at Montreal can take any later departure at least 35 minutes after it arrives. Checking every arrival against every departure shows only three places where two aircraft can exchange their onward flights: the morning bank (the 06:55 arrival from Quebec City can continue on 401 at 07:30 or on 102 at 07:45), the 17:00 bank (arrivals at 17:00 and 17:10 feed departures at 17:45 and 17:50), and the 19:00 bank (arrivals at 19:10 and 19:25 feed departures at 20:00 and 20:10). Every other connection is forced. That gives $2 \times 2 \times 2 = 8$ ways to route the day. Following each aircraft over several days shows that one of the eight satisfies $k = 2$, four more satisfy only $k = 3$, and three, including the lines as built, leave some aircraft away from Montreal forever.

The string model finds the feasible routing. With $k = 2$ there are 11,817 legal strings of at most two days (175 one-day and 11,642 two-day strings), because an aircraft at Montreal can wait for many later departures. The integer program (15.1) picks two two-day strings, drawn in Figure 15.1. Each string spans the count time twice ($r_s = 2$: one outstation night, one check night), so together they use exactly four aircraft. The routing differs from the lines as built only at the morning bank: the aircraft arriving from Quebec City at 06:55 turns in 35 minutes onto the Ottawa round trip.

Size grows fast. Allowing three-day strings raises the count to 249,004, for 32 flights. Real fleets have hundreds of flights a day; this is why strings are generated, not enumerated.

The planner's lesson is that the overnight pattern, a network planning choice, sets the maintenance frequency the fleet can achieve, and that the swap opportunities at banks decide whether the aircraft can actually rotate through it.

Pricing strings Column generation needs, at each iteration, a string with negative reduced cost. That is a shortest path from a maintenance station to a maintenance station on the connection network, with arc lengths adjusted by the dual prices of the flights. A maximum elapsed-time rule is cheap to enforce, because elapsed time is fixed by the start and end nodes; a maximum flying-hours rule requires a resource label and dominance tests, which in the lecture's long-haul case study multiplied run times by roughly 2 to 7 (Barnhart 2003). The same machinery appears in more detail for crews below.

15.3 Crew pairings and what they cost

Crews are scheduled in layers. Flights are strung into duties, duties into pairings, pairings into monthly lines for individual crew members (Figure 15.2).

DEFINITION 15.2 Duty period, pairing and time away from base

A **duty period** (duty) is a sequence of consecutive flights, roughly a working day, that one crew can operate before it must rest. It runs from **report** (before the first departure) to **release** (after the last arrival); its **block time** is the time from gate departure to gate arrival summed over its flights, and the gaps between flights are **sits**. A **pairing** is a sequence of duties separated by rest periods that starts and ends at the crew's base (domicile); a rest away from base is a **layover**. The **time away from base** (TAFB) of a pairing runs from its first report to its last release (Barnhart 2003).

Legality rules come from regulation and from collective agreements. The lecture groups them by level (Barnhart 2003). Duty rules: flights must chain in space and time; flying time and duty time have maxima; sits have a minimum (time to change aircraft) and a maximum. Pairing rules: the first duty leaves and the last returns to the base; rest between duties has a minimum, layovers

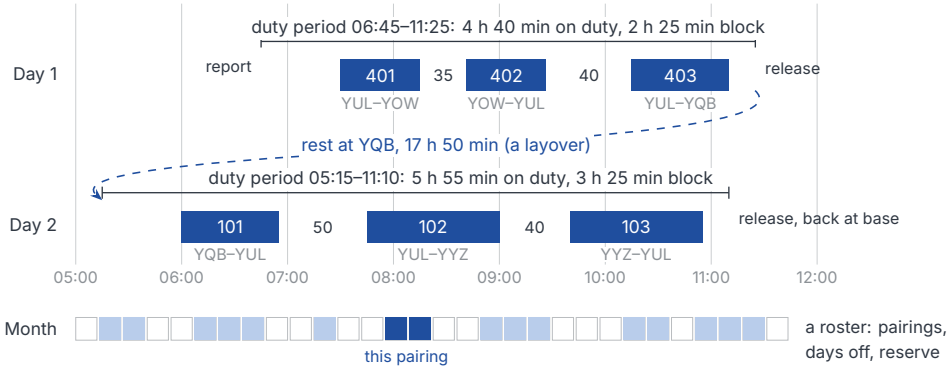


Figure 15.2 A two-day pairing from the Laurentian example. Day 1 is a 4 h 40 min duty with 2 h 25 min of block time ending in Quebec City; the crew rests 17 h 50 min there and flies a 5 h 55 min duty home the next morning. Numbers between flights are sit times in minutes. A roster strings pairings like this one, days off and reserve days into a month.

a maximum; pairings have a maximum number of days. Cockpit crews stay together as a unit, have a fixed composition and earn more; cabin crews are scheduled as individuals, their number can vary with demand and they can work a wider range of aircraft types. Domestic hub networks are mostly daily, with many connections; international networks are often weekly and need more deadheads, crews travelling as passengers to reposition. Table 15.3 lists the rules of the Laurentian example. They are adapted from the duty rules of the course's third problem set (Barnhart 2003) and simplified; they are not Canada's regulations, which Section 15.9 discusses.

Table 15.3 Illustrative crew rules for the Laurentian A220 example. The problem set's rules were 8 h of flying, 14 h of duty, sits of 30 to 90 minutes, rests of 8 to 18 hours and pairings of at most 4 days.

Rule	Laurentian example
Report before first departure; release after last arrival	45 minutes; 15 minutes
Maximum duty, report to release	13 h up to four flights, 30 min less per extra flight
Maximum block time in a duty	8 h
Sit between flights	at least 30 min on the same aircraft, at least 50 min when changing aircraft, at most 90 min
Rest between duties (layover)	10 h to 18 h
Duties per pairing	at most 3

The pay function Crews are not paid for block time alone. The lecture's cost function pays the largest of several measures at two levels (Barnhart 2003):

$$\begin{aligned} b_d &= \max\{\,\mathrm{BLK}_d,\, f_d\,\mathrm{DUT}_d,\, g_d\,\}, \\ \gamma_p &= \max\Big\{\sum_{d\in p} b_d,\, f_p\,\mathrm{TAFB}_p,\, g_p\Big\}, \\ c_p &= w\,\gamma_p + h\,n_p. \end{aligned} \tag{15.2}$$

Here BLK_d and DUT_d are the block time and duty time of duty d, f_d and f_p are contractual fractions below one, g_d and g_p are minimum guarantees, b_d is the **credit** (pay hours) of the duty and γ_p the credit of the pairing. The lecture also mentions a variant in which the third pairing term is a fraction of the number of duties. For the Laurentian example we price credit at $w = \mathrm{CAD}\ 400$ per hour for a two-pilot crew and add $h = \mathrm{CAD}\ 300$ per layover night n_p for hotel and allowances; we set $f_d = 1/2$ (an hour of pay for every two on duty), $f_p = 2/7$ (an hour for every 3.5 away from base), $g_d = 4$ h and no pairing guarantee. All are illustrative.

Each term of (15.2) compensates for a different kind of unproductive time. The duty guarantee pays for short days, the duty-time fraction for long sits, and the TAFB fraction for long layovers. The ratio of credit to block time is a quick measure of how crew-efficient a schedule is: a ratio of 1.00 means the crews are paid only for flying.

WORKED EXAMPLE 15.2 Pricing one pairing

Price the pairing of Figure 15.2. Day 1 flies 401 YUL–YOW at 07:30, 402 back at 08:50 and 403 to Quebec City at 10:15, arriving 11:10. Day 2 flies 101 YQB–YUL at 06:00, 102 to Toronto at 07:45 and 103 back at 09:40, arriving 10:55.

Duty 1. Report 06:45, release 11:25: 4.67 h on duty, 2.42 h of block (2 h 25 min). Credit $b_1 = \max\{2.42,\ 0.5 \times 4.67,\ 4.00\} = 4.00$ h: the guarantee binds.

Rest. From 11:25 to the next morning's 05:15 report: 17 h 50 min, inside the 10 to 18 hour window.

Duty 2. Report 05:15, release 11:10: 5.92 h on duty, 3.42 h of block. Credit $b_2 = \max\{3.42,\ 0.5 \times 5.92,\ 4.00\} = 4.00$ h: the guarantee binds again.

Pairing. TAFB runs from 06:45 on day 1 to 11:10 on day 2, 28 h 25 min, so the trip term is $\frac{2}{7} \times 28.42 = 8.12$ h. Credit $\gamma = \max\{8.00,\ 8.12\} = 8.12$ h (8 h 07 min), against 5 h 50 min of block: a pay-to-block ratio of 1.39. Cost: CAD $400 \times 8.119 + 300 = 3{,}548$.

The pairing is legal but expensive. Its crew sits in a Quebec City hotel from late morning to dawn because the 06:00 departure must be crewed by someone who is already there and rested. Schedules that put a midday arrival and an early departure at the same outstation buy that layover every day.

15.4 Crew pairing as set partitioning

The crew pairing problem chooses a minimum-cost set of pairings such that every flight belongs to exactly one of them (Barnhart 2003). With P the set of legal pairings, $a_{fp} = 1$ if flight f is in pairing p, and $x_p = 1$ if pairing p is chosen:

$$\min \sum_{p \in P} c_p x_p \quad \text{s.t.} \quad \sum_{p \in P} a_{fp}\, x_p = 1 \ \ \forall f \in F, \qquad x_p \in \{0,1\} \ \ \forall p \in P. \tag{15.3}$$

This is a **set partitioning** problem. If crews may deadhead, the equality is relaxed to "at least one" with a cost for each extra crew on board, a set covering model. The model is solved for one fleet type or fleet family at a time, and in practice decomposed by period: a daily problem covers every flight that operates at least four times a week,

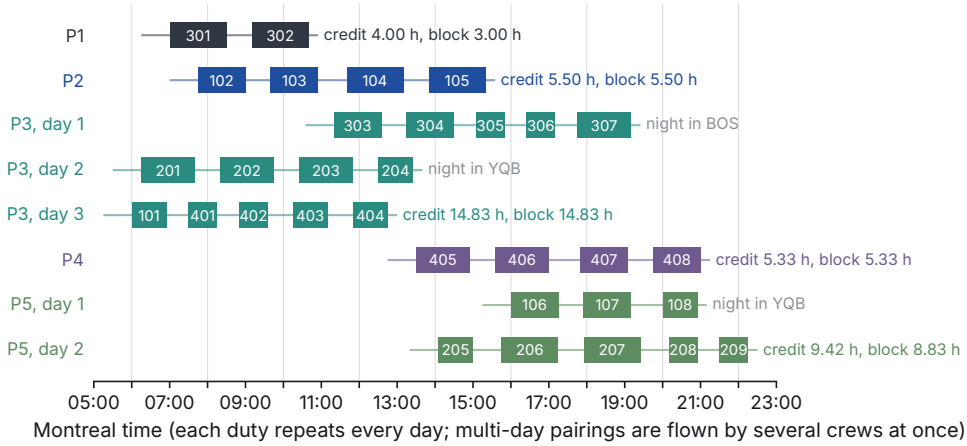


Figure 15.3 The optimal crew plan for the Laurentian A220 day under the routing of Figure 15.1: eight duties forming five pairings. Thin bars show duty periods from report to release; boxes are flights. Multi-day pairings are flown every day by several crews at once, so the plan keeps eight crews at work each day.

and its pairings are repeated every day, so a three-day pairing keeps three crews busy at once; exceptions (flights operating at most three times a week, and the flights of daily pairings that break on days when one of their flights does not operate) and transitions across the month boundary are solved afterwards (Barnhart 2003).

The model looks easy: a linear objective, no complicated constraints and one row per flight. Everything difficult, the rules and the nonlinear pay, has been moved into the definition of a column. The difficulty is the number of columns. The course's third problem set states that a typical crew pairing problem has hundreds of millions of legal pairings, and the lecture cites a duty network with 363 flights, 7,838 duties and 1.65 million duty-to-duty connections (Barnhart 2003). Even the Laurentian day shows the explosion. Under the rules of Table 15.3, with the short connects of the routing of Example 15.1, its 32 flights form 193 legal duties and 652 legal pairings; allowing four-day pairings raises the count to 1,189; dropping the 90-minute cap on sits gives 703 duties and 44,778 pairings. Rules do not only protect crews; they also keep the problem finite.

WORKED EXAMPLE 15.3 Crewing the Laurentian day

Crews connect under the rules of Table 15.3, and the routing of Example 15.1 decides which tight connections (30 to 49 minutes) a crew may make by staying with its aircraft. Enumerating every legal pairing gives 652 columns: 38 one-day, 174 two-day and 440 three-day pairings. Pricing each with (15.2) and solving (15.3) with an integer programming solver (HiGHS through `scipy.optimize.milp`) takes a fraction of a second.

The optimum, drawn in Figure 15.3, uses five pairings and costs CAD 16,533 per day:

- P1, 301–302, a 3 h LaGuardia round trip paid at the 4 h guarantee (CAD 1,600);
- P2, 102 to 105, Toronto and LaGuardia round trips, 5.50 h of block and credit (CAD 2,200);
- P3, a three-day pairing: 303 to 307 ending in Boston, 201 to 204 ending in Quebec City, 101 and 401 to 404 back to Montreal, 14.83 h of credit for 14.83 h of block (CAD 6,533 with two hotel nights);
- P4, 405 to 408, Boston and Toronto round trips, 5.33 h (CAD 2,133);
- P5, a two-day pairing: 106 to 108 ending in Quebec City (block 3.42 h, paid at the 4 h guarantee), then 205 to 209 (CAD 4,067 with one hotel night).

Credit totals 39.08 h for 37.50 h of block, a ratio of 1.042; the excess is the two guarantees. With three hotel nights, the cost is CAD $400 \times 39.08 + 900 = 16{,}533$. The linear programming relaxation has the same value, so here the relaxation is integral.

Read the plan as a planner. The 06:00 departure from Quebec City and the 06:15 from Boston must be flown by crews who slept there, which forces layovers. The evening arrival at Quebec City (108, in at 20:55) cannot feed the 06:00 departure, because the crew would rest only 8 h 05 min; that crew flies 205 the next afternoon instead, and 101 is flown by the crew that arrived on 204 at 13:25. The Boston night works only because 307 arrives at 19:10, leaving a rest of 10 h 05 min before 201: five minutes to spare, a fragility Section 15.10 returns to.

15.5 Column generation and the pricing problem

Enumeration works for 32 flights and fails for 300. **Column generation** solves the linear relaxation of (15.3) while looking at only a small fraction of its columns (Barnhart 2003; Desrosiers and Lübbecke 2005). Start with a **restricted master problem** (RMP): the relaxation restricted to a subset of columns, made feasible if necessary with artificial columns of high cost. Solve it and read the dual price π_f of each flight's covering row. A pairing p outside the RMP would improve the solution only if its **reduced cost**

$$\bar{c}_p = c_p - \sum_{f \in F} a_{fp}\, \pi_f \tag{15.4}$$

is negative, so a **pricing problem** searches for the most negative one. If it finds some, they join the RMP and the loop repeats. The duals have a useful reading: π_f is the crew cost that covering flight f adds at the margin, and a pairing is worth adding when its cost is below the value of the flights it covers. Two facts make the loop rigorous. The LP duality they rest on is reviewed in Appendix B.

RESULT 15.2 Optimality test and bound for column generation

Let z_{RMP} be the optimal value of the restricted master problem with duals π, and z_{LP} the optimal value of the full linear relaxation of (15.3).

1. If $\min_{p \in P} \bar{c}_p \geq 0$ and the RMP solution uses no artificial column, it is optimal for the full relaxation: $z_{\text{RMP}} = z_{\text{LP}}$.
2. If every pairing covers at least two flights, then

$$z_{\text{LP}} \;\geq\; z_{\text{RMP}} + \frac{|F|}{2}\, \min\Big\{0,\, \min_{p \in P} \bar{c}_p\Big\}.$$

Proof. For any feasible solution x of the full relaxation, substitute $c_p = \bar{c}_p + \sum_f a_{fp}\pi_f$:

$$\begin{aligned} \sum_p c_p x_p &= \sum_p \bar{c}_p x_p + \sum_f \pi_f \sum_p a_{fp} x_p \\ &= \sum_p \bar{c}_p x_p + \sum_f \pi_f = \sum_p \bar{c}_p x_p + z_{\text{RMP}}, \end{aligned}$$

using the covering rows and, in the last step, LP duality for the RMP, whose optimal value equals $\sum_f \pi_f$. If all reduced costs are nonnegative the first sum is nonnegative, so no solution beats z_{RMP}, which

the RMP solution attains; this is part 1. For part 2, sum the covering rows: $\sum_p (\sum_f a_{fp}) x_p = |F|$, and since each column has at least two ones, $\sum_p x_p \le |F|/2$. Hence $\sum_p \bar{c}_p x_p \ge (|F|/2)\min\{0, \min_p \bar{c}_p\}$. Apply this to an optimal x. ∎

Part 2 is a **Lagrangian bound**: at every iteration the pricing step, which must be solved anyway, also yields a lower bound on the optimum (Lübbecke and Desrosiers 2005). In a pairing problem every column covers at least two flights because a crew must leave its base and come back.

Pricing as a resource-constrained shortest path The pricing problem never lists pairings. It searches a network in which nodes are flights and arcs are legal connections, with one copy of the flight network per day of the pairing, rest arcs linking a late arrival to next-day departures at the same station, and a source and a sink at the crew base (Figure 15.4). A pairing is a source-to-sink path, and in a daily problem a path may not repeat a flight (Barnhart 2003).

The path's cost is not a sum of arc costs, because of the maxima in (15.2). The search therefore carries a **label** at each node, a vector of the quantities needed to finish the computation and check legality later: the block time and start time of the current duty, the credits of completed duties, the pairing's start (for TAFB), the number of duties, the flights already used and the sum of their duals. When a path reaches the sink, the label gives the pairing's credit and its reduced cost (15.4). A label is extended along every legal arc in time order, and a label is discarded when another label at the same node is at least as good in every component: no more block, no earlier duty start, no earlier pairing start, no more duties or completed credit, no more flights used, and at least as much dual value collected. Because each final cost is nondecreasing in these components, a dominated label can never lead to a cheaper pairing. This is the **resource-constrained shortest path problem**, the workhorse of crew and vehicle scheduling (Irnich and Desaulniers 2005).

Two network designs are common. In a flight network each arc is one connection, typically no more than about 30 per flight, but labels must enforce the duty rules; in a duty network each node is a whole legal duty, so duty rules are built in, but arcs explode, as the 1.65 million connections above show (Barnhart 2003). The third problem set asks which design suits which network; dense domestic hub networks with short flights favour the flight network, while long-haul networks with one or two flights per duty and few duties make the duty network attractive.

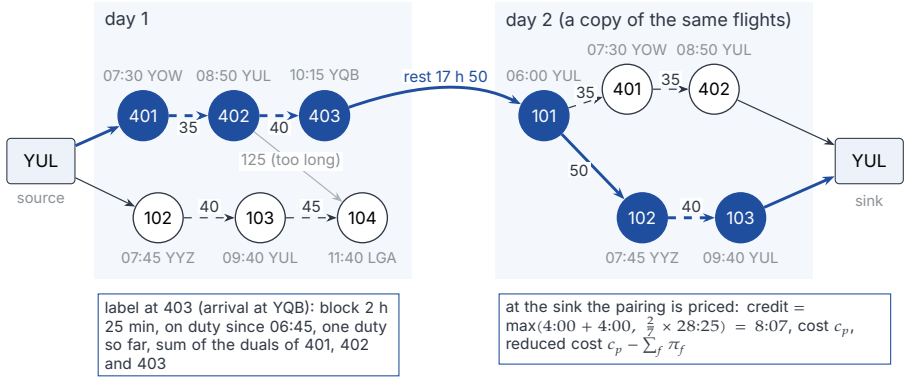


Figure 15.4 The pricing network for one Laurentian pairing. Flights are nodes; arcs are legal sits (minutes shown) within a day and rest arcs from day 1 to the copy of the flight network for day 2. Dashed arcs are short connects, legal only if the aircraft continues too; the grey arc is an illegal 125-minute sit. The filled path is the pairing of Example 15.2; its label collects block, duty time and duals until the sink, where the maxima of (15.2) are evaluated.

WORKED EXAMPLE 15.4 Column generation on the Laurentian day

Start the RMP with 32 artificial columns, one per flight, each costing CAD 5,000, so the first RMP costs CAD 160,000. At every iteration the labelling algorithm prices all legal pairings implicitly, and the five most negative are added. A check against brute-force enumeration confirms that the labelling finds the same minimum reduced cost.

At iteration 10 the RMP is worth CAD 17,083 and its solution is fractional, with six pairings at value one half. The pricing step then finds pairing 401 to 405 followed the next day by 201 to 203: its cost is CAD 3,900, the duals of its eight flights sum to CAD 5,417, so its reduced cost is $3{,}900 - 5{,}417 = -1{,}517$. It enters.

Figure 15.5 tracks the iterations. The RMP reaches the optimum, CAD 16,533, at iteration 19, but the pricing step still finds negative columns, so two more iterations are needed to prove optimality; at iteration 21 the minimum reduced cost is zero and the Lagrangian bound of Result 15.2 meets the RMP value. Column generation looked at 96 of the 652 legal pairings, 15 percent of them. On real networks the share is far smaller.

The final duals are worth a look. One optimal set puts the highest values on 307 to Boston (CAD 2,614), 108 to Quebec City (CAD 1,948) and 208, 301 and 408 (CAD 1,600 each): the

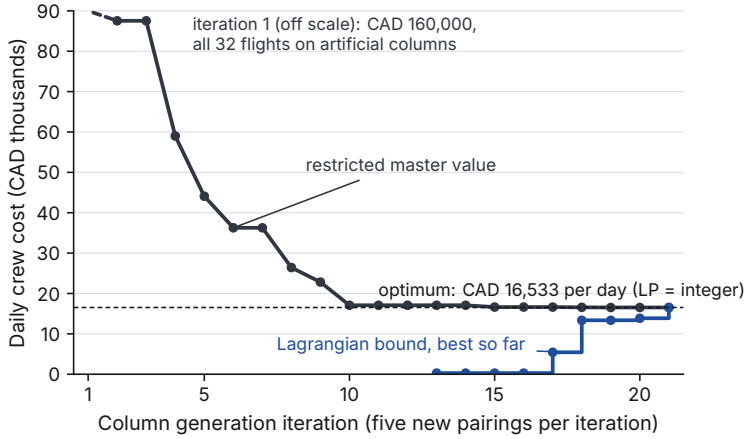


Figure 15.5 Column generation on the Laurentian crew problem. The restricted master value falls quickly and then flattens (tailing off) while the Lagrangian bound, which stays negative (and off the chart) until reduced costs become small, rises to meet it at the optimum.

evening outstation departures and the edges of the day are the crew-expensive flights. Duals of degenerate problems are not unique, so read them as a ranking, not a price list.

The shape in Figure 15.5 is typical: rapid early progress and a long tail in which the RMP barely moves while the method proves optimality. Practitioners stop early when the bound gap is small enough, which Result 15.2 makes safe to do.

15.6 Branch-and-price

Column generation solves the linear relaxation. To get integer solutions the relaxation must be embedded in branch-and-bound, with column generation at every node: **branch-and-price** (Barnhart, Johnson, et al. 1998). The difficulty is the branching rule. Fixing a fractional column to zero is easy to state but hard to enforce, because the pricing problem will regenerate the forbidden pairing unless it is modified; the resulting tree is also unbalanced, since the zero branch barely changes the problem (Barnhart 2003). A rule is acceptable only if it can be imposed while keeping the pricing problem a shortest path problem of the same kind.

Follow-on branching satisfies that requirement. For flights f and g, let ϕ_{fg} be the total value of the pairings in which g immediately follows f.

RESULT 15.3 Fractional solutions have fractional follow-ons

If a feasible solution x of the linear relaxation of (15.3) is fractional, there are flights f and g with $0 < \phi_{fg} < 1$.

Proof. Suppose every ϕ_{fg} is 0 or 1 and take a flight f. The pairings with $x_p > 0$ that contain f have total value 1. If $\phi_{fg} = 1$ for some g, all of them continue from f to g; if $\phi_{fg} = 0$ for every g, all of them end at f. The same argument applies to the flight before f. Following these forced successors and predecessors, every positive pairing that contains f contains the same sequence of flights, with the same breaks between duties, since a sit and a rest cannot both be legal for the same connection. So a single pairing contains f and its value is 1. As f was arbitrary, x is integral. ∎

The two branches are "g must follow f" and "g must not follow f". Both are imposed by deleting arcs in the pricing network: to force the follow-on, delete every arc out of f except the one to g and every arc into g except the one from f; to forbid it, delete the arc from f to g. The pricing algorithm is unchanged. The rule is a specialization of the branching rule of Ryan and Foster (1981) to consecutive flights, and the lecture recommends branching on the follow-on with the largest fractional value (Barnhart 2003). The same rule, applied to strings, solves the routing model (15.1).

The fractional RMP at iteration 10 of Example 15.4 shows the pattern. Flight 101 is followed by 102 in a pairing at value one half and by 401 in another pairing at value one half, so $\phi_{101,102} = \phi_{101,401} = 1/2$. Had this been the final relaxation at a node, the tree would branch on whether 401 follows 101: in the Laurentian routing that is the morning aircraft swap, so the crew branch would decide whether the crew stays with its aircraft.

A cheaper alternative is **price-and-branch**: generate columns at the root only, then solve the integer program over those columns. It may find no feasible integer solution at all, and folklore has it that a good integer solution needs some columns that look bad in the relaxation (Barnhart 2003). The method that proved itself in practice is full branch-and-price: in the lecture's historical example, American Airlines' crew pairing optimization covered more than 25,000 crew members in 1993, saved more than USD 20 million a year and solved

in 4 to 10 hours (Barnhart 2003). Those numbers are three decades old; they still explain why airlines invest in this software.

15.7 Rostering and bidlines

Pairings are built for anonymous crews. The next step assigns them to people for a month, together with days off, reserve days, vacations, training and medical leave. The lecture distinguishes two traditions (Barnhart 2003). In the **bidline** approach, pairings are packed into generic monthly lines that are posted, and crew members bid for them in order of seniority; the lecture describes it as common in the United States. In **rostering**, personalized lines are built directly, honouring requests and training needs with priority by seniority; the lecture describes it as typical outside the United States. Both are again set partitioning models, now sequencing pairings into monthly schedules under rules such as minimum rest between pairings, maximum monthly flying and duty time and minimum days off. The objective changes from cost to preferences and fairness, and the unit changes from a complete crew to an individual. The lecture's geographic split is twenty years old; which method applies to each crew group is a question for crew planning.

For the network planner, rostering matters through its constraints: cumulative flight time limits over 28-day and 365-day windows cap how much flying crews can absorb in a peak month, reserve crews are the buffer for disruptions (Chapter 16), and short rests between pairings can make lines unattractive or illegal.

15.8 Routing and crews together

Routing and crew pairing touch at one point. A crew moving between aircraft needs time to walk through the terminal; a crew that stays with its aircraft only needs the aircraft's turn.

DEFINITION 15.3 **Short connect**

A **short connect** is a connection between two flights that is legal for a crew only if the same aircraft flies both flights, because it is shorter than the minimum connection time for crews changing aircraft (Barnhart 2003). In the Laurentian rules, connections of 30 to 49 minutes are short connects.

In the 2003 practice the lecture describes, routing came first and crews used whatever short connects it happened to contain (Barnhart 2003); reversing the order, crews first with every short connect allowed, can leave routing infeasible. Cordeau, Stojković, Soumis and Desrosiers (2001) integrate the two string models with linking constraints and solve them by Benders decomposition. Cohn and Barnhart (2003) observe that only the short connects matter to the crew problem and propose the **extended crew pairing** model: add to the crew problem a choice among complete maintenance-feasible routings r, and forbid any pairing that uses a short connect the chosen routing does not provide. With T the set of short connects, $\alpha_{tr} = 1$ if routing r contains short connect t, $\beta_{tp} = 1$ if pairing p uses it, and $y_r = 1$ if routing r is chosen:

$$
\begin{aligned}
\min \ & \sum_{p} c_p x_p \\
\text{s.t.} \ & \sum_{p} a_{fp}\, x_p = 1 && \forall f \in F \\
& \sum_{r} \alpha_{tr}\, y_r \ \geq \ \sum_{p} \beta_{tp}\, x_p && \forall t \in T \\
& \sum_{r} y_r = 1, \qquad x_p,\ y_r \in \{0,1\}.
\end{aligned}
\tag{15.5}
$$

The model looks larger than the crew problem, but most routings are redundant. Two routings with the same set of short connects are interchangeable for the crews, and a routing whose short connects are a subset of another feasible routing's set is never needed. In the lecture's examples, more than 8,700 maintenance solutions of a 41-flight instance collapsed to one column and more than 25,000 solutions of a 61-flight instance to four; on two airline instances the method closed the gap to its lower bound completely and to within 1.7 percent (Barnhart 2003). The routing variables can even be relaxed to be continuous once the pairings are integral, so the model has no more binary variables than the crew problem alone (Barnhart 2003; Cohn and Barnhart 2003).

Table 15.4 Crew cost of the Laurentian day under each of the eight routings. "Unbounded" means some aircraft never returns to Montreal for the night.

Routing (morning, 17:00, 19:00 bank)	Most nights away from YUL in a row	Crew cost (CAD per day)
as built: 101→102, 406→407, 107→108	unbounded	16,633
as built, 107→208 at 19:00	2	16,467
406→307 at 17:00	unbounded	16,400
406→307 and 107→208	2	16,233
101→401 in the morning	1	16,533
101→401 and 107→208	2	16,367
101→401 and 406→307	2	16,181
101→401, 406→307 and 107→208	unbounded	16,014
every short connect allowed (a bound)		16,014

WORKED EXAMPLE 15.5 Which routing should the crews get?

In the Laurentian day, 23 short connects are fixed by every routing (they are forced turns at outstations and at Montreal), and six depend on the swaps: 101→401 at the morning bank, 306→307, 306→407 and 406→307 at the 17:00 bank, and 207→108 and 207→208 at the 19:00 bank. Each of the eight routings provides a different set, so each is a distinct column of the extended model (15.5). Table 15.4 solves the crew problem for each.

Under the every-second-night rule only one routing is feasible, and crews cost CAD 16,533 a day. Under the every-third-night rule five routings qualify, at crew costs from CAD 16,181 to CAD 16,533 a day. The most expensive of the five is the routing the tighter rule forces, so choosing the routing with crews in mind is worth up to CAD 352 a day; even among the four routings that only the looser rule admits, the spread is CAD 286 a day. The cheapest crew plan overall, CAD 16,014, needs

a routing that strands aircraft away from Montreal, so it is unusable.

The constraint-generation approach of the third problem set reaches the same answers from the other side. Solve the crew problem with every short connect allowed, check whether some maintenance-feasible routing provides the short connects it used, and if not, add a cut forbidding a minimal infeasible subset of them (Barnhart 2003). Under the every-third-night rule the first solution (CAD 16,014) uses 101→401, 207→108 and 406→307 together, which no feasible routing provides; one cut later the solver returns CAD 16,181, which is optimal. Under the every-second-night rule two cuts lead to CAD 16,533.

Annualized over 365 days, the CAD 352 daily gap between the best and worst every-third-night routing is about CAD 129,000 for four aircraft: small against the airline, large against the effort of asking the routing question with crews in mind.

The lesson extends beyond crews. When two planning problems interact through a narrow interface, here a few dozen short connects, the efficient integration offers one problem a menu of the other's interface patterns instead of merging the two full models.

15.9 Fatigue rules in Canada

The rules in Table 15.3 are a teaching device. In Canada, flight crew hours of work and rest are governed by Division III of Part VII of the Canadian Aviation Regulations, published in 2018 and amended in 2022, which apply to commuter, air taxi and airline operators. Transport Canada's Advisory Circular AC 700-047 explains their intent and one acceptable means of compliance (Transport Canada 2024). At a conceptual level the framework has six parts (Transport Canada 2024):

- **Flight duty period limits** read from tables that depend on the time the duty starts at the place where the crew is acclimatized, the number of flights and their average duration, so an early start or many short sectors shorten the allowed duty.
- **Cumulative flight time limits** over rolling windows, for example 112 hours in any 28 consecutive days and 1,000 hours in any 365.
- **Cumulative limits on hours of work**, which count training, positioning and other duties as well as flying.

- **Minimum rest periods**, for instance 10 hours in suitable accommodation after a duty that ends away from home base, with longer rest after crossing time zones or between late-night and early-morning duties.
- **Time free from duty** over weekly and monthly windows.
- **Unforeseen operational circumstances**, which permit limited extensions on the day of operation; a schedule whose maximum duty is exceeded this way on more than 10 percent of occasions in 90 days must be re-planned.

Two consequences matter for planning. First, legality depends on the clock time and sector count of each duty, not on a single maximum, which is why the handbook advises network planners to obtain a crew-feasibility and cost review rather than apply one remembered limit to every trip (Network Planning Handbook 2026). Second, collective agreements add their own rules and pay terms on top of the regulation, and the pay function (15.2) is where they enter the cost.

IN PRACTICE **Crew and maintenance as commercial constraints in 2026**

The handbook treats crew and maintenance as commercial constraints: a small schedule change can create a discrete extra crew or hotel requirement, and scheduled capacity and approved maintenance availability are different inputs (Network Planning Handbook 2026). Air Canada's financial reporting shows the cost lines involved. Its second-quarter 2026 MD&A reports "crew cycle" as a separate line of other operating expenses, CAD 98 million in the quarter against CAD 86 million a year earlier, an increase of 14 percent, and records charges related to collective agreements, including the pilots' agreement with ALPA ratified in 2024 and pension amendments agreed with CUPE in 2025 (Air Canada 2026e). A network analyst does not build pairings, but should ask crew planning three questions about any proposal: which pairings and layovers it changes, which rule binds, and what the change does to credit per block hour.

15.10 What breaks when a planner moves a flight

The handbook closes its discussion of crew and maintenance constraints with a warning: a 20-minute departure change may alter a connection, a duty period and a maintenance window simulta-

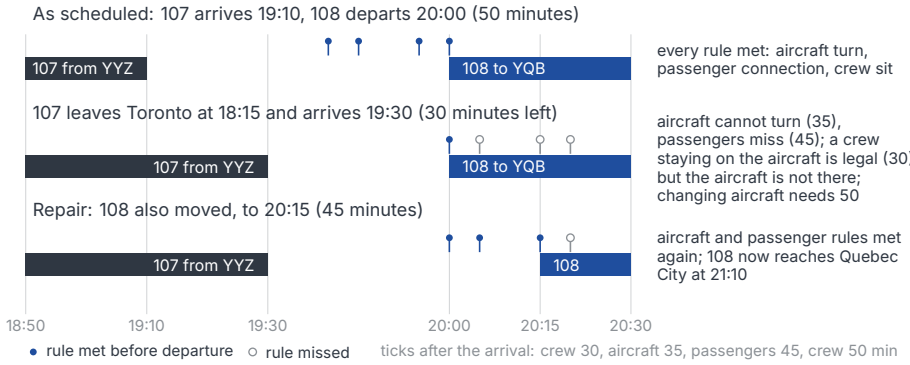


Figure 15.6 What a twenty-minute change does at Montreal's 19:00 bank. Ticks mark the earliest time each rule allows the onward departure after flight 107 arrives: a crew staying with its aircraft (30 minutes), the aircraft turn (35), a passenger connection (45) and a crew changing aircraft (50). Moving 107 breaks three of the four; moving 108 fifteen minutes later restores the aircraft turn and the passenger connection, so the crew can again stay with its aircraft.

neously (Network Planning Handbook 2026). The Laurentian day shows each of these failures.

WORKED EXAMPLE 15.6 Twenty minutes later from Toronto

The commercial team asks to move 107, the Toronto–Montreal evening flight, from 17:55 to 18:15 so that Toronto business travellers can finish a later meeting. It now arrives at 19:30 instead of 19:10. Check the four layers in turn (Figure 15.6).

Passenger connections. At an illustrative minimum connecting time of 45 minutes, the Toronto–Quebec City connection onto 108 at 20:00 shrinks from 50 to 30 minutes and is lost; so is the connection onto 208 to Ottawa, from 60 to 40 minutes.

Aircraft. The aircraft from 107 can no longer turn onto 108 (30 minutes against a 35-minute minimum), so at the 19:00 bank only one assignment remains: 107's aircraft takes 208 and 207's aircraft takes 108.

Maintenance. The only routing that satisfied the every-second-night rule (Table 15.4) used the 107→108 turn. After the change, no routing does: the aircraft that sleeps in Boston would sleep in Quebec City the next night. Under the every-third-night rule three routings survive.

Crews. The plan of Example 15.3 had a crew fly 106, 107 and 108 on one aircraft. With 30 minutes between 107 and 108 the

crew could stay on board, but the aircraft does not continue, and changing aircraft needs 50 minutes. The duty breaks and the crew plan must be rebuilt for whichever routing survives.

Repair. Moving 108 fifteen minutes later, to 20:15, restores the aircraft turn (45 minutes), the Quebec City passenger connection (exactly 45 minutes; the Ottawa connection on 208 stays lost at 40) and the every-second-night routing; the crew plan and its cost, CAD 16,533 a day, come back unchanged, and the crew that brings 108 in still rests 15 h 55 min in Quebec City before flying 205 the next afternoon. The price of the repair is commercial: 108 now reaches Quebec City at 21:10. It also opens a LaGuardia–Quebec City connection (207 to 108, 50 minutes), which Chapter 16 examines for robustness.

A variant that cannot be repaired so easily. Had the planner instead moved 307, the evening Boston departure, 20 minutes later, nothing would break for aircraft or passengers. But the crew that arrives in Boston would rest only 9 h 45 min before the 06:15 departure, below the 10-hour minimum. The only other crew that could fly 201, the one arriving on 405 at 14:55, is needed for the 15:35 return, and no legal crew plan exists under any of the eight routings. The fix would be another retiming, an extra crew positioned to Boston, or an extra hotel night, each with a cost.

The example generalizes into a short checklist for any schedule change. List every connection the flight is part of for passengers, aircraft and crews, against the applicable minimum for each; trace the aircraft rotation to the next maintenance night; ask crew planning which duties, rests and layovers the change touches; and look for the second change that repairs the first. The network planner is not expected to solve these problems, but to recognize them early and to price them with the teams that do (Network Planning Handbook 2026).

> **PITFALL**
>
> Treating legality as a property of a flight. Whether a flight can be crewed, routed or connected depends on its neighbours: the turn before it, the rest after it, the bank around it. A change that is legal for the flight in isolation can make the whole day infeasible, as the 20-minute move of 307 does.

Routing and crew problems turn a timetable into tails and people; their cost and feasibility depend on the shape of the schedule, its overnight pattern, banks and edges, as much as on its block hours.

15.11 Summary

- Aircraft routing and crew scheduling follow fleet assignment. A feasible fleet assignment guarantees aircraft rotations exist, not that any of them is maintenance feasible.
- The string model makes maintenance feasibility part of each column. A counting bound ($B \geq N/k$ base overnights) tells the planner whether an overnight pattern can support a maintenance rule at all; the swap points at banks decide whether it actually does.
- Crew pay is a maximum of block time, a fraction of duty time and guarantees at the duty level, and of the sum of duties, a fraction of time away and guarantees at the pairing level. Its excess over block time measures schedule quality.
- Crew pairing is set partitioning with one row per flight and astronomically many columns. Legality rules both protect crews and keep the column count finite.
- Column generation prices columns through duals; the pricing problem is a resource-constrained shortest path with labels and dominance. A Lagrangian bound comes free at each iteration.
- Branch-and-price needs branching rules that keep pricing intact; follow-on branching does, and every fractional solution has a fractional follow-on.
- Routing and crews interact only through short connects. Offering the crew problem a menu of routings (the extended crew pairing model) or cutting infeasible short-connect sets finds plans that sequential planning misses.

- Canadian fatigue rules depend on start time, sector count and cumulative windows; legality is never a single number. A 20-minute change can break a connection, a rest and a maintenance night at once.

Table 15.5 Key formulas of the chapter.

Quantity	Formula
Duty credit	$b_d = \max\{\mathrm{BLK}_d,\ f_d\,\mathrm{DUT}_d,\ g_d\}$
Pairing credit and cost	$\gamma_p = \max\{\sum_d b_d,\ f_p\,\mathrm{TAFB}_p,\ g_p\}$, $c_p = w\gamma_p + hn_p$
Crew pairing	$\min \sum_p c_p x_p,\ \sum_p a_{fp} x_p = 1,\ x_p \in \{0,1\}$
Reduced cost	$\bar{c}_p = c_p - \sum_f a_{fp}\pi_f$
Lagrangian bound	$z_{\mathrm{LP}} \geq z_{\mathrm{RMP}} + \frac{1}{2}\lvert F\rvert \min\{0, \min_p \bar{c}_p\}$
Follow-on value	$\phi_{fg} = \sum\{x_p : g \text{ follows } f \text{ in } p\}$
Base overnights for maintenance	$B \geq N/k$

15.12 Further reading

Barnhart's lectures on crew scheduling, maintenance routing and their integration (Barnhart 2003) are the backbone of this chapter, and the course's third problem set is a good exercise on duty-based pricing and constraint generation. Gopalan and Talluri (1998) analyze the maintenance routing problem, and Barnhart and coauthors (1998) introduce the string models for fleeting and routing. Vance, Barnhart, Johnson and Nemhauser (1997) give a crew scheduling formulation and decomposition, Barnhart and coauthors (1998) the general theory of branch-and-price, and Desrosiers and Lübbecke (2005) and Irnich and Desaulniers (2005) the column generation and pricing machinery. Cohn and Barnhart (2003) and Cordeau, Stojković, Soumis and Desrosiers (2001) integrate routing and crew pairing. For Canadian rules, read Transport Canada's AC 700-047 (Transport Canada 2024) with the crew planning team, not alone.

CHAPTER

16

Robustness and Operations Recovery

IN THIS CHAPTER

- Where delays come from, and how one late arrival travels through aircraft, crews and passengers.
- Slack, propagated delay and independent delay, and why propagation along a rotation behaves like a queue.
- What a minute of slack is worth: a corrected closed form for lognormal delays and a newsvendor rule for buffers.
- A simulated Laurentian rotation: on-time performance against buffer, and the utilization that buffer costs.
- Robust routing and retiming, which move existing slack to where delay arrives.
- Operations control on the day: delays, swaps and cancellations judged by aircraft, crews and passengers.
- Passenger-centred measures, Canada's passenger protection categories, and disruption cost in the business case.

DRAWS ON Barnhart, Airline operations lectures 1 to 3 and Barnhart and Clarke, New approaches to add robustness into airline schedules (MIT 1.206J, 2003); Belobaba, Airline operations control (MIT 16.75J, 2006); EUROCONTROL, Arriving on time (2020); Lufthansa Systems press release on Air Canada (2025); handbook; Canadian Transportation Agency

Every plan in Part III assumes that the airline will fly it as built, and the plans were built to be efficient: Chapter 15, for example, packed a day of Laurentian flying into four aircraft and five crew pairings by using tight turns and short connects, and each saving removed idle time that could have absorbed a shock. Then the day happens: a thunderstorm at LaGuardia, a de-icing queue in Montreal, a hydraulic leak on the late-morning arrival from Toronto. The operations lectures in Barnhart's course state the problem plainly:

optimizers squeeze slack out of aircraft and crew plans, so an optimized plan is easier to disrupt, and it is an open question how much airlines will pay for robustness (Barnhart 2003).

This chapter follows the plan out of the planner's hands and asks what the planner can do beforehand. Two kinds of decisions matter. Before the day, the planner decides how much slack the schedule carries and where: block times, turn times, connection times, which aircraft follows which flight. On the day, the operations control centre repairs what went wrong with delays, swaps and cancellations, choosing among the options the schedule left it. The handbook asks the network analyst to put reliability into the business case, because spare minutes have different values in different places and a connection onto the last flight of the day is more fragile than one with later options (Network Planning Handbook 2026).

We keep the Laurentian A220 day of Chapter 15: its 32 flights, the routing that meets the every-second-night maintenance rule, and the crew plan that costs CAD 16,533 a day. To this we add a delay model, a simulation of aircraft and passengers over thousands of days, an aircraft that breaks at noon, and the evening flight to Quebec City whose timing Chapter 15 left open.

16.1 Where delays come from

A flight needs an aircraft, a legal crew, and often a slot, a gate and an open runway. Anything that takes one of them away creates a disruption. The lectures sort the causes into shortages of airline resources (a mechanical problem, a crew that is missing or would break a rule) and shortages of airport and airspace resources (weather that cuts runway capacity for every aircraft at an airport, a security event) (Barnhart 2003). Belobaba's list from the operations control side adds diversions, weight restrictions, fuel stops, strikes and facility problems (Belobaba 2006j). Banked hubs are especially exposed: airlines schedule more movements into a bank than the airport can handle under instrument flight rules, so a day of low cloud turns into ground delay programs that hold flights at their origins (Barnhart 2003).

Delays then travel. A late arrival delays the next departure of the same aircraft, but its cockpit crew and its cabin crew may go to different next flights, so a single late arrival can make three later flights late through three different resources; the lecture's delay-chain example shows exactly this (Barnhart 2003). Delays also accumulate during the day. In US data for August 2000 shown in the lectures, the average lateness of flights arriving in the morning was 5 to 7 minutes,

against 21 to 26 minutes for flights arriving after 19:00, a historical pattern that the recursion of Section 16.2 explains (Barnhart 2003). The same lectures put the cost of disruptions at 2 to 3 percent of an airline's annual revenue, an estimate from around 2000 (Barnhart 2003).

What should a schedule do about this? The word robust is used in at least four senses.

DEFINITION 16.1 Robustness

A schedule is **robust** in one or more of four senses: it has the lowest expected cost once disruptions are counted; it suffers the fewest delays and disruptions of aircraft, crews and passengers; it is easy to recover when something goes wrong; and it contains disruptions so that their effect downstream stays small (Barnhart and Clarke 2003).

The senses conflict. Spare aircraft make recovery easy and cost money; short crew duties contain disruptions and need more crews. This chapter measures robustness in the second sense, with propagated delay, on-time arrivals and disrupted passengers, and prices it in the first sense at the end.

A delay model for the Laurentian day To simulate the day we need a model of what goes wrong. Each departure suffers an independent departure delay: zero with probability $1 - q$, and otherwise lognormal with a median of 10 minutes and shape parameter 0.9. The probability q depends on the departure station: 0.45 at LaGuardia, 0.40 at Toronto, 0.35 at Boston and Montreal, 0.25 at Ottawa and Quebec City, and 0.15 for the five early departures that can begin an aircraft's day (101, 102, 201, 301 and 401), so that every routing faces the same draws. With probability 0.008 a departure also suffers a large disruption of 60 to 180 minutes. The block time actually flown differs from the schedule by a normal error with mean -4 minutes and standard deviation 6, because scheduled blocks carry a little padding. Every value is illustrative. Together they give the day about 78 percent of arrivals within 15 minutes of schedule, close to the 77.7 percent of the US routings tested in the robustness lecture, a historical benchmark from 2000 (Barnhart and Clarke 2003).

16.2 Slack and propagated delay

Propagation happens at the turn. Flight j cannot leave before its aircraft has arrived on flight i and been turned, and the minimum turn time (MTT) is 35 minutes for the Laurentian A220. Any planned ground time beyond that is slack, and an inbound delay larger than the slack spills into the next departure.

DEFINITION 16.2 Slack, propagated delay and independent delay

Let flight j follow flight i on the same aircraft. The **planned turn time** is $\mathrm{PTT}_{ij} = \mathrm{PDT}_j - \mathrm{PAT}_i$, the planned departure time of j minus the planned arrival time of i, and the **slack** is $s_{ij} = \mathrm{PTT}_{ij} - \mathrm{MTT}$. If i arrives TAD_i minutes late (its **total arrival delay**), the **propagated delay** PD_{ij}, the part of j's departure delay caused by the late aircraft, is given by (16.1) below. The **independent departure delay** IDD_j and **independent arrival delay** IAD_j are the parts of j's total departure and arrival delays that the inbound aircraft did not cause (Lan, Clarke, and Barnhart 2006; Barnhart and Clarke 2003).

$$\begin{aligned} \mathrm{PD}_{ij} &= \max(\mathrm{TAD}_i - s_{ij},\ 0), \\ \mathrm{TDD}_j &= \mathrm{PD}_{ij} + \mathrm{IDD}_j, \qquad \mathrm{TAD}_j = \mathrm{PD}_{ij} + \mathrm{IAD}_j. \end{aligned} \tag{16.1}$$

Figure 16.1 draws the definitions for one evening turn at Montreal. Independent delay is everything that would have happened with an on-time aircraft: late passengers, a slow fuel truck, an air traffic control hold, a headwind. The independent arrival delay is the independent departure delay plus the deviation from the scheduled block. The same construction applies to crews, with the minimum sit replacing the minimum turn, and to passengers, with the minimum connecting time.

The split matters for data. Independent delay does not depend on the routing, so it can be measured on the routings actually flown and reused for any other. Propagated delay depends on the routing and cannot be observed for a sequence of flights that was never flown. Lan's recipe therefore runs in two steps: from history, compute PD with (16.1) and recover $\mathrm{IAD}_j = \mathrm{TAD}_j - \mathrm{PD}_{ij}$; then, for any candidate routing, rebuild the delays forward, starting each aircraft's day with its first flight's independent delay (Barnhart and Clarke 2003). That forward rebuild has a familiar structure.

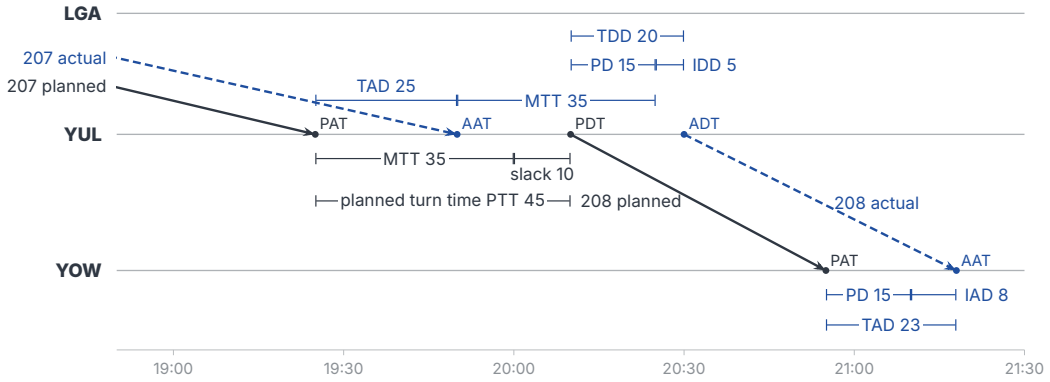


Figure 16.1 Slack and propagated delay at Montreal for the LaGuardia inbound 207 and the Ottawa departure 208, with illustrative times. The planned turn of 45 minutes is the minimum turn time of 35 plus 10 minutes of slack. Flight 207 arrives 25 minutes late, the aircraft is ready 35 minutes later at 20:25, and so 208 inherits 15 minutes of propagated delay; 5 minutes of independent delay make it leave at 20:30, and it arrives 23 minutes late, of which 8 are its own.

RESULT 16.1 Propagation along a rotation is a Lindley recursion

Number the flights of one aircraft's day $1, \dots, n$. Let s_k be the slack between flights k and $k+1$, $a_k = \mathrm{IAD}_k - s_k$, and W_k the propagated delay into flight k, with $W_1 = 0$. Then $W_{k+1} = \max(W_k + a_k,\ 0)$, and

$$W_{k+1} = \max\Big\{0,\ \max_{1 \le h \le k} \sum_{m=h}^{k} a_m\Big\}.$$

The propagated delay into each flight is nondecreasing in every upstream independent delay and nonincreasing in every upstream slack.

Proof. Flight k arrives $\mathrm{TAD}_k = W_k + \mathrm{IAD}_k$ minutes late by (16.1), so $W_{k+1} = \max(\mathrm{TAD}_k - s_k, 0) = \max(W_k + a_k, 0)$. For the closed form, use induction on k. For $k = 1$, $W_2 = \max(a_1, 0)$. If the formula holds for W_k, then

$$W_k + a_k = \max\Big\{a_k,\ \max_{h \le k-1} \sum_{m=h}^{k} a_m\Big\} = \max_{h \le k} \sum_{m=h}^{k} a_m,$$

and taking the maximum with 0 gives the formula for W_{k+1}. Every partial sum increases with each IAD_m and decreases with each s_m, and a maximum of such functions inherits both properties. ∎

This is the recursion Lindley wrote in 1952 for the waiting time of customers in a single-server queue (Lindley 1952). The aircraft is the server, each flight a customer, and propagated delay the time a customer waits because the one before it ran long. The closed form says that the delay inherited by a late-evening flight is set by the worst stretch of the whole day before it: the largest excess of independent delay over slack on any run of consecutive flights. A morning problem that the day's slack never absorbs is still there at night, consistent with the historical pattern of delays growing through the day.

WORKED EXAMPLE 16.1 A late start on line B

Line B of the Laurentian day starts in Boston with 201 to Montreal at 06:15 and flies out and back from Montreal until 209 lands there at 22:15. Its eight turns have slacks of 5, 5, 5, 5, 10, 5, 10 and 0 minutes over the 35-minute minimum. Suppose 201 arrives 60 minutes late and nothing else goes wrong.

As planned. By Result 16.1, the propagated delays into 202 to 209 are 55, 50, 45, 40, 30, 25, 15 and 15 minutes. Every later flight of the day leaves late, 275 minutes of propagated delay in total, and the last arrival is still 15 minutes late.

With 10 more minutes on every turn. The delays become 45, 30, 15 and then zero: three late flights and 90 minutes.

Eighty extra minutes of ground time removed 185 minutes of propagated delay on this day, and would have removed nothing on a day when 201 arrived on time. The planner learns that slack is an insurance purchase, and that its value has to be computed in expectation over the days that go wrong.

16.3 What a minute of slack is worth

Look at a single turn. The inbound arrival delay $X = \mathrm{TAD}_i$ is random, the slack s is chosen by the planner, and the propagated delay is $[X - s]^+$.

RESULT 16.2 Expected propagated delay is convex and decreasing in slack

If X has a finite mean, then

$$\mathbb{E}\,[X - s]^+ = \int_s^\infty \Pr(X > x)\,\mathrm{d}x,$$

$$\frac{\mathrm{d}}{\mathrm{d}s}\,\mathbb{E}\,[X - s]^+ = -\Pr(X > s),$$

so expected propagated delay is decreasing and convex in the slack. If a minute of slack costs c_s and a minute of propagated delay costs $c_d > c_s$, the slack that minimizes $c_s s + c_d \mathbb{E}\,[X - s]^+$ satisfies $\Pr(X > s^*) = c_s / c_d$.

Proof. For any realization, $[X - s]^+ = \int_s^\infty \mathbb{1}\{X > x\}\,\mathrm{d}x$. The integrand is nonnegative, so taking expectations inside the integral is allowed, which gives the first formula. Its derivative with respect to the lower limit is $-\Pr(X > s)$. Because $\Pr(X > s)$ falls as s grows, the derivative rises, so the function is convex. The derivative of the total cost, $c_s - c_d \Pr(X > s)$, is therefore nondecreasing; it is negative while $\Pr(X > s) > c_s / c_d$ and positive afterwards, so the minimum is where the two are equal. ∎

The first minute of slack saves about $\Pr(X > 0)$ minutes of expected propagated delay, a number equal to the probability that the inbound is late at all. The thirtieth saves only the probability that it is more than 30 minutes late. The optimality condition is the newsvendor rule of inventory theory, a critical fractile of the same kind as Littlewood's rule for protection levels in Chapter 18: add slack until the chance that the delay exceeds it equals the ratio of the two costs. Turns behind delay-prone inbounds and in front of expensive departures deserve more slack, which is the handbook's point that spare minutes are worth different amounts in different places (Network Planning Handbook 2026).

To use the rule we need the distribution of X. In the robustness work, the normal, exponential, gamma, Weibull and lognormal distributions were tried on total arrival delays, and a lognormal fitted best (Barnhart and Clarke 2003). Because arrivals can be early, the lognormal is shifted: $X = \theta + L$, where $\theta < 0$ and L is lognormal with median m and shape δ.

RESULT 16.3 Expected propagated delay under a shifted lognormal

Let $X = \theta + L$ with $\ln L \sim \mathcal{N}(\ln m, \delta^2)$, and let $s > \theta$. Write $z_s = (\ln(s-\theta) - \ln m)/\delta$. Then $\Pr(X > s) = 1 - \Phi(z_s)$ and

$$\mathbb{E}\,[X - s]^+ = m e^{\delta^2/2}[1 - \Phi(z_s - \delta)] - (s-\theta)[1 - \Phi(z_s)].$$

Proof. Let $a = s - \theta > 0$, so $[X - s]^+ = [L - a]^+$ and $\mathbb{E}\,[L-a]^+ = \mathbb{E}[L\,\mathbb{1}\{L > a\}] - a\Pr(L > a)$. Write $L = m e^{\delta Z}$ with Z standard normal; then $L > a$ exactly when $Z > z_s$, which gives $\Pr(L > a) = 1 - \Phi(z_s)$. Completing the square, $e^{\delta z}\varphi(z) = e^{\delta^2/2}\varphi(z - \delta)$, so

$$\mathbb{E}[L\,\mathbb{1}\{L > a\}] = m e^{\delta^2/2} \int_{z_s}^{\infty} \varphi(z - \delta)\, \mathrm{d}z = m e^{\delta^2/2}[1 - \Phi(z_s - \delta)].$$

■

A warning about the source. The lecture slides print a closed form for the expected propagated delay that amounts to $\Pr(X > 0)\,\mathbb{E}[X]$, the probability of a late arrival times the mean delay, rather than the expectation of the positive part (Barnhart and Clarke 2003). The two differ, and the printed expression can be negative: with $\theta = -25$, $m = 10$, $\delta = 1$ and no slack it gives -1.53 minutes, while Result 16.3 gives 4.30 minutes and a simulation of two million arrivals gives 4.30. Use the result above.

WORKED EXAMPLE 16.2 How much slack after a LaGuardia arrival?

Suppose Laurentian's analysts fit the arrival delays of its LaGuardia inbounds with $\theta = -15$, $m = 14$ and $\delta = 0.9$ (illustrative). The mean delay is 6.0 minutes with a standard deviation of 23.4; 47 percent of arrivals are late at all and 20 percent are more than 15 minutes late. By Result 16.3, expected propagated delay falls from 9.6 minutes with no slack to 7.6 with 5 minutes, 6.1 with 10, 5.0 with 15 and 2.9 with 30.

Take the evening turn of 207 onto 208 at Montreal, which has 10 minutes of slack (Figure 16.1). Price a minute of slack at CAD 20 (aircraft and crew time that could be used elsewhere) and a minute of propagated delay at CAD 80 (onboard passengers' time, crew overtime and the delay passed further down the

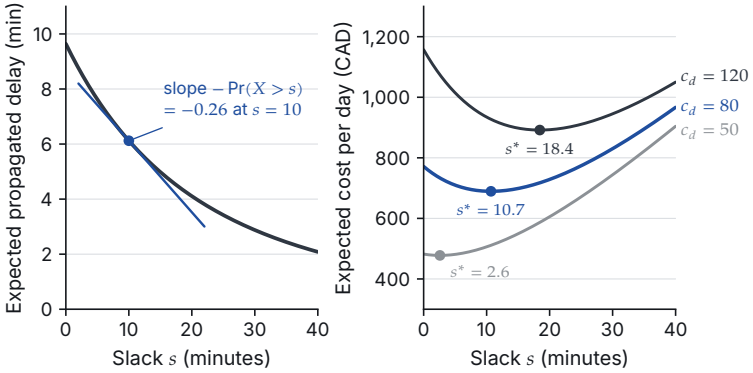


Figure 16.2 The value of slack after a LaGuardia arrival with a shifted-lognormal delay ($\theta = -15$, $m = 14$, $\delta = 0.9$). Left: expected propagated delay falls convexly as slack grows; its slope is minus the probability that the arrival delay exceeds the slack. Right: daily cost of slack at CAD 20 per minute plus propagated delay at c_d per minute; each minimum sits where $\Pr(X > s) = 20/c_d$.

line); both values are illustrative. The critical ratio is $20/80 = 0.25$, so

$$s^* = \theta + m\, e^{\delta\, \Phi^{-1}(0.75)} = -15 + 14\, e^{0.9 \times 0.674} = 10.7 \text{ minutes.}$$

The expected cost of the turn is CAD 689 a day at s^*, against CAD 771 with no slack and CAD 830 with 30 minutes (Figure 16.2). If a minute of delay costs CAD 50, the best slack falls to 2.6 minutes; at CAD 120 it rises to 18.4.

The planner learns two things. The cost curve is flat near its minimum (5 minutes of slack costs only 3 percent more than 10.7), so precision about s matters little. The delay cost matters a great deal, and it depends on what follows the turn: a crew change, a bank of connections or the last flight of the day.

16.4 Buffers and utilization

One turn at a time is a simplification, because by Result 16.1 slack on one turn also shields every later flight. To see the whole effect we simulate line B with the delay model of Section 16.1, adding the same buffer b to each of its eight turns, over 40,000 days.

Figure 16.3 shows the result. As built, 77.6 percent of line B's arrivals are within 15 minutes of schedule and only 66.8 percent of its

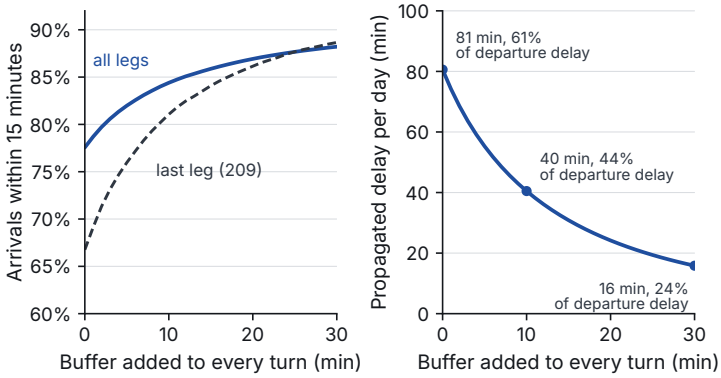


Figure 16.3 Buffer study on line B (nine A220 legs, 40,000 simulated days). Left: share of arrivals within 15 minutes of schedule, for all legs and for the last leg of the day. Right: propagated delay per day and its share of all departure delay. The first minutes of buffer buy the most.

last flight's; propagated delay averages 81 minutes a day and makes up 61 percent of all departure delay on the line. That share is far above the 20 to 30 percent that the robustness lecture reports for Continental Airlines at the time (Barnhart and Clarke 2003), because line B was built with the minimum turn plus 0 to 10 minutes everywhere. Ten minutes of buffer on each turn lift on-time arrivals to 84.4 percent (81.1 percent for the last flight) and halve propagated delay to 40 minutes; twenty more minutes add only 3.8 points. Returns diminish, as Result 16.2 predicts.

The buffer is not free. Ten minutes on each of eight turns is 80 minutes of aircraft time a day: line B's last arrival would move from 22:15 to 23:35, or a flight would have to go. EUROCONTROL's 2020 think paper documents this trade at the scale of a continent. As air traffic flow management delays grew in Europe after 2013, more than doubling per flight for traditional scheduled carriers and almost tripling for low-cost carriers, airlines responded with schedule buffers. The share of flights departing within 5 minutes of schedule fell from 63 percent in 2011 to 53 percent in 2019, while the share arriving within 5 minutes fell only from 64 to 59 percent; the widening gap is delay recovered in the air, mostly by buffer, and it levelled off from 2017 at about 7 points for low-cost carriers and 5 for traditional scheduled carriers. The paper is explicit about the price: resources planned but not used, lower aircraft utilization, and early arrivals that make the network less predictable (EUROCONTROL 2020). Two cautions apply: these are pre-2020 European figures with a 5-minute threshold, not comparable with North American 15-minute

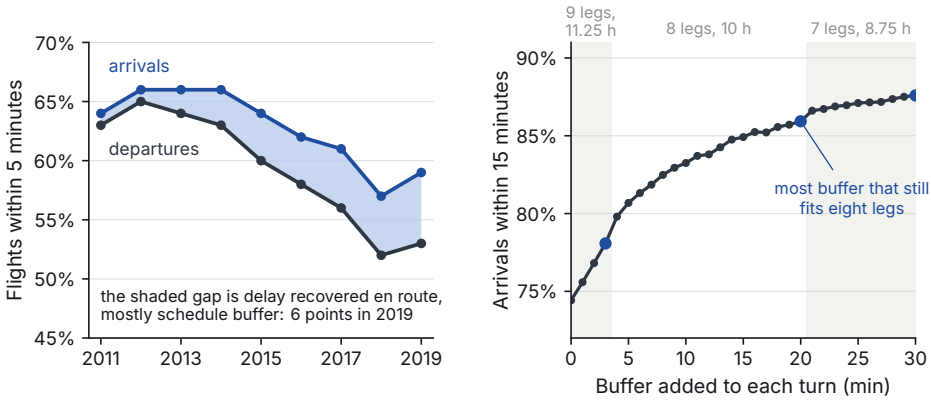


Figure 16.4 Buffers buy arrival punctuality and cost utilization. Left: share of flights within 5 minutes of schedule in the EUROCONTROL network, 2011 to 2019, redrawn from EUROCONTROL (2020); the shaded gap between arrivals and departures is delay recovered en route. Right: an illustrative Laurentian A220 shuttling between Montreal and Toronto (75-minute blocks, 40-minute base turns) inside a 06:00 to 23:00 window; legs and block hours fall in steps as the buffer per turn grows.

statistics, and the handbook treats EUROCONTROL's delay-cost references as methodological context rather than Canadian cost parameters (Network Planning Handbook 2026).

The right panel of Figure 16.4 shows the utilization side for a single aircraft shuttling between Montreal and Toronto. With no buffer it fits nine legs and 11.25 block hours into its 17-hour day, and 74.4 percent of its arrivals are within 15 minutes. A buffer of up to 3 minutes per turn keeps nine legs (78.1 percent on time); from 4 to 20 minutes it flies eight legs and 10 block hours; beyond 20 minutes, seven. Two lessons follow. Within a given number of legs, buffer is free, so an eight-leg day should use all 20 minutes (85.9 percent on time). Between leg counts, buffer is expensive: the step from nine legs to eight gives up 1.25 block hours, 11 percent of the aircraft's flying, for 7.8 points of on-time performance, and the step to seven buys at most 2.4 points more, even with all 42 minutes of buffer that seven legs leave room for. Whether the ninth leg is worth more than the reliability it destroys is a route evaluation question (Chapter 12) with an operations input.

16.5 Robust routing and retiming

Buffers buy reliability with aircraft time. Lan, Clarke and Barnhart showed two ways to obtain reliability from the slack a schedule al-

ready has (Lan, Clarke, and Barnhart 2006; Network Planning Handbook 2026).

Robust routing Maintenance routing (Section 15.2) is mostly a feasibility problem with a small direct financial impact (Barnhart and Clarke 2003): when many routings satisfy the maintenance rules, their direct costs hardly differ. Among them, choose the routing that pairs delay-prone inbounds with long turns. In the string model (15.1), give each string s the cost

$$c_s = \mathbb{E}\Big[\sum_{(i,j)\in s} \mathrm{PD}_{ij}^{s}\Big] = \sum_{(i,j)\in s} \mathbb{E}[\mathrm{PD}_{ij}^{s}],$$

the expected propagated delay over its consecutive pairs of flights. Linearity of expectation turns a stochastic objective into a deterministic one (Barnhart and Clarke 2003). Two features make the model harder than it looks. The expected delay on a connection depends on everything earlier in the string, because arrival delay accumulates (Result 16.1), so string costs are not sums of fixed arc costs and the pricing problem of column generation must carry the delay history along the path, as a richer label in the sense of Section 15.5. And delays on connections never flown must be rebuilt from independent delays. The model was solved by branch-and-price with follow-on branching (Section 15.6) (Barnhart and Clarke 2003).

Retiming The second lever moves departure times within a small window, typically 5 to 15 minutes either way, after fleeting and routing are fixed. Shifting a departure later lengthens its inbound connections and shortens its outbound ones, so slack moves from where it is idle to where passengers need it. A connection from flight i to flight j disrupts its c_{ij} passengers when the actual connection time $\mathrm{ADT}_j - \mathrm{AAT}_i$ falls below the minimum connecting time (MCT), so the expected number of disrupted passengers is $c_{ij}\,p$ with $p = \Pr(\mathrm{ADT}_j - \mathrm{AAT}_i < \mathrm{MCT})$ (Barnhart and Clarke 2003).

Give each flight i a set of copies n, each a candidate departure time, with binary $f_{i,n} = 1$ if copy n is chosen, and for each connection $(i,j) \in C$, every passenger connection and every aircraft turn, a variable $x_{i_n j_m}$ for the pair of copies chosen. With $\mathrm{DP}_{i_n j_m}$ the disrupted passengers on that pair, the connection-based model of Lan, Clarke and Barnhart chooses one copy per flight to minimize expected disrupted passengers (Lan, Clarke, and Barnhart 2006; Barnhart and Clarke 2003). Pairs of copies that would break an aircraft turn get an upper bound of zero, which keeps the routing feasible.

$$
\begin{aligned}
\min \quad & \sum_{(i,j)\in C} \sum_{n,m} \mathbb{E}[\mathrm{DP}_{i_n j_m}]\, x_{i_n j_m} \\
\text{s.t.} \quad & \textstyle\sum_n f_{i,n} = 1 && \forall i, \\
& \textstyle\sum_m x_{i_n j_m} = f_{i,n} && \forall (i,j) \in C,\ n, \\
& \textstyle\sum_n x_{i_n j_m} = f_{j,m} && \forall (i,j) \in C,\ m, \\
& f_{i,n} \in \{0,1\},\ \ x_{i_n j_m} \geq 0.
\end{aligned}
\tag{16.2}
$$

RESULT 16.4 Integral copies make the connection variables integral

In (16.2), if every $f_{i,n}$ is 0 or 1, the constraints force $x_{i_n j_m} = f_{i,n} f_{j,m}$. Only the copy variables need to be declared integer.

Proof. Fix a connection (i,j) and let copies n^* of i and m^* of j be the chosen ones. For $n \neq n^*$, $\sum_m x_{i_n j_m} = 0$ with nonnegative terms, so every $x_{i_n j_m}$ is zero; likewise $x_{i_n j_m} = 0$ whenever $m \neq m^*$. The constraint for n^* then reads $x_{i_{n^*} j_{m^*}} = 1$. In every case $x_{i_n j_m} = f_{i,n} f_{j,m}$. ∎

The disaggregated linking constraints matter. The lecture compares this formulation with an aggregated one that sums the linking constraints over connections: on a 278-flight test network the disaggregated model was integral at the root and solved in 13 seconds, while the aggregated one was still unsolved after 107 hours (Barnhart and Clarke 2003). The strength of a formulation, how close its linear relaxation comes to the integer optimum (Appendix B), mattered more than its size.

Table 16.1 Historical results of robust routing and retiming on US networks of 20 to 278 flights, built on July 2000 data and tested on August 2000 data (adapted from Barnhart and Clarke, 2003).

Method and test	Measure	Reported change
Robust routing, July 2000 (data used to build)	propagated delay	−47%
Robust routing, August 2000 (new data)	propagated delay	−44%
Robust routing, test networks	disrupted passengers	−10.8%
Robust routing, test networks	arrivals within 15 min	77.7% to 79.3%
Retiming ±15 min, MCT 30 min, August	expected disrupted passengers	−39.7%
Retiming ±10 min, MCT 30 min, August	expected disrupted passengers	−32.3%
Retiming ±5 min, MCT 30 min, August	expected disrupted passengers	−20.0%
Retiming ±15 min, MCT 20 min, August	expected disrupted passengers	−27.1%

Table 16.1 summarizes the reported results; all are historical and come from four sub-networks of one US carrier. One inconsistency in the source should be known: for the largest network, the reported new propagated delay does not equal the old value minus the reported reduction. The percentages in the table follow the reduction column, which agrees with its totals; computed from the new-delay column instead, the reductions would be about 38 and 37 percent (Barnhart and Clarke 2003). Either way, re-routing alone removed well over a third of propagated delay and a tenth of passenger disruptions at almost no cost, and with a 30-minute MCT, small retiming moves removed a fifth to two-fifths of expected disruptions. Benefits shrink with the MCT, because short connection standards leave less risk to remove, and with the window. Copies every 5 minutes were enough: copies every minute multiplied the problem size 10 to 17 times for less than 0.6 percent more benefit (Barnhart and Clarke 2003).

WORKED EXAMPLE 16.3 Which Laurentian routing is robust?

Chapter 15 found eight ways to route the four aircraft through the day's three swap points and chose among them by maintenance feasibility and crew cost. Here we simulate all eight over the same 20,000 days (the same random delays for each routing, so that differences are not noise) and count propagated delay, on-time arrivals and disrupted passengers. Passengers follow an illustrative pattern: 80 local passengers on every flight; 12 connecting passengers on every pair of an inbound and an outbound flight at Montreal between 45 minutes and 3 hours apart that does not return to the inbound's origin; and 79 passengers a day connecting from five inbounds to two evening transatlantic departures, an assumed 20:30 to Seville and 21:45 to Paris. One illustrative minimum connecting time of 45 minutes applies to every connection, a simplification: Chapter 13 uses 60 minutes for a transborder arrival, whose passengers clear customs at Montreal. A passenger who misconnects takes the next flight to the same destination that leaves at least 45 minutes after arrival (seats are assumed to be free), or the first flight the next morning; a missed transatlantic connection costs 24 hours. Same-aircraft connections never break, because the outbound waits for its own aircraft. Table 16.2 shows the results.

Propagated delay barely moves. It ranges from 269 to 282 minutes a day, 5 percent, because every routing inherits the same thin turns and the three swap points offer little slack to redistribute. The difference is nonetheless real: the as-built routing with 107 swapped to 208 has 9.4 minutes a day less propagated delay than the every-second-night routing, with a simulation standard error of 0.3.

Passengers rank the routings differently. The routing with the least propagated delay is not the one with the fewest disrupted passengers. Keeping 107 and 108 on one aircraft at the 19:00 bank protects the Toronto to Quebec City connection: each routing that does so disrupts 6.5 fewer passengers a day than the same routing with the 19:00 swap.

Crew cost pulls the other way. Under the every-third-night rule, Chapter 15 would pick the routing that saves CAD 352 a day of crew cost. It disrupts 4.3 more passengers a day than the every-second-night routing, so the saving is worth having only if a disrupted passenger costs Laurentian less than CAD

82. With an average delay for a disrupted passenger of 17 hours in this model, it almost certainly costs more.

The planner learns that a routing chosen for one objective is rarely the best for the others, and that a few disrupted passengers can outweigh a crew saving that looked decisive in Table 15.4.

Table 16.2 The eight Laurentian routings of Chapter 15 simulated over the same 20,000 days. Arrivals within 15 minutes range only from 77.5% to 78.1%.

Routing (morning, 17:00, 19:00 bank)	Maintenance rule met	Crew cost (CAD per day)	Propagated delay (min per day)	Disrupted passengers per day
as built: 101→102, 406→407, 107→108	none	16,633	270.1	22.8
as built, 107→208 at 19:00	every third night	16,467	268.9	29.2
406→307 at 17:00	none	16,400	275.0	27.0
406→307 and 107→208	every third night	16,233	273.9	33.5
101→401 in the morning	every second night	16,533	278.3	24.5
101→401 and 107→208	every third night	16,367	276.7	31.0
101→401 and 406→307	every third night	16,181	282.3	28.8
101→401, 406→307 and 107→208	none	16,014	280.7	35.3

16.6 Operations control and recovery

On the day, the plan belongs to operations control. Belobaba describes three levels: system operations control, which manages the whole day in real time and reassigns aircraft, crews and other resources when operations become irregular; dispatch, which plans each flight's route and fuel; and station control, which runs the turnarounds (Belobaba 2006j). The objective is to fly as close to the plan as possible, subject to safety, on-time performance and cost.

The recovery considerations Belobaba lists from a US carrier include the number of customers inconvenienced and their rebooking, the downline effect on other customers, the work of rebalancing aircraft and crews, maintenance requirements, completion factor and on-time performance, slots and weather, with route profitability only as a tie-breaker. At a hub, whether to hold a departure for late connecting passengers depends on how many later flights go to the same destination, how many seats they have, where the aircraft goes next and what hotels cost (Belobaba 2006j).

The controllers' levers are few (Barnhart 2003). They can delay a flight until its aircraft and crew are ready. They can swap aircraft between flights, which turns one long delay into several short ones and can keep a bank together. They can cancel, and in a hub network a round trip from the hub to a spoke and back almost always exists, a cancellation that leaves every aircraft where it should be; it is cheapest when both flights belong to the same crew duty. They can call reserve crews, deadhead crews or lay them over. In the practice the lectures describe, decisions were taken in sequence, aircraft first, then crews, then passengers, with fleet-specific coordinators able to veto a proposal (Barnhart 2003).

The classic optimization model for the aircraft step is a time-space network (Chapter 14) in which each flight has copies at candidate departure times and a cancellation option. The lectures attribute this aircraft recovery model to Yu and coauthors (Barnhart 2003). With $x_f^t = 1$ if flight f departs at time t, $z_f = 1$ if it is cancelled, delay cost d_f^t, cancellation cost c_f, and ground arcs y as in fleet assignment, it reads

$$\begin{aligned} \min\ & \sum_f \sum_{t \in T_f} d_f^t\, x_f^t + \sum_f c_f\, z_f \\ \text{s.t. } & \textstyle\sum_{t \in T_f} x_f^t + z_f = 1 \quad \forall f, \\ & \text{aircraft balance at every station and time,} \\ & \text{start-of-day positions and end-of-day requirements,} \\ & x_f^t,\ z_f \in \{0,1\}, \quad y \geq 0. \end{aligned} \tag{16.3}$$

The weakness of (16.3) is its cost coefficients. A delay of a given length or a cancellation has a fixed cost, but its true cost depends on the passengers it strands and the crews it breaks, which depend on the whole solution; tests on airline data found the solution very sensitive to the coefficients. In the lecture's two-aircraft example, holding both aircraft so that the bank stays together is best for passengers,

and the model never chooses it (Barnhart 2003). Passenger-centred models, developed by Bratu and Barnhart (Bratu and Barnhart 2006) and presented in the lectures, therefore put passengers in the objective: one minimizes the number of disrupted passengers, another their total delay, both over flight copies and cancellations. In the lectures, crew feasibility is kept by constraints that forbid combinations of copies breaking a planned duty and by not generating copies that would exceed the maximum elapsed duty time (Barnhart 2003). Only the copy variables need to be integer in the first model, which solves quickly; in simulation it reduced disrupted passengers by 10 to 20 percent on bad days without disrupting additional crew duties, a historical result (Barnhart 2003).

WORKED EXAMPLE 16.4 An aircraft goes technical at noon

On the Laurentian day flown with the every-second-night routing, tail 1 flies 102 to Toronto and 103 back, lands in Montreal at 10:55 and is found to need a two-hour repair. It is due to fly 104 to LaGuardia at 11:40, then 105, 106, 107 and 108 to Quebec City. Tail 2, which flies line B, lands from Boston on 203 at 11:50 and is due out on 204 to Quebec City at 12:30. Figure 16.5 and Table 16.3 compare four options.

(**a**) **Wait for the repair.** 104 leaves at 12:55, 75 minutes late, and the delay rolls through tail 1's afternoon: 105, 106, 107 and 108 leave 70, 65, 60 and 45 minutes late. 107 now lands at 20:10, too late for the 20:30 Seville departure (10 passengers miss it) and for the 20:10 to Ottawa (12 passengers, stranded until morning).

(**b**) **Swap at noon.** Tail 2 takes 104 at 12:25 and continues tail 1's day; repaired tail 1 takes 204 at 12:55 and continues line B. The delay is shared: nine late flights instead of five, but 225 delay minutes instead of 315, and every transatlantic connection holds. The crew plan survives unchanged, because each crew either stays with its aircraft or has at least 50 minutes to change. Maintenance does not: tail 2 slept in Boston last night and now ends the day in Quebec City, a second night away from Montreal, which the every-second-night rule forbids.

(**c**) **Swap, then swap back at the 19:00 bank.** Returning 108 to tail 1 and 208 to tail 2 sends tail 2 home to Montreal for the night. It breaks two crew connections instead: the crew of 107 would have 25 minutes to change aircraft for 108, and the crew of 207

45 minutes for 208, both below the 50-minute minimum. Crew control can re-pair them, exchanging their evening flights so that each crew stays with its aircraft: one crew flies 106, 107, 208 and 209 (a 6 h 45 min duty, counting report 45 minutes before the delayed first departure) and the other 205, 206, 207 and 108 (7 h 35 min), both legal under Table 15.3. The two crews trade places: one comes home a day early, the other gains a night in Quebec City and a longer pairing.

(d) Cancel the LaGuardia round trip. Cancelling 104 and 105 puts tail 1 back on schedule with 106 at 16:00. No flight leaves late, but 208 passengers lose their flights; the later LaGuardia flights have seats for 48 of them, and 160 wait until the next day.

An aircraft-centred model that prices only delay minutes and cancellations could choose (d). Passengers rank the options (c), (b), (a), (d), and option (d) costs them more than five times the passenger-hours of (b). The best option for passengers is feasible only because crew control can re-pair two duties, and (b) is feasible only under a looser maintenance rule. The planner learns that the menu the controllers face was written months earlier: the noon swap exists because the schedule brings two A220s to Montreal within an hour of each other, and the cancellation is costly because only one later round trip to LaGuardia that day can absorb the stranded passengers.

Table 16.3 Recovery options for the noon repair. Departure delay is summed over flights; passengers stranded overnight are counted as 18 hours late and missed transatlantic connections as 24 hours. Rebooking uses Laurentian's own flights only.

Outcome	(a) Wait	(b) Swap	(c) Swap and back	(d) Cancel
Late departures	5	9	11	0
Departure delay (min)	315	225	225	0
Disrupted passengers	22	12	12	208
Stranded overnight	22	12	12	160
Passenger delay (hours)	946	571	566	3,076
Crew plan	as planned	as planned	two duties re-paired	as planned
Maintenance plan	as planned	broken for tail 2	as planned	as planned

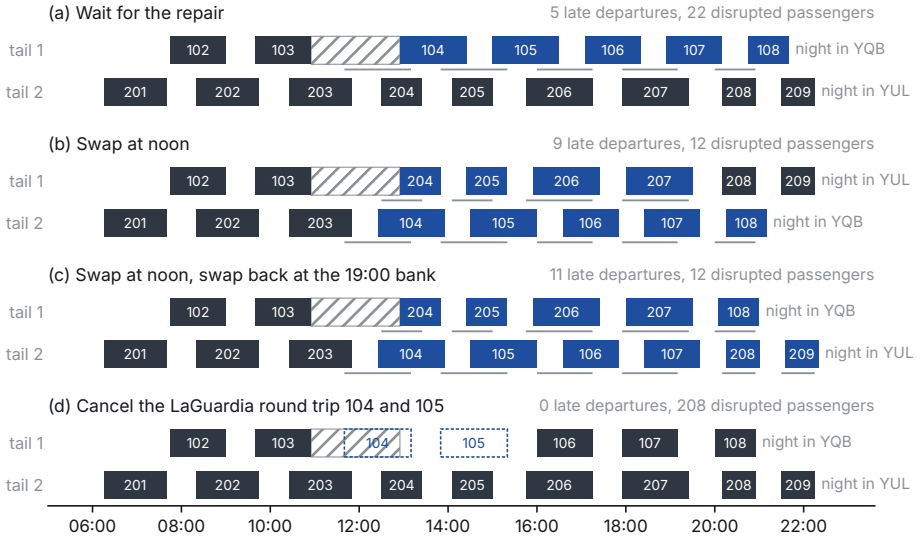


Figure 16.5 Four ways to recover from tail 1's two-hour repair at Montreal (hatched), drawn for the two tails involved. Bars are flights as operated, late ones in the accent colour; grey lines under late flights mark their planned times; dashed outlines are cancelled flights. The note beside each panel title counts late departures and disrupted passengers.

IN PRACTICE Operations control at Air Canada in 2026

Air Canada's operations control runs on Lufthansa Systems' NetLine/Ops ++, which the vendor announced as live at the airline in January 2025, together with a move of Air Canada's Lufthansa Systems applications to the vendor's cloud; the release names the vendor's Lido Flight 4D as another tool in use (Lufthansa Systems 2025a). The press release describes the benefits in qualitative terms only (situational awareness, real-time decision support, faster responses to disruptions, punctuality) and gives no measured outcomes, so it says nothing about which recovery models the system runs; the Air Canada spokesperson quoted is the senior vice-president responsible for global airports and operations control (Lufthansa Systems 2025a). Disruption also shows up in the airline's financial reporting: in the second quarter of 2026, operated capacity grew 0.3 percent year over year, below guidance of 0.5 to 1 percent, mainly because weather disruptions reduced completion late in the quarter (Air Canada 2026e). For a network analyst the practical points are that controllers see the schedule the planners publish, as a set of rotations, crew duties and connections, and that

every swap opportunity and two-flight cancellation in that schedule is a recovery option they may need.

16.7 Measuring what passengers experience

Punctuality is usually stated per flight, for example as the share of flights arriving within 15 minutes of schedule (A15) (Network Planning Handbook 2026), and the flight is the wrong unit for passengers. A passenger cares about arrival at the final destination, and the passengers who suffer most are those whose trip breaks.

> **DEFINITION 16.3 Disrupted passenger**
>
> A passenger is **disrupted** when the planned itinerary becomes infeasible, because a flight in it is cancelled or because a connection becomes shorter than the minimum connecting time. A passenger's **delay** is the actual minus the scheduled arrival time at the final destination (Barnhart and Clarke 2003; Bratu and Barnhart 2006).

If a share σ of passengers is disrupted, with average delay D_d, and the rest have average delay D_{nd}, then the average passenger delay and the disrupted passengers' share of it are

$$\bar{D} = \sigma D_d + (1-\sigma) D_{nd}, \qquad \text{share of delay} = \frac{\sigma D_d}{\bar{D}}. \tag{16.4}$$

Because D_d is measured in hours and D_{nd} in minutes, a small σ can dominate. In the data for one US carrier in August 2000 presented in the lectures, 3.2 percent of passengers were disrupted, waited 320 minutes on average against 16 for the others, and accounted for 40 percent of all passenger delay; the average passenger was delayed 25.7 minutes while the average flight was delayed 15.4, a ratio of 1.7 (Barnhart 2003). Connecting passengers were 2.75 times as likely to be disrupted as local ones, disruptions late in the day were much longer, and passengers on routes with more daily frequencies were re-accommodated faster. All of these figures are historical.

The Laurentian simulation tells the same story (Figure 16.6). Under the every-second-night routing the day carries 2,963 passengers, 403 of them connecting. Flights arrive on average 13.2 minutes late (counting early arrivals as on time) and passengers 21.4 minutes

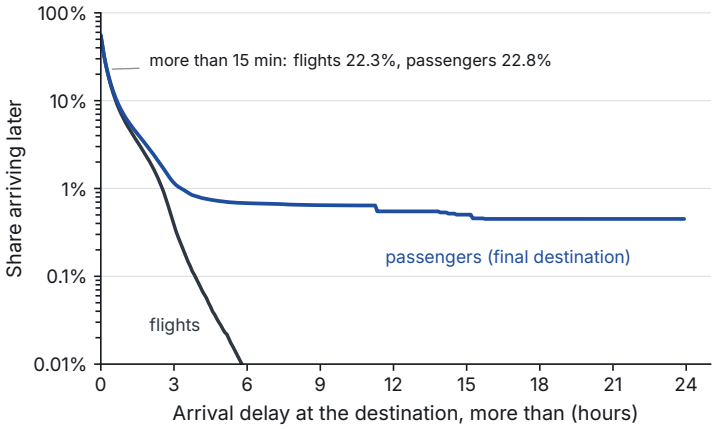


Figure 16.6 Share of flights and of passengers arriving more than a given delay late on the Laurentian day (every-second-night routing, 20,000 simulated days, logarithmic scale). The two curves agree up to about an hour; beyond that, missed connections, the end of the day and transatlantic misconnects give passengers a long tail.

late, a ratio of 1.63. Only 24.5 passengers a day, 0.83 percent, are disrupted, but they wait 17 hours on average and, by (16.4), account for 39 percent of all passenger delay. A flight measure cannot see this tail: at 15 minutes flights and passengers look alike, at six hours almost no flight is that late while 0.7 percent of passengers are.

The handbook gives the operational rule: measure passenger arrival at the final destination alongside flight punctuality, because a flight-level improvement can hide worse outcomes for connecting itineraries (Network Planning Handbook 2026). EUROCONTROL adds that targets and staff incentives built on departure punctuality are slow to change even though arrival punctuality is what passengers value (EUROCONTROL 2020). The definitions behind a punctuality figure (departure or arrival; a threshold of 0, 5, 14 or 15 minutes; which flights the completion factor excludes) should be read before two reports are compared (Network Planning Handbook 2026).

> **PITFALL**
>
> Reading flight punctuality as passenger experience. On-time performance can be improved by padding blocks or by cancelling the latest flights, and both can make passengers worse off: in Example 16.4, cancelling the LaGuardia round trip pro-

duces no late departure at all and the most disrupted passengers by far. Report disrupted passengers and delay at the final destination next to A15.

16.8 Disruption in the business case

A schedule change that adds a connection, tightens a turn or moves the last departure of the day changes expected disruption, and the handbook's first-pass framework puts that change into the decision (Network Planning Handbook 2026):

$$\text{expected net contribution} = \text{planned contribution} - \mathbb{E}[\text{incremental disruption cost}]. \tag{16.5}$$

The expected cost should come from delay scenarios, missed connections, cancellations, re-accommodation and rotation recovery, and it should not double count disruption costs that the base cost forecast already contains (Network Planning Handbook 2026). The handbook's invented Montreal to Seville bridge carries an expected incremental disruption cost of CAD 2,000 per departure, which turns an initial flight contribution of CAD 13,000 into a network contribution of CAD 8,000 after other network effects (Chapter 12) (Network Planning Handbook 2026).

Canadian passenger protection Part of the disruption cost is set by regulation. Under the Air Passenger Protection Regulations, the Canadian Transportation Agency distinguishes three categories of flight disruption, and an airline's obligations differ by category (Table 16.4) (Canadian Transportation Agency 2023; Government of Canada 2019). In all three the airline must tell passengers the reason and their entitlements, update them every 30 minutes, and offer rebooking, or in some cases a refund, after a cancellation or a delay of three hours or more. Standards of treatment (food, drink, a means of communication and, for an overnight wait, a hotel) apply once a passenger told of the disruption less than 12 hours before departure has waited two hours, but only when the cause is within the airline's control, safety-related causes included. Compensation for inconvenience is owed only for disruptions within the airline's control that are not required for safety and were announced 14 days or less before departure; it grows with the delay at arrival, with thresholds at three, six and nine hours. For disruptions within their control, large airlines, those that carried at least two million

passengers in each of the two previous years, must rebook on their own or a partner's flight leaving within nine hours and otherwise on another airline's flight within 48 hours (Canadian Transportation Agency 2023). Overbooking and denied boarding, which have their own rules, are covered in Chapter 19.

Table 16.4 The three disruption categories of the Air Passenger Protection Regulations as described by the Canadian Transportation Agency's guide, which is not a legal text; the regulations prevail, and the compensation amounts they set are not reproduced here.

Category	Examples given by the regulator	Standards of treatment	Compensation
Within the airline's control	commercial decisions, staffing, flights consolidated for low demand, scheduled maintenance	yes, after a two-hour wait	yes, by arrival delay
Within its control, required for safety	safety problems found in checks before or after a flight	yes, after a two-hour wait	no
Outside the airline's control	weather, security, medical emergencies, air traffic control instructions, orders of authorities	no	no

The handbook draws the planning consequence: rebooking, assistance and compensation exposures are not interchangeable, so a model should neither treat every delay as compensable nor assume that weather costs nothing commercially, and the costs should be estimated from operational experience under the approved rules (Network Planning Handbook 2026). A tight connection may improve the published schedule while raising expected misconnections, hotel nights and rerouting.

IN PRACTICE Passenger protection rules in October 2026

The rules are in flux. In June 2023 Parliament amended the Canada Transportation Act to eliminate the three categories, and proposed regulations published in the Canada Gazette on 21 December 2024 would put this into effect: compensation would be owed unless the disruption falls in an exhaustive list of exceptional circumstances, assistance would start after a two-hour delay, and the refund deadline would shorten from 30 to 15 days (Canadian Transportation

Agency 2024). In September 2026 the Agency's transition book for its new chair still described the amendments as proposed and reported a backlog of more than 95,000 complaints (Canadian Transportation Agency 2026b). Bill C-31, a budget implementation bill tabled in May 2026, would transfer the making of these regulations from the Agency to the Minister of Transport and let designated third parties resolve complaints (Canadian Transportation Agency 2026b; McCarthy Tétrault 2026); it passed second reading on 3 June 2026, and in late October 2026 it was still at committee stage in the House of Commons (Parliament of Canada 2026). A business case written today should state which version of the rules it assumes.

WORKED EXAMPLE 16.5 Selling a connection onto the last flight to Quebec City

In Example 15.6 the commercial team moved 107 from Toronto to 18:15, and the planner repaired the day by moving 108, the last departure to Quebec City, to 20:15. The repair opened a 50-minute connection from 207, the LaGuardia inbound due at 19:25, which clears this chapter's 45-minute MCT (though not the 60-minute transborder MCT of Chapter 13) and could be sold. Suppose 12 passengers a day would buy it, each adding an illustrative CAD 150 of contribution, so the planned contribution is CAD 1,800 a day. A passenger who misconnects onto the last flight of the day is stranded until the next morning; let K be the cost of each such passenger to Laurentian (hotel, meals, rebooking, goodwill and any compensation owed), illustratively CAD 400.

The simulation of Section 16.5 gives the misconnection probability. Flight 207 comes late in line B's day, and by the evening its aircraft carries the day's accumulated delay: with 108 at 20:15 the connection fails 36 percent of the time, stranding 4.3 passengers a day on average (Table 16.5). The expected disruption cost is CAD 1,709 a day and the expected net contribution (16.5) of selling the connection only CAD 91; at any K above CAD 421 it loses money.

Moving 108 later buys protection. At 20:30 the connection is 65 minutes, fails 20 percent of the time and earns CAD 838 a day; at 20:45 it fails 13 percent of the time and earns CAD 1,177.

The later departure also helps 108's own punctuality, because its aircraft gets a longer turn (81.5 percent of its arrivals within 15 minutes at 20:30, against 77.1 percent at 20:15), and the crew plan and the maintenance routing are unchanged. What it costs is commercial: every local passenger to Quebec City arrives up to half an hour later, which the demand models of Chapter 13 would have to price.

The planner learns that a legal connection is not a reliable one. The MCT is a minimum, not a target, and a connection onto the last flight of the day, from an inbound at the end of a tight rotation, needs either more time or a decision not to sell it.

Table 16.5 Selling the LaGuardia to Quebec City connection onto the last departure (20,000 simulated days). In the original plan 107 leaves Toronto at 17:55; in the other columns at 18:15. A contribution of CAD 150 per connecting passenger and K = CAD 400 per stranded passenger are illustrative; breakeven K is the stranding cost at which the connection stops paying.

Departure of 108	20:00, original	20:15, repair	20:30	20:45
Connection from 207 (min)	35, not sold	50	65	80
Misconnect probability		36%	20%	13%
Passengers stranded per day	0	4.3	2.4	1.6
Expected cost (CAD/day)	0	1,709	962	623
Net contribution (CAD/day)	0	91	838	1,177
Breakeven K (CAD)		421	748	1,157
108 within 15 minutes	74%	77%	82%	84%

Slack is worth most where delay arrives and where a failure is expensive: behind late-prone inbounds, in front of connections onto the last flight of the day, and at the turns that feed a bank. The same minutes spread evenly buy much less.

16.9 Summary

- Disruptions come from airline resources (aircraft, crews) and from airports and airspace (weather, capacity), and they travel through aircraft, cockpit crews, cabin crews and passengers along different paths.
- Propagated delay is the part of a departure delay caused by a late inbound aircraft: $\mathrm{PD}_{ij} = \max(\mathrm{TAD}_i - s_{ij}, 0)$. Along a rotation it follows Lindley's queueing recursion, so an evening flight inherits the worst stretch of the day.
- Expected propagated delay is convex and decreasing in slack, with slope $-\Pr(X > s)$; the cost-minimizing slack satisfies the newsvendor condition $\Pr(X > s^*) = c_s/c_d$. For a shifted log-normal there is a closed form, and the version printed in the robustness lecture is wrong.

- Buffers have diminishing returns and a stepwise cost in utilization: buffer is free until it costs a leg. EUROCONTROL's 2011 to 2019 data show European airlines buying arrival punctuality with buffers.
- Robust routing and retiming reallocate the slack a schedule already has; historical tests removed over a third of propagated delay and up to two-fifths of expected disrupted passengers. A tight retiming formulation solved in seconds where a weaker one failed after days.
- Operations control recovers with delays, swaps, cancellations and crew repairs. Aircraft-centred models with fixed delay and cancellation costs miss what passengers lose; passenger-centred models put disrupted passengers in the objective.
- Flight punctuality understates passenger delay: a few disrupted passengers carry a large share of it. Report delay at the final destination and disrupted passengers alongside A15.
- Expected net contribution subtracts expected incremental disruption cost from planned contribution. Canada's passenger protection categories decide which costs apply, and the rules were under revision in 2026.

Table 16.6 Key formulas of the chapter.

Quantity	Formula
Slack	$s_{ij} = \mathrm{PDT}_j - \mathrm{PAT}_i - \mathrm{MTT}$
Propagated delay	$\mathrm{PD}_{ij} = \max(\mathrm{TAD}_i - s_{ij}, 0)$
Delay decomposition	$\mathrm{TDD}_j = \mathrm{PD}_{ij} + \mathrm{IDD}_j$, $\mathrm{TAD}_j = \mathrm{PD}_{ij} + \mathrm{IAD}_j$
Lindley recursion	$W_{k+1} = \max(W_k + \mathrm{IAD}_k - s_k, 0)$
Value of slack	$\mathbb{E}\,[X - s]^+ = \int_s^\infty \Pr(X > x)\,\mathrm{d}x$, slope $-\Pr(X > s)$
Newsvendor slack	$\Pr(X > s^*) = c_s / c_d$
Lognormal case	$m e^{\delta^2/2}[1 - \Phi(z_s - \delta)] - (s - \theta)[1 - \Phi(z_s)]$
Expected disrupted passengers	$c_{ij} \Pr(\mathrm{ADT}_j - \mathrm{AAT}_i < \mathrm{MCT})$
Passenger delay	$\bar{D} = \sigma D_d + (1 - \sigma) D_{nd}$
Business case	expected net contribution = planned contribution − expected disruption cost

16.10 Further reading

Barnhart and Clarke's robustness lecture (Barnhart and Clarke 2003) and the three airline operations lectures of Barnhart's course (Barnhart 2003) are the core of this chapter, and Belobaba's lecture on operations control (Belobaba 2006j) describes the organization

that uses the plan. The published versions are Lan, Clarke, and Barnhart (2006) on robust routing and retiming and Bratu and Barnhart (2006) on passenger-centred recovery. Clausen et al. (2010) survey disruption management models for aircraft, crews and passengers, and the chapter on irregular operations in Belobaba, Odoni and Barnhart's textbook, which the handbook recommends, gives the industry context (Belobaba, Odoni, and Barnhart 2015; Network Planning Handbook 2026). EUROCONTROL's think paper is a short, data-based account of buffers and punctuality (EUROCONTROL 2020). For Canadian obligations, start from the Canadian Transportation Agency's guide and the regulations themselves (Canadian Transportation Agency 2023; Government of Canada 2019).

PART IV

Revenue

Part III decided how many seats to fly, between which cities and at what times. Part IV is about what those seats earn: how airlines build fare structures, how revenue management decides which fares are on sale on each flight and each day, and how overbooking and network control turn a schedule into revenue. For a network planner this is not someone else's business, because the value of every seat in the schedule, and therefore of every capacity decision, depends on how it is priced and sold.

CHAPTER

17

Pricing and Fare Structures

IN THIS CHAPTER

- Why one economy cabin is sold at many prices, and how much price discrimination can add to revenue.
- One price versus two on a linear demand curve, derived, and what changes when the aircraft is full.
- Fences and self-selection: the incentive constraints a fare structure must satisfy, and when a restriction pays for itself.
- Fare products, restrictions and the disutility each restriction imposes on different travellers.
- From restricted to simplified to branded fares: the low-cost challenge, the ePODS evidence and current practice.
- Who decides what: pricing sets fares and rules, revenue management sets availability; competitive fare matching.
- Laurentian's YUL–SVQ fare ladder, the running example of Chapter 18.

DRAWS ON Belobaba, Airline economics review; Introduction to revenue management; Pricing challenges: ePODS and reality; Willingness to pay and competitive revenue management; O&D control: what have we learned? (MIT 16.75J, 2006); Competition Bureau (2025); Air Canada Q2 2026 MD&A; 2026 research digest

When Laurentian opens sales for a Montreal–Seville departure, it does not post one price for a seat in economy. It posts five, from CAD 419 to CAD 1,249, each with its own conditions on when the ticket must be bought, whether it can be changed or refunded and how much baggage it includes. Two passengers in adjacent seats, eating the same meal and landing at the same minute, may have paid fares that differ by a factor of three. This chapter explains why an airline sells one cabin at many prices, how it stops the traveller who would pay CAD 1,249 from buying the CAD 419 ticket, and how those arrangements have changed under pressure from low-cost carriers and online shopping.

The planner's interest is direct. Route evaluation (Chapter 12) multiplies passengers by an average fare, and the spill models of Chapter 9 value a seat at that average. Both numbers are produced by the fare structure: by how many price points there are, how well the restrictions separate travellers, how competitors respond, and how revenue management allocates seats among fares. An average fare of CAD 620 on YUL–SVQ is not a property of the route. It is the outcome of a pricing design and a control policy, and it moves when either changes.

The chapter starts from the economics of price discrimination, which an industrial organization economist will recognize, and shows how airlines implement it with fare products and fences. It then follows fare structures through their history: the heavily restricted structures of the 1990s, their simplification under low-cost competition in the 2000s, and the branded fares and unbundled services of today. It closes with the division of labour between pricing and revenue management, competitive fare matching, and the YUL–SVQ fare ladder that Chapter 18 will optimize.

17.1 One cabin, many prices

Two features of the airline business make differential pricing natural. The first is the cost structure. Once a flight is scheduled, almost all of its cost is committed: Laurentian's avoidable cost of a YUL–SVQ departure is about CAD 88,000 with 150 passengers on board, of which only CAD 50 per passenger varies with the number carried (Chapter 3). An empty seat earns nothing and cannot be stored, so the relevant question for each seat is not what it cost but what someone will pay for it. A single fare set at the average cost, about CAD 587 per passenger, would turn away travellers willing to pay CAD 450 for a seat that will otherwise leave empty, and would charge CAD 587 to travellers willing to pay CAD 1,500.

The second feature is heterogeneity. Travellers differ in how much they are willing to pay and, just as important, in how they trade money against convenience. Business travellers decide late, change plans and value flexibility, and their demand responds little to price. Leisure travellers plan ahead, accept conditions and shop on price. The reading library quotes typical market price elasticities of about -0.8 for business and -1.6 for leisure travel, within an observed range of -0.8 to -2.0 (Belobaba 2006d); these are historical estimates, and Chapter 4 treats elasticities properly.

Economists distinguish two reasons for charging different prices. **Price discrimination** charges different prices for the same product,

with the same cost of production, according to willingness to pay. **Product differentiation** charges different prices for products that differ in characteristics and in cost (Belobaba 2006d). Airline fare structures do both. A flexible economy ticket and a restricted one are different products, but the difference in their cost to the airline is the expected cost of refunds and changes, a few dollars, while the difference in their price is hundreds of dollars. The gap is priced on willingness to pay, which makes it price discrimination in the economist's sense.

Differential pricing raises revenue from both ends of the market. Discount travellers who would not fly at a single fare fill seats that would otherwise leave empty, and high-fare travellers pay more than a single fare would extract. Belobaba reports studies showing that most traditional high-cost airlines could not cover their total operating costs with any single fare level (Belobaba 2006d). Consumers can gain too: discount travellers fly who would otherwise stay home, and the extra traffic supports frequencies that high-fare travellers value. How unequal revenue per passenger is in practice shows in current results: Air Canada's premium cabins produced about 30 percent of its passenger revenue in the second quarter of 2026 (Air Canada 2026e).

Revenue maximization has two components (Belobaba 2006l). **Pricing** designs the fare products and sets their levels; **revenue management** decides, flight by flight and day by day, how many seats each fare may sell. This chapter is about the first. Chapters 18 to 21 are about the second.

17.2 Price discrimination in theory

The classical taxonomy sorts price discrimination by what the seller knows about each buyer (Tirole 1988; Varian 1989).

DEFINITION 17.1 Three degrees of price discrimination

- **First-degree** (perfect) discrimination charges each customer exactly his or her willingness to pay. The seller captures all consumer surplus.
- **Second-degree** discrimination offers the same menu of options to everyone and lets customers sort themselves. The seller cannot tell customers apart, but customers with different tastes choose different options.
- **Third-degree** discrimination charges different prices to groups the seller can identify, such as students, residents of a country, or customers buying in a given market or point of sale.

Airlines use all three, at least approximately. Fares are set by origin-destination market and often by point of sale, which is third-degree discrimination: there is no economic reason for the Montreal–Seville fare to bear a fixed relation to the Montreal–Madrid fare, even when both markets share a flight (Belobaba 2006d). Fare products with restrictions are second-degree discrimination. And frequent revision of prices as departure approaches, the subject of Chapter 21, is a coarse step toward the first degree.

17.2.1 One price or two on a linear demand curve

The simplest model shows what is at stake. Suppose that in a market of N potential travellers per departure, willingness to pay is spread uniformly between 0 and a maximum P, so that the number willing to pay at least p is

$$Q(p) = N\left(1 - \frac{p}{P}\right), \qquad 0 \le p \le P. \tag{17.1}$$

With one price and no capacity limit, revenue $pQ(p)$ is maximized at $p^* = P/2$: half the market flies and revenue is $NP/4$. Now let the airline offer several prices $p_1 > p_2 > \cdots > p_n$ and suppose, for the moment, that fences are perfect: everyone willing to pay at least p_1 pays p_1, everyone between p_2 and p_1 pays p_2, and so on. The k-th price then sells $Q(p_k) - Q(p_{k-1})$ tickets, with $Q(p_0) = 0$.

RESULT 17.1 Fenced prices on a linear demand curve

Let demand be (17.1), let each passenger cost $c \geq 0$ to carry and let there be no capacity limit. The n perfectly fenced prices that maximize contribution are equally spaced between P and c,

$$p_k = P - \frac{k}{n+1}(P-c), \qquad k = 1, \dots, n, \tag{17.2}$$

and the maximum contribution is

$$\Pi_n = \frac{n}{n+1} \cdot \frac{N(P-c)^2}{2P}. \tag{17.3}$$

With $c = 0$, one price earns $NP/4$, two prices earn $NP/3$, one third more, and as n grows revenue approaches $NP/2$, the whole area under the demand curve.

Proof. Write $q_k = Q(p_k) - Q(p_{k-1}) = N(p_{k-1} - p_k)/P$ for the tickets sold at price p_k, with $p_0 = P$. Contribution is $\Pi = \sum_k (p_k - c)\, q_k$. The price p_k enters the k-th term and, for $k < n$, the $(k+1)$-th, so

$$\frac{\partial \Pi}{\partial p_k} = \frac{N}{P}\left[(p_{k-1} - p_k) - (p_k - c) + (p_{k+1} - c)\right] = \frac{N}{P}(p_{k-1} - 2p_k + p_{k+1}),$$

where $p_{n+1} \equiv c$ in the last condition. Setting every derivative to zero makes each price the average of its neighbours, so the prices are equally spaced from $p_0 = P$ to $p_{n+1} = c$, which is (17.2). The Hessian is N/P times a tridiagonal matrix with -2 on the diagonal and 1 beside it, which is negative definite, so this is the maximum. Each block then holds $N(P-c)/(P(n+1))$ passengers paying a margin of $(n+1-k)(P-c)/(n+1)$, and summing over k gives (17.3). ∎

The extra revenue has two sources, both visible in Figure 17.1. The high fare extracts part of the surplus that a single price leaves with travellers who would have paid more, and the low fare sells seats to travellers who were priced out. Price discrimination raises output as well as price, which is why it suits an industry whose product perishes at departure.

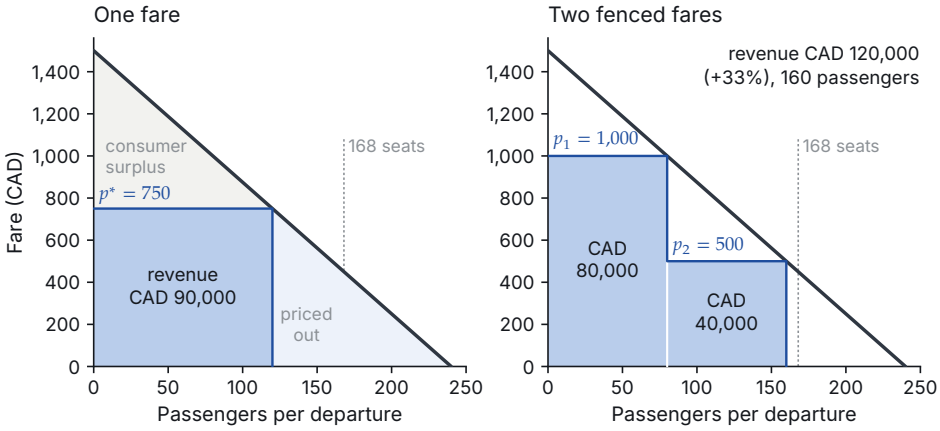


Figure 17.1 Revenue from one fare and from two fenced fares on a linear demand curve with $N = 240$ potential travellers and a maximum willingness to pay of CAD 1,500. One fare leaves surplus with travellers who would have paid more and prices out the rest; a second fare recovers part of both and fills 160 of 168 seats.

WORKED EXAMPLE 17.1 One fare or two

Take a market of $N = 240$ potential travellers per departure with $P =$ CAD 1,500, roughly the scale of a transatlantic leisure market served by one narrowbody (illustrative), and ignore the small cost of carrying a passenger.

With one fare, $p^* = 750$ and $Q = 120$ passengers fly, so revenue is CAD 90,000 and an aircraft with 168 economy seats leaves 71 percent full.

With two fares, (17.2) with $n = 2$ gives $p_1 = 1{,}000$ and $p_2 = 500$. Each fare sells 80 tickets, so 160 passengers fly (95 percent of the seats) and revenue is 80,000 + 40,000 = CAD 120,000, a third more than with one fare.

With three fares, the prices would be 1,125, 750 and 375 with 60 passengers each. That is 180 passengers for 168 seats: the capacity constraint now binds, and the problem changes, as discussed next.

The aircraft and the demand are the same in all three cases. The fare structure alone moves revenue by a third and the load factor from 71 to 95 percent.

When capacity binds, the lowest fare must ration seats. If the unconstrained prices would sell more than C seats, the optimal lowest price is the one that just fills the cabin, $p_n = P(1 - C/N)$. Revenue

can then be written $\sum_{k<n}(p_k - p_n)[Q(p_k) - Q(p_{k-1})] + p_n C$, and maximizing it over the higher prices is the problem of Result 17.1 with $n-1$ prices and a unit cost equal to p_n. For two fares,

$$p_2 = P\left(1 - \frac{C}{N}\right), \qquad p_1 = \frac{P + p_2}{2}, \qquad \text{when } \tfrac{2}{3}N > C. \tag{17.4}$$

With $N = 300$ and $C = 168$, the unconstrained fares of 1,000 and 500 would sell 200 seats. The constrained optimum raises the low fare to $p_2 =$ CAD 660 and sets $p_1 =$ CAD 1,080, with 84 passengers at each fare and revenue of CAD 146,160, against CAD 112,500 for the best single fare (750, which sells 150 seats). The interpretation is the one that runs through Part IV: once seats are scarce, selling a seat at the high fare has an opportunity cost, the low fare the seat would otherwise earn, and the high fare is priced as if each seat cost that much. Chapter 18 turns the same idea into Littlewood's rule for allocating seats, and Chapter 20 into bid prices.

Third-degree discrimination across groups the airline can observe follows the monopolist's inverse elasticity rule. If group s has demand elasticity η_s and carrying a passenger costs c, setting marginal revenue $p_s(1 + 1/\eta_s)$ equal to c gives

$$\frac{p_s - c}{p_s} = -\frac{1}{\eta_s}, \qquad \text{so} \qquad p_s = c\,\frac{\eta_s}{1 + \eta_s} \quad (\eta_s < -1). \tag{17.5}$$

The less elastic group pays more. With $c =$ CAD 250, for instance, a point of sale facing an elasticity of -2.5 would be priced at CAD 417 and one facing -1.5 at CAD 750 (illustrative values; the relevant elasticities are those facing one airline, which are larger in magnitude than the market elasticities quoted earlier). The welfare effect is ambiguous in general: for third-degree discrimination to raise total surplus it must raise total output (Varian 1989), which in airline markets it often does by opening cheap fares to travellers a uniform price would exclude.

17.3 Fences and self-selection

Third-degree discrimination needs a characteristic the airline can see. Willingness to pay is not one: a traveller asking for the lowest fare does not reveal that her employer would have paid four times as much. Second-degree discrimination solves the problem with a menu designed so that each type of traveller prefers the option meant for it. The airline device is the **fence**, a condition attached

to a cheap fare that costs the airline little but costs travellers with a high willingness to pay a lot.

Consider two types of traveller. There are n_T time-sensitive travellers and n_L leisure travellers. A flexible ticket is worth v_i to a traveller of type i; a restricted ticket, for example one that must be bought 21 days ahead and cannot be refunded, is worth $v_i - d_i$, where d_i is the **disutility** of the restriction to type i. The fence works only if $d_T > d_L$: the restriction must hurt the time-sensitive traveller more, the single-crossing condition of contract theory. The airline posts a flexible fare f_1 and a restricted fare f_2 and wants each type to buy its own product. Four constraints must hold:

$$v_L - d_L - f_2 \geq 0, \qquad v_T - f_1 \geq 0,$$
$$f_1 - f_2 \geq d_L, \qquad f_1 - f_2 \leq d_T.$$

The first row are the participation constraints of the two types: each is willing to buy its product. The second row are the incentive-compatibility constraints: the leisure traveller prefers the restricted ticket and the time-sensitive traveller prefers the flexible one.

RESULT 17.2 The optimal fenced menu

Suppose $d_T > d_L$ and $v_T \geq v_L$. Among menus that sell the flexible fare to time-sensitive travellers and the restricted fare to leisure travellers, revenue is maximized by

$$f_2 = v_L - d_L, \qquad f_1 = \min(v_T,\, f_2 + d_T). \tag{17.6}$$

When the incentive constraint of the time-sensitive traveller binds ($f_1 = f_2 + d_T$), the menu earns more than the single price v_L at which everyone buys if and only if

$$n_T\,(d_T - d_L) > n_L\,d_L. \tag{17.7}$$

Proof. If the leisure traveller's participation constraint were slack, both fares could rise by the same small amount without changing any comparison or, if the time-sensitive traveller already pays $f_1 = v_T$, the restricted fare alone could rise (this only loosens her incentive constraint, and the leisure traveller's holds strictly because $f_1 - f_2 > v_T - v_L + d_L \geq d_L$); so it binds: $f_2 = v_L - d_L$. Given f_2, revenue rises with f_1, which can increase until the time-sensitive traveller is indifferent between the two tickets ($f_1 = f_2 + d_T$) or unwilling to fly

($f_1 = v_T$). The leisure traveller's incentive constraint then holds, because $f_1 - f_2 = \min(v_T - v_L + d_L,\, d_T) \geq d_L$. Compared with selling a flexible ticket to everyone at v_L, the menu gains $f_1 - v_L = d_T - d_L$ on each time-sensitive traveller and loses $v_L - f_2 = d_L$ on each leisure traveller, which gives (17.7). ∎

The result carries three lessons. First, a fence works through the difference in disutilities, not their level. Second, the time-sensitive traveller keeps a surplus $v_T - f_1$, an information rent: any higher flexible fare would send her to the restricted ticket. Third, a fence is not free. It lowers what the leisure traveller will pay by d_L, so it pays only if it hurts the high type much more than the low type, weighted by their numbers. Restricted fares are an instance of what Deneckere and McAfee call **damaged goods**: the seller deliberately degrades a cheap version of its product to make the expensive version worth buying (Deneckere and McAfee 1996).

WORKED EXAMPLE 17.2 **Designing a fence for YUL–SVQ**

Suppose a departure attracts $n_T = 40$ time-sensitive travellers who value a flexible ticket at CAD 1,400 and $n_L = 120$ leisure travellers who value it at CAD 550. A 21-day advance purchase with no refunds costs the first group CAD 600 and the second CAD 100 (all values illustrative).

Without a fence, the airline can sell a flexible ticket to everyone at 550, earning $160 \times 550 =$ CAD 88,000, or to the time-sensitive travellers only at 1,400, earning CAD 56,000.

With the fence, (17.6) gives $f_2 = 550 - 100 = 450$ and $f_1 = \min(1{,}400,\, 450 + 600) = 1{,}050$. Revenue is $40 \times 1{,}050 + 120 \times 450 =$ CAD 96,000, which is CAD 8,000 more than the best single price, as (17.7) predicts: $40 \times 500 - 120 \times 100 = 8{,}000$. Each time-sensitive traveller keeps CAD 350 of surplus. Perfect discrimination would earn $40 \times 1{,}400 + 120 \times 550 =$ CAD 122,000.

Now weaken the fence, for instance by allowing changes for a small fee, so that the restriction costs a time-sensitive traveller only CAD 250. The flexible fare can then be at most 700, and the menu earns CAD 82,000, less than the single price. By (17.7) the fence pays only if it costs the time-sensitive travellers more than CAD 400.

The lesson for a planner is that the revenue a route can earn depends on restrictions that never appear in the schedule. A

market whose business travellers can easily meet leisure restrictions supports a much smaller fare premium.

Real travellers do not come in two types. Suppose the time-sensitive travellers' disutility d of the restricted fare has distribution function H and density h, and that all of them value a flexible ticket above any fare considered. With the restricted fare fixed at f_2, a time-sensitive traveller buys the flexible fare when $d > f_1 - f_2$, so expected revenue per time-sensitive traveller is

$$r(f_1) = f_2 + (f_1 - f_2)[1 - H(f_1 - f_2)].$$

The airline is in effect selling flexibility, at the price $\Delta = f_1 - f_2$, to buyers whose valuation of it is d.

RESULT 17.3 The price of flexibility

The revenue-maximizing fare gap $\Delta^* = f_1 - f_2$ satisfies

$$\Delta^* = \frac{1 - H(\Delta^*)}{h(\Delta^*)}. \tag{17.8}$$

If $d \sim \mathcal{N}(m, \sigma_d^2)$, then $\Delta^* = \sigma_d / \lambda(z^*)$, where $z^* = (\Delta^* - m)/\sigma_d$ and $\lambda(z) = \varphi(z)/(1 - \Phi(z))$ is the inverse Mills ratio of Chapter 9.

The proof is one line: maximizing $\Delta[1 - H(\Delta)]$ gives the first-order condition $1 - H(\Delta) - \Delta h(\Delta) = 0$, the monopoly pricing rule with the survival function of d as the demand curve. With a normal distribution, $h(\Delta) = \varphi(z)/\sigma_d$ and $1 - H(\Delta) = 1 - \Phi(z)$. Because the normal hazard rate $\lambda(z)/\sigma_d$ increases with Δ, the condition has a single root, and it is the maximum.

Figure 17.2 shows the trade-off for $f_2 = $ CAD 450 and $\sigma_d = $ CAD 150. With a strong fence ($m = 600$) the best flexible fare is CAD 914, a fare ratio of 2.03; 82 percent of time-sensitive travellers pay it and revenue per time-sensitive traveller is CAD 829. With a moderate fence ($m = 400$) the best flexible fare falls to CAD 765 (ratio 1.70, revenue CAD 675). With a weak fence ($m = 200$) it falls to CAD 644 (ratio 1.43): only 52 percent of time-sensitive travellers pay it, and revenue per traveller is CAD 550. Weaker fences compress the optimal ratio between the highest and lowest fares and cut the revenue from the most valuable travellers. That is the mechanism behind the fare simplification of the 2000s.

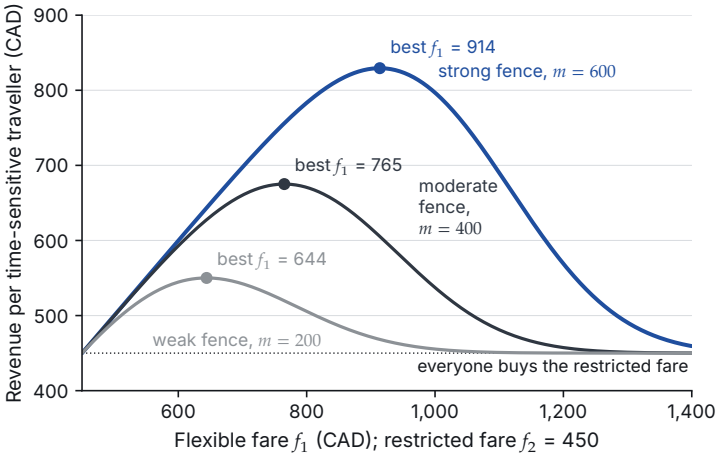


Figure 17.2 Expected revenue per time-sensitive traveller as the flexible fare rises above a restricted fare of CAD 450, when the disutility of the restrictions is normal with mean m and standard deviation CAD 150. Strong fences support high flexible fares; weak fences make travellers buy down as soon as the gap grows.

A fare structure is only as strong as its fences: the premium an airline can charge for flexibility is bounded by how much its customers dislike the restrictions on the cheaper fares.

17.4 Fare products and restrictions

Airlines implement the menu through fare products.

DEFINITION 17.2 Fare product, fare basis and booking class

A **fare product** is a fare level for travel in an origin-destination market together with its rules: when it must be bought, how long the traveller must stay, whether it can be changed or refunded, which flights and sales channels it is valid on, and, in branded structures, which services it includes. Each fare is published under a **fare basis** code and assigned to a **booking class** (reservations booking designator), a one-letter inventory bucket on each flight. Revenue management controls booking classes, not fares: when a class closes on a flight, every fare product mapped to it becomes unavailable on that flight.

The classic restrictions each target a different difference between business and leisure trips.

- **Advance purchase** requires the ticket to be bought a minimum number of days before departure, typically 7, 14 or 21. Leisure trips are planned early and business trips often are not. Advance purchase also orders bookings in time, with the cheap fares booking first, which is the arrival pattern the models of Chapter 18 assume.
- **Minimum stay**, classically a Saturday night away, excludes the business round trip that returns home for the weekend. The reading library describes it as historically the most effective fence (Belobaba 2006d).
- **Refund and change conditions** separate travellers whose plans are certain from those whose plans may change.
- **Round-trip requirements, maximum stays, day and flight restrictions, and channel or point-of-sale conditions** add further dimensions.

Each restriction costs the airline little to impose, and each costs travellers an amount that varies across them; the reading library calls this amount the restriction's **disutility cost** (Belobaba 2006d). It is the d of Section 17.3.

PODS, the MIT Passenger Origin-Destination Simulator used throughout the reading library, turns the idea into a choice rule (Chapter 6 describes it in full) (Belobaba 2006o). Each simulated traveller draws a maximum willingness to pay and a disutility for each restriction, discards the fare products priced above her willingness to pay, and buys the remaining product with the lowest **generalized cost**

$$\mathrm{GC}_{nk} = f_k + \sum_{r \in \mathcal{R}_k} \delta_{nr}, \tag{17.9}$$

where f_k is the fare of product k, $\mathcal{R}_k$ its set of restrictions and δ_{nr} the disutility of restriction r to traveller n. The full model adds terms for schedule inconvenience and path quality. Table 17.1 gives the baseline fare structure of the ePODS classroom game and the mean generalized costs that the PODS team calibrated for an average business and an average leisure traveller.

Table 17.1 The ePODS baseline fare structure and mean generalized costs (historical, adapted from Belobaba, 2006). Mean disutilities, read from the lecture charts and therefore approximate, are about 225 for the Saturday stay and 75 for each of the two penalty restrictions for business travellers, and about 88 and 13 for leisure travellers.

Class	Fare (USD)	Restrictions	Business GC	Leisure GC
Y	350	none	350	350
M	200	7-day advance purchase, Saturday stay	425	288
B	150	14-day advance purchase, Saturday stay, non-refundable	450	251
Q	100	21-day advance purchase, Saturday stay, non-refundable, change fee	475	214

WORKED EXAMPLE 17.3 What the Saturday-night stay was worth

In Table 17.1 the average business traveller ranks the classes Y (350), M (425), B (450), Q (475): she buys the full fare although three cheaper fares are on sale. The average leisure traveller ranks them Q (214), B (251), M (288), Y (350). The fences separate the two groups exactly as intended.

Remove the Saturday-night stay. The business traveller's generalized costs become 350, 200, 225 and 250. M is now her cheapest option in every sense, and her fare drops from 350 to 200, a loss of 43 percent on each business passenger unless revenue management closes M before she books. The leisure traveller still buys Q.

Remove the advance-purchase rules instead, and no generalized cost in the table changes. A business traveller who books late still faces a Saturday stay on every discounted class, so she still buys Y.

This is the pattern of the ePODS experiments in the next section: removing advance purchase alone cost 0.5 percent of revenue, removing the Saturday-night stay 16.8 percent (Belobaba 2006o).

17.5 From restricted to simplified fares

Through the 1990s, network carriers in North America sold economy through heavily restricted structures: a full fare with no conditions, then a cascade of discounts, each with a longer advance purchase, a Saturday-night stay and stiffer penalties (Belobaba 2006d). Low-fare airlines attacked this structure with simple one-way fares, few restrictions and modest ratios between their highest and lowest fares. In the early 2000s, as business travel fell and internet shopping made every fare visible, the restricted structures eroded. America West and then Alaska simplified their fares, and by 2005 Delta had introduced simplified fares in all its US domestic markets, matched selectively by other network carriers (Belobaba 2006o). The simplified structures dropped the Saturday-night stay, kept advance purchase, non-refundability and change fees, and compressed the ratio of the highest to the lowest fare to about 4 to 1 or less in the affected markets. Revenue management still controlled the number of seats at each fare level. Table 17.2 shows an example.

Table 17.2 Delta's simplified coach fares between Boston and Atlanta, April 2005 (historical, adapted from Belobaba, 2006). The ratio of the highest to the lowest fare is 3.3 to 1.

Booking class	One-way fare (USD)	Advance purchase	Minimum stay	Change fee (USD)	Refund
T	124	21 days	none	50	no
U	139	14 days	none	50	no
L	184	7 days	none	50	no
K	209	3 days	none	50	no
B	354	3 days	none	50	no
Y	404	none	none	none	full fare

The consequences were predictable from Section 17.3. With weaker fences, many more travellers bought lower fares: simplified structures produced higher load factors but revenues 10 to 15 percent lower (Belobaba 2006d). Belobaba summarizes the change as a drop in business demand and in the willingness to pay the highest fares, a greater readiness to accept restrictions on low fares, and cuts in the lowest fares to stimulate traffic and answer low-cost carriers; together they lowered total revenue and unit revenue while load factors held steady (Belobaba 2006d).

The PODS team measured the effect of each restriction in a simulated market with two competing airlines, six fare classes with fares from USD 125 to USD 425 (a ratio of 3.4 to 1) and leg-based revenue management with EMSR-b (Chapter 18) at both airlines (Belobaba 2006o). Starting from a fully restricted structure, they removed the restrictions step by step. Figure 17.3 redraws the results.

Three results stand out. Removing advance purchase alone cost almost nothing (0.5 percent), because the Saturday stay still fenced business travellers out of the discount classes. Removing the Saturday stay cost 16.8 percent, removing every restriction except advance purchase 29.6 percent, and removing everything 45 percent, while the load factor rose from 81.6 to 88.1 percent. With no fences at all, nearly the whole load was in the lowest class: about 88 of 88.1 load factor points, against 26 of 81.6 in the fully restricted case (approximate values read from the lecture chart).

The value of revenue management behaved differently. EMSR-b earned 8.2 percent more than first come, first served in the restricted base and 11.8 percent more once the Saturday stay was gone, because protecting seats matters more when fences leak. With everything removed its advantage fell to 0.1 percent. A system that forecasts demand by class from booking history sees almost all demand

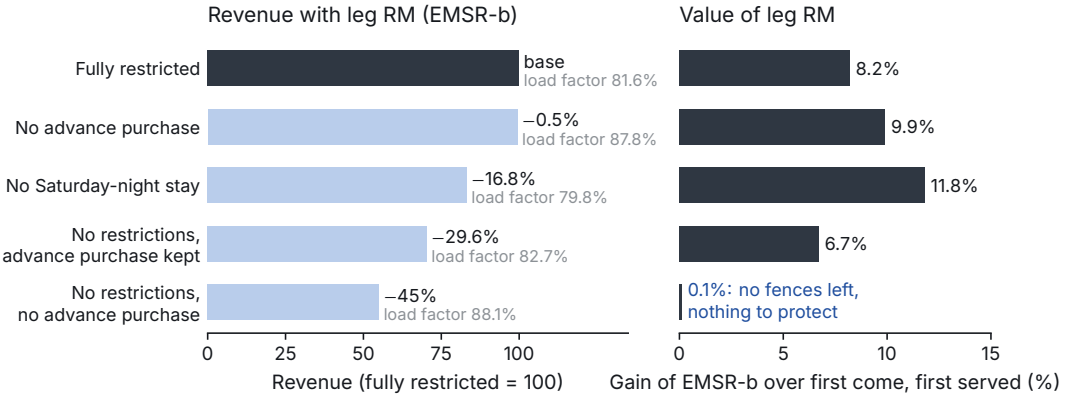


Figure 17.3 The ePODS fare simplification experiments (historical, adapted from Belobaba, 2006). Left: revenue relative to the fully restricted structure, and load factor, as restrictions are removed. Right: the revenue gain of leg revenue management (EMSR-b) over first come, first served.

in the lowest class, protects almost nothing, and sells almost everything at the lowest fare, which confirms its own forecast. This self-reinforcing collapse is **spiral down**, and Chapter 21 shows how forecasting willingness to pay rather than class demand repairs it (Belobaba 2006a).

> **PITFALL**
>
> Reading a higher load factor as a better fare structure. In the ePODS experiments the structure with the highest load factor, 88.1 percent, earned 45 percent less than the fully restricted structure, whose load factor was 81.6 percent. Fences are worth more than fullness.

17.6 Branded fares and bundles

Simplification did not end differential pricing; it moved the fences. From the late 2000s carriers rebuilt product differentiation around branded fare families and unbundled services. Instead of restricting when and how a cheap ticket may be used, the airline removes services from it: checked bags, seat selection, changes, refunds, priority boarding, loyalty earning.

DEFINITION 17.3 Branded fares and ancillaries

A **branded fare** (or fare family) is a named bundle of fare rules and services, such as a basic, a standard and a flexible economy product, sold under the same names across markets. Each brand is offered at one or more price points, each mapped to a booking class, so that revenue management controls the price within a brand while the traveller chooses among brands. An **ancillary** is a service sold separately from the fare, such as a checked bag or an advance seat assignment.

The economics of Section 17.3 apply directly. A basic fare is damaged goods in its purest form: the seat, the flight and the arrival time are identical to those of the standard fare, and the difference is the removal of services that some travellers value much more than others (Deneckere and McAfee 1996). Selling services both in bundles and separately is mixed bundling, which extracts more surplus than pure bundling or pure unbundling when valuations vary across customers (Adams and Yellen 1976). Unbundling also lowers the base fare displayed in price-ranked shopping screens; the Competition Bureau notes that base fares which exclude services travellers usually want, such as a carry-on bag or seat selection, make fares harder to compare (Competition Bureau Canada 2025).

The sums involved are large. IdeaWorksCompany estimates global airline ancillary revenue at USD 157 billion in 2025, 15.7 percent of airline revenue against 9.1 percent in 2016, and reports that about 45 percent of travellers buy only the base fare (IdeaWorksCompany 2025). Air Canada sells economy under five brands, Basic, Standard, Flex, Comfort and Latitude, which differ in baggage, seat selection, change and refund conditions and loyalty earning (Prince of Travel 2026); Comfort was added in 2018 (Travelweek 2018). On 28 July 2026 Air Canada extended the idea to its premium cabins with Premium Economy Basic and Business Class Basic fares in selected international markets: non-refundable, changeable for a fee, with paid seat selection, but with the same seat, meals and priority services (Open Jaw 2026a).

Brands restore some of the separation that simplification removed, but they do not restore the independence of demand across booking classes that classical revenue management assumes. A traveller who finds Standard closed may buy Flex instead (sell-up) or settle for Basic (buy-down), so the demand for each class depends on which

other classes are open. Modern systems forecast that choice explicitly, which is the subject of Chapter 21.

IN PRACTICE Fare structures and pricing in 2026

Three changes shape airline pricing in 2026. Branded fares and paid ancillaries now do much of the fencing that advance purchase and minimum stays once did, as the Air Canada brands above illustrate (Prince of Travel 2026). The 26-letter booking-class system, which limits a carrier to at most 26 booking classes, and so 26 price points, on a flight, is being relaxed: IATA distinguishes dynamic pricing (prices set from the shopping context without filing them in advance), continuous pricing (an unlimited number of price points) and dynamic bundling (offers assembled for each request) (International Air Transport Association 2021). ATPCO, the industry's fare-filing utility, committed in 2022 to help airlines reach 80 percent of offers created dynamically by 2026, up from 6 percent at the time; it reported 23 percent by mid-2023 (ATPCO 2022, 2023), and we found no published figure for 2026. Distribution is moving with it: NDC connections carried 21.6 percent of US agency transactions settled through ARC in May 2026 (Travel Distribution News 2026). MIT simulations found continuous pricing generally revenue positive, by up to 2 percent for a first mover (Szymański, Belobaba, and Papen 2021). On the airline side, a vendor case study states that Air Canada uses PROS real-time dynamic pricing (PROS 2025). None of this repeals the economics of this chapter: continuous prices add price points and brands add fences, but travellers still have to sort themselves.

17.7 Laurentian's YUL–SVQ fare ladder

Laurentian's economy fare structure on YUL–SVQ, used for the Laurentian examples of Chapter 18, is in Table 17.3. It follows the pattern of current branded fares, with three brands, Flex, Standard and Basic, the first two each sold at two price points.

Table 17.3 Laurentian's YUL–SVQ economy fare ladder for the A321XLR economy cabin of 168 seats. Fare: one-way net fare in CAD. AP: advance purchase in days. Changes and refunds: the change fee is CAD 150, the refund fee CAD 200, and the credit option gives a travel credit instead of a refund. Free checked bags: two with Flex, one with Standard, none with Basic. Mean and SD: unconstrained demand per departure, normal and independent by class, so the total standard deviation is the square root of the sum of the variances. All values are illustrative.

Class	Brand	Fare	AP	Changes	Refund	Mean	SD
Y	Flex	1,249	0	free	full	12	5
B	Flex	919	14	free	fee	18	7
M	Standard	709	14	fee	credit	28	11
Q	Standard	549	21	fee	none	44	17
V	Basic	419	21	none	none	64	26
Total						166	34

Each design choice has a reason. Within a brand, the two classes differ in advance purchase and refund conditions, so every class remains a distinct product with its own fences, as the independent-demand models of Chapter 18 require. The fares fall by a nearly constant ratio from one class to the next, between 0.74 and 0.77, and the highest fare is 2.98 times the lowest, a compressed ratio typical of simplified and branded structures. The advance-purchase rules also order demand in time: V and Q can be bought only until 21 days before departure and B and M until 14 days, so the cheapest fares tend to book first.

Total mean demand is 166 passengers per departure, between the shoulder (150) and peak (175) values of the running case (Chapter 12), for 168 economy seats. Each class has a coefficient of variation of about 0.4. If classes are independent, the total has a coefficient of variation of 0.21, below the 0.24 to 0.27 of the running case, because in reality demand in all classes rises and falls together with the season and the economy. Leg revenue management assumes independence, and Chapter 21 examines what happens when demand moves between classes. The demand-weighted average fare is CAD 617. Under the revenue management controls of Chapter 18, Laurentian expects to carry 150.6 passengers at an average fare of CAD 620, the value the running case uses for the route; selling first come, first served it would carry 153.5 passengers at an average of only CAD 585.

To check the fences, Table 17.4 gives illustrative disutilities for the restrictions in the ladder.

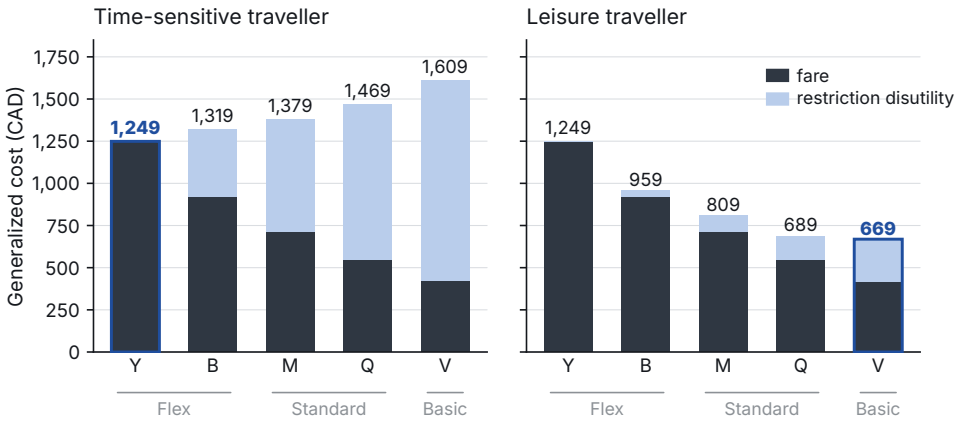


Figure 17.4 Generalized cost of Laurentian's five YUL–SVQ products for an average time-sensitive traveller and an average leisure traveller, using the disutilities of Table 17.4. The outlined bar is each traveller's cheapest option.

Table 17.4 Illustrative mean disutilities (CAD) of the restrictions in Laurentian's fare ladder for an average time-sensitive and an average leisure traveller.

Restriction	Time-sensitive	Leisure
14-day advance purchase	300	20
21-day advance purchase	450	40
refund only with a fee	100	20
travel credit instead of refund	200	40
no refund	300	60
change fee	150	40
no changes	400	80
one free bag instead of two	20	0
no free bag	40	70

Figure 17.4 checks the fences. The average time-sensitive traveller ranks the classes Y, B, M, Q, V and buys Y; the average leisure traveller ranks them V, Q, M, B, Y and buys V. Two details matter for revenue management. B is only CAD 70 behind Y in generalized cost for the time-sensitive traveller, so business travellers who can commit 14 days ahead will buy B; the 14-day rule is a soft fence, and revenue management will need to limit B. And Basic is only CAD 20 cheaper than Q for the average leisure traveller once the checked bag she needs is counted: the bag fee, not the fare, decides who flies Basic. That makes the bag a weak fence between business and leisure travellers but a useful one among leisure travellers.

The ladder is a discrete version of the demand curves of Section 17.2. In Figure 17.5 the classes are stacked in order of fare, and

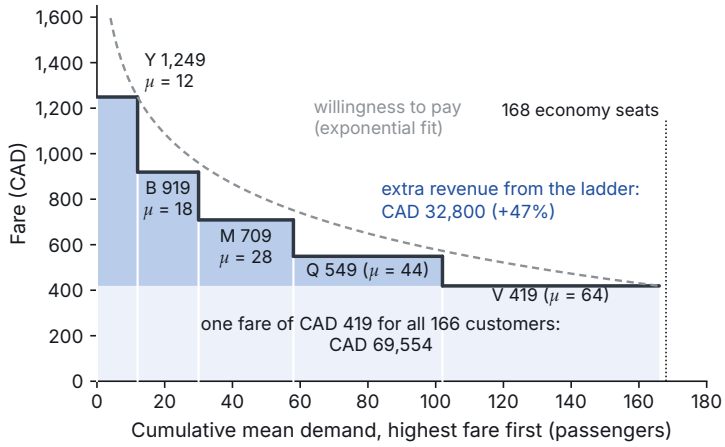


Figure 17.5 Laurentian's YUL–SVQ fare ladder drawn as a staircase: each block is a class's fare times its mean demand. The upper corners trace an exponential willingness-to-pay curve. A single fare of CAD 419 accepted by all 166 customers would earn CAD 69,554; the ladder, with every customer buying the product meant for him, would earn CAD 102,354.

their corners trace a willingness-to-pay curve that is close to exponential, falling by half for every CAD 219 of fare. This is the shape of the willingness-to-pay curves used in PODS (Chapters 6 and 21). If every customer bought the product meant for him, the ladder would earn CAD 102,354 per departure, 47 percent more than the CAD 69,554 of a single fare of CAD 419 that all 166 customers would accept. That difference is the prize the fences protect, and revenue management must defend it against capacity: expected economy revenue under its controls is CAD 93,465 (Chapter 18), enough on its own to cover the avoidable cost of the departure, about CAD 88,000, which revenue at a single fare of CAD 419 would not.

17.8 Pricing, revenue management and competition

Pricing and revenue management divide the commercial work, and the network planner supplies what both of them sell (Table 17.5).

Table 17.5 The division of labour between network planning, pricing and revenue management.

Decision	Made by	Horizon	Instrument
Capacity, schedule and aircraft	Network planning	months to years	the schedule
Fare levels, rules and booking-class mapping	Pricing	months, adjusted daily	filed fares
Seats available to each booking class	Revenue management	daily and per request	booking limits, bid prices

Pricing designs the menu. It decides, market by market and often point of sale by point of sale, which fare products exist, at what levels, under which rules and in which booking class each is sold, and it files the fares through ATPCO, which distributes them to reservation and shopping systems (ATPCO 2022). Revenue management takes the menu and the capacity as given (Belobaba 2006l) and decides which items on the menu are offered on each flight at each moment, by opening and closing booking classes. A traveller sees the lowest fare whose rules she meets among the classes that are open.

The division is not watertight. A pricing decision changes the demand for each class and therefore the best booking limits; a revenue management decision to close the cheap classes raises the effective price. Continuous pricing blurs the line further, because revenue management then sets prices directly (Chapter 21). For the planner the practical consequence is that a fare change and a capacity change must be evaluated together.

Pricing is also set against competitors. Airlines watch one another's fares through the same shopping systems their customers use, and a fare filed by one carrier can be matched by another within hours. The logic is that of price competition between close substitutes: a carrier that does not match a lower fare loses the price-sensitive part of its demand, while a carrier that does gives away revenue on customers who would have paid more. When both carriers match each other downward, both lose. The Competition Bureau found in 2025 that on Canadian routes a single additional competitor reduces average fares by about 9 percent, and that established carriers often answer entrants with lower prices and more seats, a normal competitive response unless it amounts to conduct the Competition Act prohibits (Competition Bureau Canada 2025); Chapter 10 discusses the Canadian market structure. In PODS simulations that Belobaba presented to an IATA revenue management

conference, a network carrier that matched a low-fare competitor's simplified structure lost as much as 8 to 9 percent of revenue by removing advance purchase and 13 percent or more by removing the Saturday-night stay; origin-destination control reduced the loss but did not eliminate it (Belobaba 2002). The same work notes that under leg-based control a matched low fare is open to every itinerary using the leg, whereas origin-destination control keeps it available on empty flights without letting it displace more valuable network passengers (Chapter 20).

WORKED EXAMPLE 17.4 Should Laurentian match?

A competitor's connection through Madrid starts selling YUL–SVQ at CAD 379 with a 21-day advance purchase and no refunds, CAD 40 below Laurentian's V fare. Laurentian estimates that if it does nothing, 30 percent of its V demand will switch: mean V demand falls from 64 to 44.8 passengers. If it matches by refiling V at CAD 379, it keeps its V demand but earns CAD 40 less on every V passenger (illustrative assumptions).

A naive calculation compares V revenue alone: $64 \times 379 = $ CAD 24,256 with matching and $44.8 \times 419 = $ CAD 18,771 without, so matching appears to be worth CAD 5,485 per departure.

The flight, however, is neither empty nor sold first come, first served. Re-solving the booking limits of Chapter 18 in each case gives expected revenues of CAD 91,242 with matching and CAD 91,114 without (both below the CAD 93,465 earned before the competitor arrived): matching is worth CAD 128, not 5,485. Two effects close the gap. Many of the defecting V customers would have been refused anyway, because V's booking limit of 70 seats already caps its sales; V sales fall by 12.2 passengers, not 19.2. And the seats they free are partly taken by Q, M, B and Y travellers whom a nearly full flight would otherwise have spilled.

Table 17.6 repeats the calculation for leaner and busier seasons. Off peak, matching is worth more than CAD 3,000 per departure, because the retained passengers fill seats that would fly empty. At the peak it loses CAD 1,042, because the flight fills anyway and matching only gives CAD 40 away to each V passenger. The naive value moves in the opposite direction.

The value of a fare action depends on how full the flight is, which depends on the capacity the planner assigned. A fare

war on a route where an airline is short of seats is far more expensive than one where it has seats to spare.

Table 17.6 Value per departure of matching a competitor's CAD 379 fare on YUL–SVQ when 30 percent of V demand would otherwise defect. The demand of every class is scaled to the demand factor shown (mean demand divided by 168 seats); expected revenues use EMSR-b booking limits re-solved for each case. Illustrative.

Demand factor	Load factor before the competitor	Value of matching (CAD)	Naive value (CAD)
0.69	69.1%	3,726	3,839
0.79	78.2%	3,425	4,388
0.89	85.3%	1,940	4,936
0.99	89.7%	128	5,485
1.09	91.8%	−1,042	6,033

PITFALL

Valuing a fare change with average fares and empty seats. A pricing action changes which passengers fly, and on a constrained flight each passenger gained displaces another. Before trusting a business case for a fare change, ask revenue management what the seats involved are worth: their displacement cost or bid price (Chapters 20 and 22).

17.9 Summary

- Airlines sell one cabin at many prices because seats are perishable, the cost of carrying one more passenger is small, and travellers differ in willingness to pay and in how much they mind restrictions.
- On a linear demand curve, n perfectly fenced prices are equally spaced and earn $n/(n+1)$ of the area under the curve: two prices earn a third more than one. When capacity binds, the lowest fare rations seats and the higher fares are set as if each seat cost the lowest fare.
- Fare products with restrictions are second-degree price discrimination. A fence works only if it hurts high-willingness-to-pay travellers more than others, and it pays only if $n_T(d_T - d_L) > n_L d_L$.

- The optimal premium for flexibility is the monopoly price of flexibility, $\Delta^* = (1 - H)/h$; weaker fences compress fare ratios and cut revenue from the most valuable travellers.
- Generalized cost, fare plus restriction disutilities, explains which product each traveller buys and why removing the Saturday-night stay cost far more than removing advance purchase.
- Fare simplification in the 2000s raised load factors and cut revenue; with all fences removed, leg revenue management lost almost all its value (historical ePODS results).
- Branded fares and unbundled ancillaries rebuilt the fences around services rather than travel conditions; continuous and dynamic pricing are adding price points in 2026.
- Pricing sets fare levels and rules, revenue management sets availability, and network planning sets capacity. A fare decision is worth what it does to a flight that revenue management is already filling.

Table 17.7 Key formulas of the chapter.

Quantity	Formula
n fenced prices, linear demand	$p_k = P - k(P - c)/(n + 1)$
Maximum contribution	$\Pi_n = \frac{n}{n+1} N(P - c)^2/(2P)$
Two fares with binding capacity	$p_2 = P(1 - C/N)$, $p_1 = (P + p_2)/2$
Inverse elasticity rule	$(p_s - c)/p_s = -1/\eta_s$
Optimal fenced menu	$f_2 = v_L - d_L, f_1 = \min(v_T, f_2 + d_T)$
Fence pays if	$n_T(d_T - d_L) > n_L d_L$
Price of flexibility	$\Delta^* = (1 - H(\Delta^*))/h(\Delta^*)$
Generalized cost	$\mathrm{GC}_{nk} = f_k + \sum_{r \in \mathcal{R}_k} \delta_{nr}$

17.10 Further reading

Belobaba's lectures on airline economics and on pricing challenges (Belobaba 2006d, 2006o) are the source of the ePODS fare structures, restriction disutilities and simplification experiments, and his presentation on willingness to pay (Belobaba 2006a) explains why revenue management had to change when fences weakened. For the theory of price discrimination, Tirole (1988) and Varian (1989) remain the standard treatments; Deneckere and McAfee (1996) explain damaged goods and Adams and Yellen (1976) mixed bundling. Belobaba, Odoni and Barnhart (2015) and Doganis (2019) describe airline pricing practice, Talluri and van Ryzin (2004b) treat

pricing and revenue management in one framework, and the IATA factsheet on dynamic pricing (International Air Transport Association 2021) defines the vocabulary of current offer management. The Competition Bureau's market study (Competition Bureau Canada 2025) is the best public evidence on how Canadian fares respond to competition.

CHAPTER

18

Seat Inventory Control on a Single Flight

IN THIS CHAPTER

- The seat allocation problem, and why revenue rather than yield or load factor is the objective.
- How a revenue management system turns booking data into booking limits, and how often it does so.
- Booking classes, serial nesting, booking limits and protection levels, and the identities that link them.
- Littlewood's two-class rule, derived as a marginal argument and as an optimality condition.
- EMSR-a and EMSR-b, on the lecture's example and on Laurentian's YUL–SVQ flight.
- The dynamic program that defines optimal control: the value of a seat, nested protection levels and how close the heuristics come.
- Booking curves, what revenue management is worth, and what it changes for the planner who supplies the seats.

DRAWS ON Belobaba, Introduction to revenue management: flight leg revenue optimization; Pricing challenges: ePODS and reality; O&D control: what have we learned? (MIT 16.75J, 2006); Littlewood (1972); Belobaba (1987, 1989); Brumelle and McGill (1993); Lee and Hersh (1993); Talluri and van Ryzin (2004)

Chapter 17 gave Laurentian five economy fares on YUL–SVQ, from CAD 419 to CAD 1,249. It did not say how many seats each fare may sell. If the airline accepted bookings in the order they arrive, the cheap fares, which are bought first, would take much of the cabin before the travellers willing to pay CAD 1,249 started looking. If it held back too many seats for those travellers, the aircraft would leave with empty seats that leisure travellers would gladly have bought at CAD 419. Seat inventory control decides, for each flight

and each moment of the booking period, which fares are still for sale.

A network planner rarely sets a booking limit, but every number a planner uses passes through them. The load factor of a flight, its average fare, the spill it suffers and the value of one more seat are all outcomes of the control policy (Chapter 9). This chapter derives the policy from first principles on a single flight leg: the classical heuristics that airlines ran for decades, the dynamic program that defines what is optimal, and the evidence on what the machinery earns. Chapter 19 adds no-shows, Chapter 20 extends the logic to connecting networks, where bid prices replace booking limits, and Chapter 22 shows how planners use the outputs.

18.1 The seat allocation problem

Revenue management takes the fare products and the capacity of a flight as given and decides how many seats to make available to each fare (Belobaba 2006l, 1987). With almost all the cost of a scheduled flight committed in advance, maximizing revenue is the operational route to maximizing profit. The discipline was first called yield management, and the name is misleading. Maximizing yield, revenue per passenger mile, leads to severe limits on cheap fares and low load factors; maximizing load factor leads to cheap seats and low yields. Table 18.1, from the lecture, shows three ways to sell the same 200-seat flight.

Table 18.1 Seats sold by fare class on a 2,100-mile flight with 200 seats under three control philosophies (historical, adapted from Belobaba, 2006).

Fare class	Average fare (USD)	Yield emphasis	Load factor emphasis	Revenue emphasis
Y	420	20	10	17
B	360	23	13	23
H	230	22	14	19
V	180	30	55	37
Q	120	15	68	40
Passengers		110	160	136
Load factor		55%	80%	68%
Revenue (USD)		28,940	30,160	31,250
Average fare (USD)		263	189	230
Yield (cents per RPM)		12.53	8.98	10.94

The revenue-oriented policy has neither the highest yield nor the highest load factor, but it earns 8.0 percent more than the yield-oriented policy and 3.6 percent more than the load-factor policy. Revenue is the objective; yield and load factor are diagnostics. The lecture distinguishes three revenue management techniques: overbooking, which accepts more reservations than seats to offset no-shows (Chapter 19); fare class mix, the subject of this chapter; and origin-destination control, which treats local and connecting passengers differently (Chapter 20) (Belobaba 2006l).

The single-leg problem is stated as follows. A future departure has a booking capacity C in one cabin, here Laurentian's 168 economy seats on the A321XLR. Booking classes $j = 1, \dots, n$ share those seats, numbered from the highest fare, $f_1 > f_2 > \cdots > f_n$. For each class there is a forecast of demand D_j. The airline chooses how many seats each class may sell so as to maximize expected revenue.

18.2 The revenue management system

The problem is solved millions of times. The lecture illustrates the scale with a large US carrier: 2,000 flight legs a day, 10 booking classes and 300 days of bookings before departure make 6 million booking limits in the inventory at any moment (Belobaba 2006l). Only a computer system can keep them current. Figure 18.1 shows its structure.

The **data** layer stores bookings, cancellations and no-shows by flight, booking class and day before departure for every past depar-

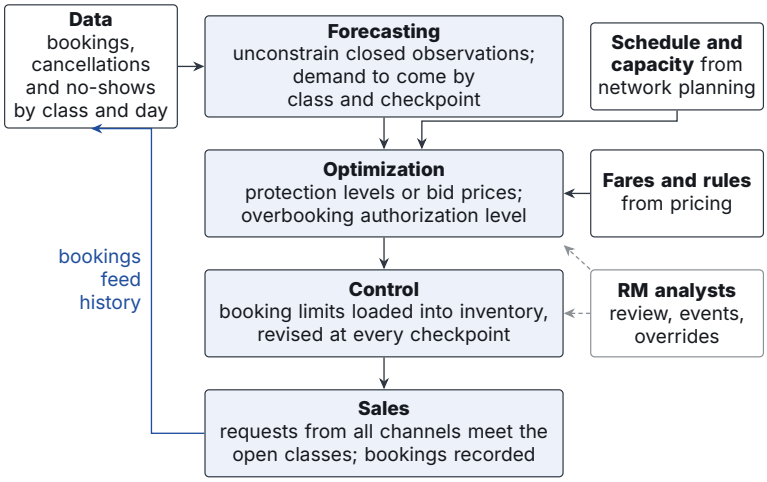


Figure 18.1 The architecture of a revenue management system, after the third-generation systems described by Belobaba (2006). Network planning supplies capacity and pricing supplies fares; bookings made under the controls become the history from which the next forecasts are built.

ture. The **forecasting** layer turns that history into forecasts of demand still to come for each future departure and class. It must first unconstrain the history, because a class that was closed recorded no bookings even when travellers wanted it (Chapter 9), and it often projects bookings to come with pickup methods (Chapter 8). The **optimization** layer computes protection levels for the higher fares and an overbooking authorization level (Chapter 19). The **control** layer loads the resulting booking limits into the airline's inventory, where every sales channel sees them as open or closed classes.

Two features matter as much as the models. The system re-forecasts and re-optimizes many times before departure, at fixed checkpoints, called data collection points, and whenever bookings depart from the forecast, so the availability of a class on one departure can change from one day to the next; the lecture notes that a substantial part of the revenue gain comes from this revision (Belobaba 2006l). And the loop is closed: limits applied today shape the bookings recorded today, which become the history behind tomorrow's forecasts. A system that under-forecasts high-fare demand protects too little, records fewer high-fare bookings and forecasts even less, the spiral down of Chapters 17 and 21. Analysts review the recommendations and intervene for events the history cannot know about.

IN PRACTICE Revenue management systems in 2026

The architecture of Figure 18.1 has survived, but each layer has changed since the reading library was written. In 2006 about a dozen leading airlines were moving from leg control to origin-destination control (Belobaba 2006l), the bid-price methods of Chapter 20. Forecasting increasingly models willingness to pay and sell-up rather than independent class demand; PROS, a major vendor, states that adding willingness-to-pay data to forecasts raises average revenue by about 1 to 3 percent, a vendor estimate (PhocusWire n.d.). Newer entrants sell machine-learning pricing engines and report larger gains, such as the 6 to 9 percent claimed by Fetcherr, which are self-reported without published methods (Boynton and Massy-Beresford 2025). Peer-reviewed evidence on continuous pricing comes mainly from simulation (Szymański, Belobaba, and Papen 2021). Which forecasting and optimization engine Air Canada runs is not public; a vendor case study states only that it uses PROS for dynamic pricing (PROS 2025). For a planner the point is unchanged: the system's outputs are forecasts and opportunity costs by flight and day, and their quality depends on unconstraining and forecasting as much as on the optimizer.

18.3 Booking classes, nesting and protection levels

Each booking class is an inventory bucket identified by a letter, to which the fare products of Chapter 17 are mapped. The way the buckets share the cabin's seats is the first design choice.

DEFINITION 18.1 Partitioned and nested inventory

In a **partitioned** inventory, each booking class has its own allocation of seats, which no other class can use. In a **serially nested** inventory, classes are ranked by fare and every seat available to a class is also available to all higher classes, so a high-fare request is never refused while any seat remains for sale at a lower fare.

The lecture's example shows why every modern system nests (Belobaba 2006l). With allocations of 30 seats to Y, 40 to B and 70 to M in a 140-seat cabin, a partitioned inventory refuses the 31st Y request even if B or M seats are still unsold. A nested inventory sells it a seat that would otherwise have gone to B or M. Nesting changes the

question the airline must answer: not how many seats to give each class, but how many seats to keep away from each lower class.

DEFINITION 18.2 Protection levels and booking limits

With classes numbered from the highest fare and capacity C:

- the **protection level** y_j is the number of seats reserved for classes 1 to j together, which classes $j+1$ to n may not use;
- the **booking limit** b_j is the number of seats that may be sold before class j closes.

They satisfy $y_1 \le y_2 \le \dots \le y_{n-1} \le C$ and

$$b_1 = C, \qquad b_{j+1} = C - y_j, \qquad j = 1, \dots, n-1. \tag{18.1}$$

Class j is open while the seats remaining exceed y_{j-1} (with $y_0 = 0$). The seats that class j can obtain and class $j+1$ cannot number $b_j - b_{j+1} = y_j - y_{j-1}$.

Implementations differ in how a high-fare booking affects the lower classes. If protection levels are applied to the seats remaining, as above, every booking counts against the limits of all lower classes; some systems instead count against b_j only the bookings in class j and below. The two coincide when lower classes book first, the assumption of the classical models (Talluri and van Ryzin 2004b).

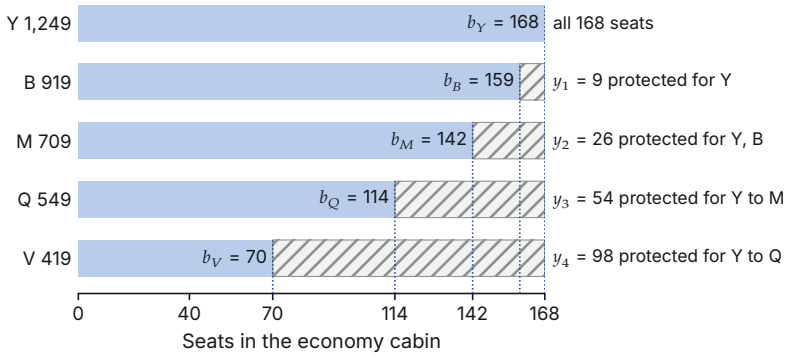


Figure 18.2 Serially nested booking limits and protection levels on Laurentian's YUL–SVQ economy cabin, using the protection levels derived in Section 18.5. Each bar shows the seats a class may sell (its booking limit b) and the seats protected for the classes above it.

WORKED EXAMPLE 18.1 Protection with known demand

The lecture's deterministic example has a 140-seat cabin with known demand of 30 for Y, 40 for B and 85 for M (Belobaba 2006l). The answer is immediate: protect 30 seats for Y, then 40 more for B, and give the remaining 70 to M, spilling 15 M requests. In the notation of Definition 18.2, $y_1 = 30$ and $y_2 = 70$, so the nested booking limits are $b_Y = 140$, $b_B = 140 - 30 = 110$ and $b_M = 140 - 70 = 70$.

With uncertain demand this rule is no longer right. Protecting exactly the forecast of Y demand would leave seats empty whenever Y demand falls short and would protect too few whenever the B fare is low relative to Y. The rest of the chapter finds the right amount.

18.4 Two classes: Littlewood's rule

Start with two classes, a high fare f_1 and a low fare $f_2 < f_1$, and assume that low-fare demand books first, as advance purchase rules encourage. The airline chooses a protection level y for class 1: class 2 may buy at most $C - y$ seats, and class 1 can use whatever is left. Demand D_1 for class 1 is uncertain.

Consider the y-th protected seat. If the airline sells it to class 2 now, it earns f_2, provided low-fare demand is there to take it. If it protects

the seat, it earns f_1 only if high-fare demand turns out to be at least y, which happens with probability $\Pr(D_1 \geq y)$. Protecting the seat is worthwhile when $f_1 \Pr(D_1 \geq y) \geq f_2$. As y grows, the probability falls, so there is a last seat worth protecting.

RESULT 18.1 Littlewood's rule

With two classes booking low before high and independent demands, the optimal protection level for class 1 is

$$y^* = \max\{y : f_1 \Pr(D_1 \geq y) \geq f_2\}, \tag{18.2}$$

and the booking limit of class 2 is $C - y^*$. For continuous demand $\Pr(D_1 > y^*) = f_2/f_1$, so that for normal demand

$$y^* = \mu_1 + \sigma_1 \, \Phi^{-1}\left(1 - \frac{f_2}{f_1}\right).$$

The protection level depends neither on class 2 demand nor on capacity, as long as $y^* \leq C$ (Littlewood 1972).

Proof. Let $R(y)$ be expected revenue with protection level y. Raising y by one changes the outcome only when class 2 would have used the extra seat, that is, when $D_2 \geq C - y$. In that event the airline loses f_2 on one low-fare sale, and class 1, which now has $y+1$ seats instead of y, gains f_1 if $D_1 \geq y + 1$. By independence,

$$R(y+1) - R(y) = \Pr(D_2 \geq C - y)\,[f_1 \Pr(D_1 \geq y + 1) - f_2].$$

The bracket decreases in y, so R rises while it is positive and falls afterwards; the largest y for which the step into y was worthwhile is (18.2). The first factor is nonnegative and does not affect the sign, which is why the answer is independent of class 2 demand and of C. ∎

The rule is a newsvendor condition, the same one that sized the aircraft in Chapter 9. Protecting one seat too few costs $f_1 - f_2$ when a high-fare traveller then finds the seat sold at the low fare; protecting one seat too many costs f_2 when the seat leaves empty. The optimal protection balances the two at $\Pr(D_1 \leq y^*) = (f_1 - f_2)/f_1$. A consequence is worth remembering: when the low fare is more than half the high fare, $\Phi^{-1}(1 - f_2/f_1) < 0$ and the airline protects fewer seats than mean high-fare demand; when it is less than half, more.

With demand counted in whole passengers, $\Pr(D_1 \geq y)$ is evaluated with a continuity correction, $1 - \Phi((y - 0.5 - \mu_1)/\sigma_1)$, which

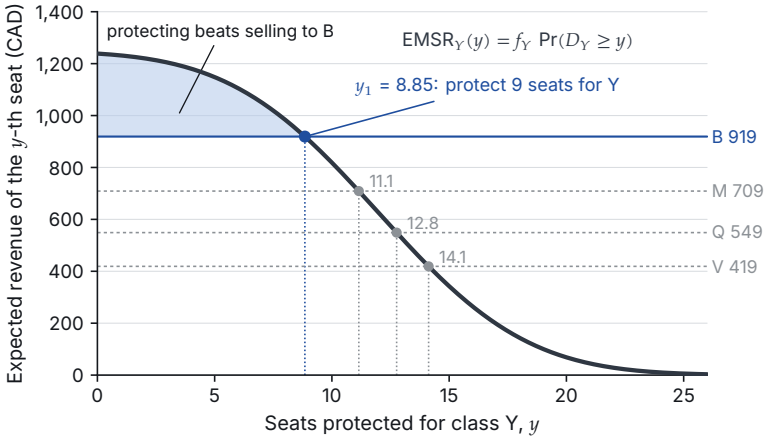


Figure 18.3 Littlewood's rule for class Y on Laurentian's YUL–SVQ flight. The curve is the expected revenue of protecting the y-th seat for Y. Where it crosses the B fare, protecting another seat stops paying. Crossings with the cheaper fares (grey) give the protection Y would need against each of them.

amounts to rounding the continuous solution to the nearest seat. The lecture's worked example uses the normal tail without the correction and rounds down; the difference is at most one seat.

WORKED EXAMPLE 18.2 Protecting Flex seats on YUL–SVQ

Class Y sells at CAD 1,249 with demand $\mathcal{N}(12, 5^2)$, and the next class, B, sells at CAD 919 (Table 17.3). The fare ratio is $919/1{,}249 = 0.736$, so the airline protects seats while $\Pr(D_Y \geq y) \geq 0.736$. With $\Phi^{-1}(1 - 0.736) = -0.630$,

$$y^* = 12 - 0.630 \times 5 = 8.85.$$

In whole seats, the ninth protected seat is expected to earn $1{,}249 \times \Pr(D_Y \geq 9) = 947$, more than the B fare, and the tenth only 864, less. Laurentian protects 9 seats for Y, and B's booking limit is $168 - 9 = 159$.

The airline protects fewer seats than mean Y demand because B pays 74 percent of the Y fare: the tenth seat would be bought at the Y fare only with probability 0.69, and an expected CAD 864 is worth less than a sure CAD 919. Against cheaper alternatives the same rule protects more: 11.15 seats against M, 12.76 against Q and 14.12 against V (Figure 18.3). Capacity does not enter. A

larger aircraft would still protect 9 seats for Y and give the extra seats to the cheaper classes.

PITFALL

Protecting the forecast. Holding back exactly the mean demand of the high fare is right only when the low fare is half the high fare. On YUL–SVQ it would protect 12 Y seats instead of 9, and on the whole cabin it costs revenue (Table 18.4).

18.5 Many classes: EMSR-a and EMSR-b

With more than two classes, protection must be set for each cut between the higher and lower classes. Belobaba's expected marginal seat revenue (EMSR) heuristics apply Littlewood's rule at every cut (Belobaba 1987, 1989).

DEFINITION 18.3 Expected marginal seat revenue

The **expected marginal seat revenue** of class j at y seats, $\mathrm{EMSR}_j(y) = f_j \Pr(D_j \geq y)$, is the expected revenue from making a y-th seat available to class j.

The two variants differ in how they treat the several classes above a cut. **EMSR-a** protects each higher class separately against the lower class and adds the results. **EMSR-b**, the later refinement that became the standard leg heuristic (Talluri and van Ryzin 2004b; Belobaba 2006l), merges the classes above the cut into one artificial class whose fare is their demand-weighted average and whose demand is the sum of their demands.

RESULT 18.2 EMSR-a and EMSR-b protection levels

With independent normal demands $D_k \sim \mathcal{N}(\mu_k, \sigma_k^2)$, the protection level for classes $1, \dots, j$ against class $j+1$ is, under EMSR-a,

$$y_j^a = \sum_{k=1}^{j} \left[\mu_k + \sigma_k \, \Phi^{-1}\!\left(1 - \frac{f_{j+1}}{f_k}\right) \right]^+, \tag{18.3}$$

and under EMSR-b,

$$y_j^b = \mu_{1:j} + \sigma_{1:j} \, \Phi^{-1}\!\left(1 - \frac{f_{j+1}}{\bar{f}_{1:j}}\right), \tag{18.4}$$

where the merged class has

$$\mu_{1:j} = \sum_{k \le j} \mu_k, \qquad \sigma_{1:j} = \Big(\sum_{k \le j} \sigma_k^2\Big)^{1/2}, \qquad \bar{f}_{1:j} = \frac{1}{\mu_{1:j}} \sum_{k \le j} f_k \mu_k .$$

The booking limits follow from (18.1).

Both are exact for two classes and heuristic beyond. EMSR-a adds protection levels computed class by class, so it ignores the fact that the uncertainties of several classes partly cancel when they are pooled: the standard deviation of a sum is less than the sum of the standard deviations. When protection levels lie below the means, as they do when fare ratios exceed one half, EMSR-a therefore protects too few seats, and when they lie above the means, too many. EMSR-b pools the uncertainty but treats a mixture of fares as a single average fare.

WORKED EXAMPLE 18.3 The lecture's EMSR-b example

The lecture works through four classes with the normal demands and fares of Table 18.2 (Belobaba 2006l).

Protect Y from B. The condition is $1{,}000 \Pr(D_Y \ge \pi) \ge 700$, that is $\Pr(D_Y \ge \pi) \ge 0.70$. With $D_Y \sim \mathcal{N}(10, 3^2)$, the probabilities at $\pi = 7$, 8 and 9 are 0.841, 0.747 and 0.63, so the largest integer that satisfies the condition is 8: protect 8 seats for Y.

Protect Y and B from M. The combined class has mean $10 + 15 = 25$, standard deviation $\sqrt{9 + 25} = 5.83$ and average fare $(10 \times 1{,}000 + 15 \times 700)/25 = 820$. The condition is 820 $\Pr(D_{YB} \geq \pi) \geq 500$, or $\Pr(D_{YB} \geq \pi) \geq 0.6098$. The probabilities at 20, 22, 23 and 24 are 0.805, 0.697, 0.633 and 0.5675, so 23 seats are protected jointly for Y and B.

With an authorized capacity of 80 seats, the booking limits are 80 for Y, $80-8 = 72$ for B and $80-23 = 57$ for M. Completing the example, which the lecture stops short of: Y, B and M together have mean 45, standard deviation $\sqrt{83} = 9.11$ and average fare $30{,}500/45 = 677.8$, and the threshold is $350/677.8 = 0.5164$. Since $\Pr(D_{YBM} \geq 44) = 0.544$ and $\Pr(D_{YBM} \geq 45) = 0.500$, 44 seats are protected and Q's booking limit is $80 - 44 = 36$.

EMSR-a protects 8, 22 and 42 seats: against M it adds 10 seats for Y and 12.2 for B, and against Q it adds 11.2, 15.0 and 16.3. Solving the problem exactly with the dynamic program of Section 18.6, with the same normal demands rounded to whole passengers, gives optimal protection levels of 8, 23 and 45 and expected revenue of USD 38,819.5. EMSR-b, as computed above, earns USD 38,817.6, 0.005 percent less; EMSR-a earns USD 38,791.8, 0.07 percent less; selling first come, first served earns USD 37,999.4, 2.1 percent less.

Table 18.2 The lecture's EMSR-b example (historical, adapted from Belobaba, 2006), completed for class Q. Mean and SD of demand by class; fares in USD. The EMSR-b and EMSR-a columns give the protection levels for the class and those above it; booking limits use EMSR-b with an authorized capacity of 80 seats.

Class	Mean	SD	Fare	EMSR-b	Booking limit	EMSR-a
Y	10	3	1,000	8	80	8
B	15	5	700	23	72	22
M	20	7	500	44	57	42
Q	30	10	350		36	

The same steps on Laurentian's ladder give Table 18.3 and Figure 18.4.

Table 18.3 EMSR-b and EMSR-a protection levels for Laurentian's YUL–SVQ economy cabin of 168 seats. Classes: the classes protected together, with the mean and standard deviation of their total demand and their demand-weighted average fare $\bar{f}$ (CAD). Next: the next lower class and its fare. Limit: the booking limit of that class, 168 minus the EMSR-b protection level rounded to the nearest seat (9, 26, 54, 98).

Classes	Mean	SD	$\bar{f}$	Next	EMSR-b	EMSR-a	Limit
Y	12	5.00	1,249	B 919	8.85	8.85	159
Y, B	30	8.60	1,051	M 709	26.11	23.94	142
Y to M	58	13.96	886	Q 549	53.74	48.75	114
Y to Q	102	22.00	741	V 419	98.36	90.19	70

WORKED EXAMPLE 18.4 EMSR-b on YUL–SVQ

Take the last cut, which decides how many seats V may sell. Classes Y, B, M and Q have mean demand $12 + 18 + 28 + 44 = 102$ and standard deviation $\sqrt{25 + 49 + 121 + 289} = 22.0$. Their demand-weighted average fare is

$$\bar{f}_{1:4} = \frac{12 \times 1{,}249 + 18 \times 919 + 28 \times 709 + 44 \times 549}{102} = \frac{75{,}538}{102} = 740.6.$$

The ratio of the V fare to this average is $419/740.6 = 0.5658$, and $\Phi^{-1}(1 - 0.5658) = -0.1656$, so

$$y_4 = 102 - 0.1656 \times 22.0 = 98.36.$$

Laurentian protects 98 seats for the four upper classes, and V may sell at most $168 - 98 = 70$ seats. The other cuts, in Table 18.3, protect 9 seats for Y, 26 for Y and B, and 54 for Y, B and M, so the nested booking limits are 168, 159, 142, 114 and 70 (Figure 18.2).

Mean V demand is 64, so the limit of 70 binds only on departures with strong leisure demand, about 40 percent of them if V demand is normal with standard deviation 26. EMSR-a protects only 90 seats against V: it adds four class-by-class protection levels without pooling their uncertainty, and the one for Q, whose fare is only a third above V's, lies 12 seats below Q's mean demand. The next section shows that EMSR-b's levels are in fact optimal here.

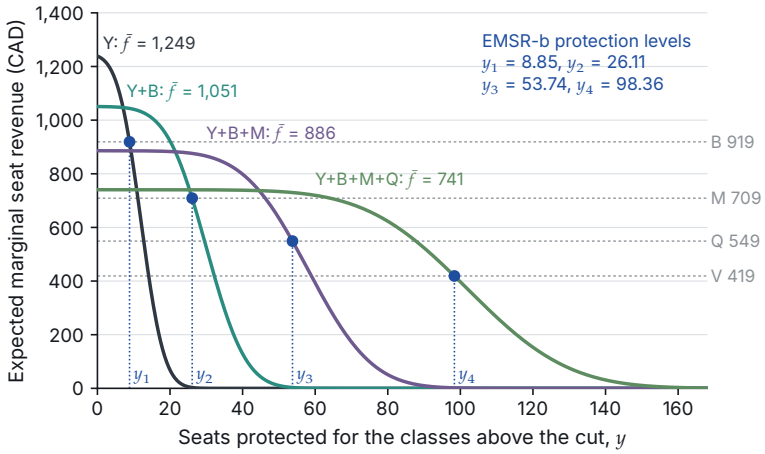


Figure 18.4 EMSR-b on Laurentian's YUL–SVQ flight. Each curve is the expected revenue of protecting a y-th seat for the classes above one cut, valued at their demand-weighted average fare; it crosses the next lower fare at the protection level.

18.6 Dynamic programming and the optimality of nested protection levels

The heuristics are built from a marginal argument. Dynamic programming makes the argument exact by defining the value of a seat.

The static model Keep the classical assumptions: independent demands, and classes booking in blocks from the lowest fare to the highest. Let $V_j(x)$ be the maximum expected revenue from classes $j, j-1, \dots, 1$ when x seats remain and class j is about to book. When class j's demand D_j arrives, the airline accepts $u \le \min(D_j, x)$ requests, earns $f_j u$ and moves on with $x - u$ seats:

$$V_j(x) = \mathbb{E}\Big[\max_{0 \le u \le \min(D_j, x)} \{f_j u + V_{j-1}(x-u)\}\Big], \qquad V_0(x) = 0. \quad (18.5)$$

The **marginal value of a seat**, $\Delta V_j(x) = V_j(x) - V_j(x-1)$, is the expected revenue that the x-th remaining seat will earn from classes 1 to j.

RESULT 18.3 Optimal nested protection levels

V_j is concave in x, and $\Delta V_j(x)$ is nondecreasing in j. When class $j+1$ books, it is optimal to accept its requests as long as more than y_j^* seats remain, where

$$y_j^* = \max\{x : \Delta V_j(x) > f_{j+1}\}, \tag{18.6}$$

and the protection levels are nested, $y_1^* \le y_2^* \le \dots \le y_{n-1}^*$ (Brumelle and McGill 1993; Talluri and van Ryzin 2004b).

The argument runs as follows. Given that V_j is concave, accepting one more class-$(j+1)$ request is worthwhile exactly when f_{j+1} is at least the marginal value of the seat it consumes, and because ΔV_j decreases in x, that holds as long as more than y_j^* seats remain. The policy is therefore a protection level. Showing that V_{j+1} inherits concavity, and that an additional class to come cannot lower the value of a seat, completes an induction; Talluri and van Ryzin give the details (Talluri and van Ryzin 2004b). Since the fares fall with j and the marginal values rise, the protection levels are nested.

Two observations tie this to the heuristics. For $j = 1$, $\Delta V_1(y) = f_1 \Pr(D_1 \ge y)$ exactly, so the first protection level is Littlewood's. For larger j, EMSR-b approximates $\Delta V_j(y)$ by $\bar{f}_{1:j} \Pr(D_1 + \dots + D_j \ge y)$. On Laurentian's flight the approximation is accurate where it matters: the exact marginal value of the 98th seat for classes Y to Q is CAD 423.5 and the EMSR-b value is CAD 430.3, both just above the V fare. Ten seats away from the protection levels the two can differ by up to CAD 100, but there the decision is not in doubt. The optimal protection levels are 9, 26, 54 and 98, exactly the rounded EMSR-b levels.

The dynamic model Real requests do not arrive in tidy blocks: Q and V book during the same months, and a Y request can come at any time. Lee and Hersh formulated the problem over the booking period itself (Lee and Hersh 1993). Divide the booking period into T short periods, counted by the number remaining, so short that at most one request arrives in each. In period t a request for class j arrives with probability $\lambda_j(t)$, with $\sum_j \lambda_j(t) \le 1$, independently of the past. Let $V_t(x)$ be the maximum expected revenue from the last t periods with x seats left, with $V_0(x) = 0$ and $V_t(0) = 0$. Accepting a class-j request earns f_j and leaves $x-1$ seats; rejecting it leaves x. Writing $\Delta V_t(x) = V_t(x) - V_t(x-1)$, the Bellman equation becomes

$$V_t(x) = V_{t-1}(x) + \sum_{j=1}^{n} \lambda_j(t)\,\left[f_j - \Delta V_{t-1}(x)\right]^+. \tag{18.7}$$

The value of the future is what it would be with no request, plus, for each class, the probability of a request times the gain from accepting it when the fare exceeds the value of the seat it uses.

DEFINITION 18.4 Marginal value of a seat

The **marginal value of a seat** $\Delta V_t(x)$ is the expected revenue lost if one of the x remaining seats is taken away with t periods to go. It is the opportunity cost of selling a seat now, the single-leg version of the bid price of Chapter 20.

RESULT 18.4 Optimal dynamic control

In the dynamic model:

1. It is optimal to accept a class-j request in period t with $x \geq 1$ seats left if and only if $f_j \geq \Delta V_{t-1}(x)$.
2. $\Delta V_t(x)$ is nonincreasing in x and nondecreasing in t.
3. Consequently, the optimal policy is a set of nested protection levels

   $$y_j(t) = \max\{x : \Delta V_{t-1}(x) > f_{j+1}\}, \qquad y_1(t) \leq \ldots \leq y_{n-1}(t),$$

   each nondecreasing in the time remaining: the earlier it is, the more seats are protected.

Proof. Part 1 is the comparison inside (18.7). For part 2, let $k(u) = \sum_j \lambda_j(t)\left[f_j - u\right]^+$, a convex, nonincreasing function whose slope lies between -1 and 0 because $\sum_j \lambda_j(t) \leq 1$. Set $\Delta V_{t-1}(0) = +\infty$ and assume, as induction hypothesis, that ΔV_{t-1} is nonincreasing; it holds at $t = 1$, where $\Delta V_0 \equiv 0$. Subtracting (18.7) at $x - 1$ from the same equation at x gives, with $a = \Delta V_{t-1}(x)$ and $b = \Delta V_{t-1}(x-1) \geq a$,

$$\Delta V_t(x) = a + k(a) - k(b).$$

Since $k(a) \geq k(b)$, $\Delta V_t(x) \geq \Delta V_{t-1}(x)$: marginal values rise with time to go. With $c = \Delta V_{t-1}(x+1) \leq a$,

$$\Delta V_t(x) - \Delta V_t(x+1) = \left[(a-c) - (k(c) - k(a))\right] + \left[k(a) - k(b)\right] \geq 0,$$

because the slope bound gives $k(c) - k(a) \le a - c$. So ΔV_t is non-increasing in x, completing the induction. Part 3 follows: since $\Delta V_{t-1}(x)$ decreases in x, class $j+1$ is refused exactly when $x \le y_j(t)$; lower fares give larger protection levels; and higher marginal values earlier in the horizon give larger protection levels. ∎

The result says that optimal control on a single leg is nested protection levels that move over time, the structure that EMSR-b reproduces, approximately, when it is re-solved at every checkpoint with the demand still to come.

WORKED EXAMPLE 18.5 Optimal control of YUL–SVQ

To solve (18.7) for Laurentian's flight we need a booking process. Assume that requests arrive over the last 210 days before departure, each class only until its advance-purchase cutoff (21 days for Q and V, 14 for B and M), with a smooth booking-time profile peaking earliest for V and latest for Y; the expected numbers of requests equal the means of Table 17.3. Half the requests have arrived 70 days before departure and 89 percent 30 days before. With 24 periods per day, $T = 5{,}040$ and no period has a request probability above 0.069. Requests are Poisson, so each class's demand has a variance equal to its mean: V's standard deviation is 8 rather than 26. This model is less uncertain than the static one, which is the price of a tractable Markov model; the comparison below is made within each model.

The program takes about a second to solve. Expected revenue under optimal control is CAD 99,537 per departure. At the opening of sales, the 168th seat is worth CAD 268, the 100th CAD 490, the 40th CAD 709 and the 10th CAD 1,012. The optimal protection levels for Y, Y and B, Y to M, and Y to Q are 15, 38, 74 and 127 seats at day 210, 14, 32, 48 and 60 at day 60, and 10, 15, 18 and 20 at day 30 (Figure 18.5). Early in the horizon the program protects more than EMSR-b computed with the same Poisson demand (10, 27, 56, 100), because it knows that a V request refused today will be replaced by another tomorrow: low-fare demand keeps arriving for months.

Table 18.4 compares the policies. In the static model EMSR-b is optimal, EMSR-a loses 0.18 percent, protecting mean demand loses 0.23 percent and first come, first served loses 3.9 percent. In the dynamic model, EMSR-b re-solved at 16 checkpoints comes within 0.43 percent of the optimum. Computed once

at day 210 and never revised, the same EMSR-b limits lose 8.2 percent, more than having no control at all. The reason is that the static calculation assumes that all V demand books before any Q demand. In fact all classes book during the same months, so Q and higher-class bookings use up part of the 68 seats that the 100-seat protection leaves to V. V closes for good once 68 seats are sold, while about a third of its requests are still to come, and the aircraft leaves with 141 passengers.

Table 18.4 Expected revenue per departure of Laurentian's YUL–SVQ economy cabin under alternative controls. Static model: independent normal demands of Table 17.3, classes booking from the lowest fare up. Dynamic model: Poisson requests over 210 days with advance-purchase cutoffs. Differences are measured within each model.

Policy	Protection levels	Revenue (CAD)	Versus optimal	Load	Average fare (CAD)
Static model					
First come, first served	none	89,831	−3.89%	153.5	585
Protect mean demand	12, 30, 58, 102	93,254	−0.23%	149.1	626
EMSR-a	9, 24, 49, 90	93,298	−0.18%	151.9	614
EMSR-b	9, 26, 54, 98	93,465	0.00%	150.6	620
Optimal (dynamic program)	9, 26, 54, 98	93,465		150.6	620
Dynamic model					
First come, first served	none	97,710	−1.84%	161.9	603
EMSR-b, set once at day 210	10, 27, 56, 100	91,351	−8.22%	140.9	648
EMSR-b, revised at 16 checkpoints	revised	99,105	−0.43%	159.1	623
Optimal (dynamic program)	time-varying	99,537		160.9	619

Figure 18.5 shows the two monotonicity properties. In the left panel each curve falls with the seats remaining, with flat steps at

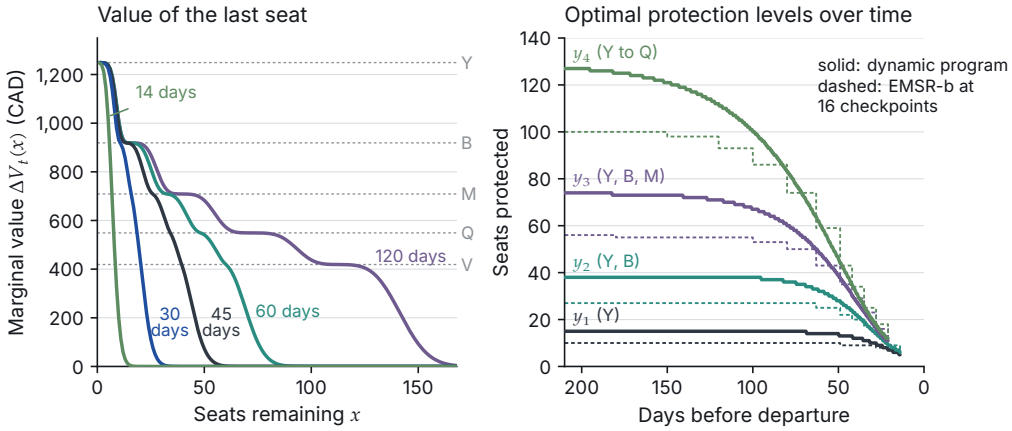


Figure 18.5 The dynamic program on YUL–SVQ. Left: the marginal value of a seat against the seats remaining, at five points of the booking period; a class is open when its fare is at least the marginal value. Right: the optimal protection levels over the booking period (solid) and the EMSR-b protection levels re-solved at 16 checkpoints (dashed).

the fares: an extra seat is worth roughly the fare of the cheapest request it would let the airline accept. As departure approaches the curves shift left, because fewer requests remain to compete for each seat. In the right panel the protection levels fall over time. At every checkpoint EMSR-b protects no more than the program, and early in the horizon much less, because it cannot see that low-fare demand will keep coming; the gap at the checkpoints closes over the last seven weeks and is zero from three weeks out. Between checkpoints EMSR-b's levels stay fixed while the program's keep falling, so there EMSR-b can protect more, as in the next example.

> **PITFALL**
>
> Computing limits once. Protection levels are about demand still to come. Limits that are not revised as bookings arrive assume an arrival order that real booking processes do not follow, and in the example above they do worse than no control. Ask how often a system re-optimizes before trusting its limits.

18.7 Booking curves and controls over time

A **booking curve** plots the cumulative bookings of a departure, or of a class, against days before departure. It is the basic object of revenue management forecasting: pickup methods forecast the bookings still to come from the bookings already on hand at each data

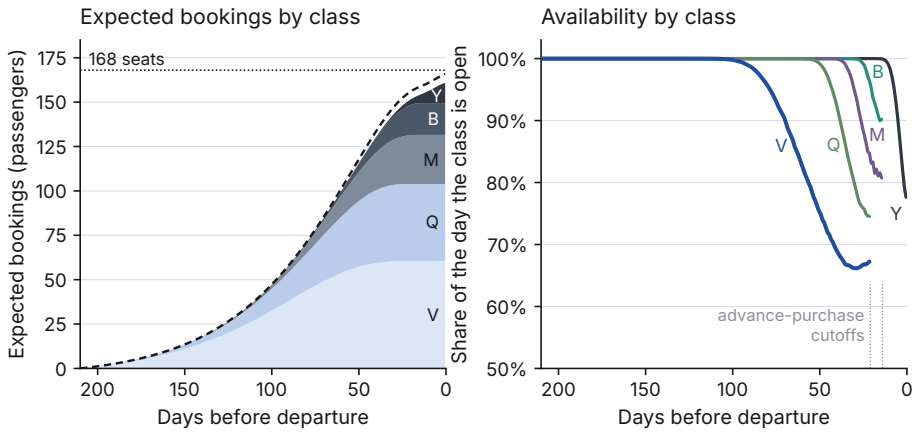


Figure 18.6 Booking curves on YUL–SVQ under optimal control. Left: expected cumulative bookings by class; the dashed line is the expected number of requests, the unconstrained booking curve. Right: share of each day during which each class is open; curves end at the class's advance-purchase cutoff.

collection point (Chapter 8), and the shape of the curve tells the system whether a departure is running ahead of or behind its usual pace.

Figure 18.6 shows the booking process of Laurentian's flight under optimal control. Expected bookings reach 100 passengers 60 days before departure, 143 at 30 days and 161 at departure. The cheap classes book first, as their advance-purchase rules require. Their availability falls as the flight fills: V is open on 79 percent of departures 60 days out and on about 65 percent when its sale ends 21 days out; Q falls to about 72 percent and M and B, which sell until 14 days out, to about 83 and 90 percent. Y stays open on 95 percent of departures a week before departure. Under first come, first served every class stays open until the flight is full, which happens on 45 percent of departures, against 24 percent under optimal control.

Controls do not only close. When cancellations return seats, or when bookings run behind the curve, the marginal value of a seat falls and a closed class can reopen at the next re-optimization. Travellers and travel agents see this as fares that go up and sometimes come back down; the system sees it as protection levels responding to the demand still to come.

WORKED EXAMPLE 18.6 Reading the controls 45 days out

Forty-five days before departure, expected bookings are 123, so a typical departure has 45 seats left. The expected requests still to come are 11.6 for Y, 11.6 for B, 9.2 for M, 4.6 for Q and 2.9 for V.

On a departure with 50 seats left, the marginal value of a seat is CAD 71 and every class is open. With 40 seats left it is CAD 402, still below the V fare: all classes remain open. With 30 seats left it is CAD 646, so Q and V close while Y, B and M stay open; with 32 requests expected in those three classes, 30 seats are not too many to keep for them.

EMSR-b re-solved at the 49-day checkpoint protects 9, 22, 35 and 43 seats. It agrees with the program at 50 and at 30 seats left, but with 40 seats left it closes V (40 is not more than 43) while the program still sells it. The two policies differ only near the boundaries, and such differences are where the 0.43 percent gap of Table 18.4 comes from.

18.8 What revenue management is worth

The rules of this chapter have a long history. Littlewood proposed the two-class rule at BOAC in 1972 (Littlewood 1972). Belobaba introduced expected marginal seat revenue in the late 1980s (Belobaba 1987, 1989) and in the early 1990s proposed the EMSR-b variant (Talluri and van Ryzin 2004b). Brumelle and McGill characterized optimal nested protection levels (Brumelle and McGill 1993), and Lee and Hersh formulated the dynamic model (Lee and Hersh 1993); McGill and van Ryzin survey the development (McGill and van Ryzin 1999).

The gains were large from the start. American Airlines estimated that yield management had generated a quantifiable benefit of USD 1.4 billion over three years (Smith, Leimkuhler, and Darrow 1992). Belobaba put the gain from effective leg-based fare class control and overbooking at 4 to 6 percent of total revenue, and PODS simulations showed origin-destination control adding another 1 to 2 percent, more at high load factors (Belobaba 2002); in the restricted ePODS market of Chapter 17, leg control with EMSR-b earned 8.2 percent more than first come, first served (Belobaba 2006o). These figures are historical and come from airline estimates and simulations rather than controlled experiments.

The models of this chapter show why the gain depends so much on circumstances. Table 18.5 scales all of Laurentian's class demands up and down and reports the gain of optimal control over first come, first served.

Table 18.5 Revenue gain of seat inventory control on YUL–SVQ as mean demand varies, all class demands scaled together. The demand factor is mean demand divided by 168 seats.

Demand factor	0.79	0.89	0.99	1.09	1.19
Static model: EMSR-b over first come, first served	0.4%	1.5%	4.0%	7.8%	12.6%
Dynamic model: optimal over first come, first served	0.0%	0.1%	1.9%	7.8%	15.6%
Load factor, first come, first served (dynamic)	79.0%	88.7%	96.4%	99.5%	100.0%
Load factor, optimal control (dynamic)	79.0%	88.7%	95.8%	97.9%	98.0%

Revenue management earns almost nothing when the flight has empty seats and a great deal when it does not. With a demand factor of 1.19, optimal control earns 15.6 percent more than first come, first served in the dynamic model. Selling first come, first served, revenue even falls as demand grows past capacity, from CAD 97,710 at a demand factor of 0.99 to CAD 92,664 at 1.19, because cheap early bookings fill the cabin before the high fares arrive. Real gains also depend on forecasting quality and on fences: the PODS work found that better detruncation and forecasting could roughly double the gains of origin-destination control (Belobaba 2002), and the ePODS experiments of Chapter 17 showed leg control losing almost all its value when fences disappear.

18.9 The network planner's takeaway

Revenue management changes the numbers a planner reads in three ways.

It changes loads. An RM-controlled flight carries fewer passengers than one sold first come, first served, at higher fares: 150.6 passengers at CAD 620 against 153.5 at CAD 585 in the static model, 160.9 at CAD 619 against 161.9 at CAD 603 in the dynamic one. A lower load factor on a well-managed flight is a sign that seats were held for high fares, not that demand is weak.

It changes who is spilled. Under EMSR-b control, 7.6 of the 15.4 passengers spilled per departure in the static model are V travellers and 3.8 are Q travellers; under first come, first served, 5.0 Y and 4.7 B travellers are among the 12.5 turned away. Revenue management moves spill to the bottom of the fare ladder, which is why Chapter 9 warned against valuing spilled passengers at the average fare.

It changes the value of capacity. On a controlled flight, one more seat is worth its marginal value, the expected revenue of the last seat. For Laurentian's flight it is CAD 268 at the opening of sales in the dynamic model, and about CAD 295 in the static model: far below the average fare of CAD 620, and below even the lowest fare of CAD 419, because the last seat is filled only on some departures. Value is concave in capacity. Ten more seats would add CAD 1,896 per departure in the dynamic model and ten fewer would cost CAD 3,229. And in the static model the protection levels for the upper classes do not change with capacity at all (9, 26, 54 and 98 seats with 158, 168 or 178 seats): extra seats go to the cheapest fares. On this flight the classical spill estimate, the average fare times the probability that demand reaches the 168th seat, gives a similar number, CAD 278, because capacity is close to mean demand, where the two valuations of Chapter 9 cross. On a tighter flight they diverge: with 10 percent more demand in every class, the marginal seat is worth CAD 405 and the spill estimate gives CAD 548.

On a flight controlled by revenue management, one more seat is worth the marginal value of the last seat, not the average fare.

These marginal values are the bridge to the rest of Part IV. On a network, a seat on one leg is used by many itineraries, and its marginal value becomes the leg's bid price (Chapter 20). Chapter 22 shows how a planner uses bid prices and marginal seat values to price an upgauge, a frequency change or a new connection.

18.10 Summary

- Revenue management takes fares and capacity as given and allocates seats among fares to maximize revenue, which is neither the yield-maximizing nor the load-maximizing policy.
- An RM system forecasts demand to come from unconstrained booking history, optimizes protection levels, loads booking lim-

its into inventory and repeats the cycle at checkpoints until departure; its own controls shape the data it learns from.

- Classes are serially nested: protection levels y_j reserve seats for classes 1 to j, and booking limits satisfy $b_{j+1} = C - y_j$.
- Littlewood's rule protects the y-th seat for the high fare while $f_1 \Pr(D_1 \geq y) \geq f_2$: a newsvendor condition that protects less than mean demand when the fare ratio exceeds one half and does not depend on capacity.
- EMSR-a adds pairwise protection levels and ignores the pooling of uncertainty; EMSR-b merges the classes above each cut and is close to optimal, here exactly optimal.
- The dynamic program values each seat at $\Delta V_t(x)$, accepts a request when its fare is at least that value, and implies nested protection levels that fall as departure approaches.
- Limits must be re-solved as bookings arrive: on YUL–SVQ, EMSR-b revised at 16 checkpoints comes within 0.43 percent of optimal, while limits set once lose 8.2 percent.
- RM lowers loads, raises average fares, moves spill to the cheapest classes and values the marginal seat well below the average fare; its gains grow steeply with the demand factor.

Table 18.6 Key formulas of the chapter.

Quantity	Formula
Booking limits	$b_1 = C$, $b_{j+1} = C - y_j$
Littlewood's rule	$y^* = \max\{y : f_1 \Pr(D_1 \geq y) \geq f_2\}$
Littlewood, normal demand	$y^* = \mu_1 + \sigma_1 \Phi^{-1}(1 - f_2/f_1)$
EMSR-a	$y_j = \sum_{k \leq j} \left[\mu_k + \sigma_k \Phi^{-1}(1 - f_{j+1}/f_k)\right]^+$
EMSR-b	$y_j = \mu_{1:j} + \sigma_{1:j} \Phi^{-1}(1 - f_{j+1}/\bar{f}_{1:j})$
Static DP	$V_j(x) = \mathbb{E}[\max_{u \leq \min(D_j, x)} \{f_j u + V_{j-1}(x-u)\}]$
Dynamic DP	$V_t(x) = V_{t-1}(x) + \sum_j \lambda_j(t) \left[f_j - \Delta V_{t-1}(x)\right]^+$
Optimal acceptance	accept class j if $f_j \geq \Delta V_{t-1}(x)$

18.11 Further reading

Belobaba's lecture on flight leg revenue optimization (Belobaba 2006l) is the source of the system description and of the EMSR-b example, and his presentations on pricing challenges and on origin-destination control (Belobaba 2006o, 2002) report the PODS evidence on what revenue management earns. The original papers

are short and readable: Littlewood (1972) for the two-class rule, Belobaba (1987, 1989) for EMSR, Brumelle and McGill (1993) for the optimality of nested protection levels and Lee and Hersh (1993) for the dynamic model. Smith, Leimkuhler and Darrow (1992) describe the system that made American Airlines' reputation. Talluri and van Ryzin (2004b, ch. 2) give the definitive treatment of single-resource capacity control, and McGill and van Ryzin (1999) survey the field's first decades. Wittman and Belobaba (2019) set out the mechanisms by which airlines are moving from booking-class availability toward dynamic pricing.

CHAPTER

19

Overbooking

IN THIS CHAPTER

- Why selling more reservations than the aircraft has seats is the rational answer to no-shows, and what it is worth.
- The accounting of bookings, no-shows, go-shows, denied boardings and spoiled seats.
- Show-ups as a binomial random variable, its normal approximation, and why real show-up rates vary more than the binomial allows.
- From rules of thumb to the cost-minimizing authorization level, derived by marginal analysis, with the lecture's examples reproduced.
- Overbooking Laurentian's 182-seat A321XLR to Seville: the optimal limit, the denied-boarding risk and its sensitivity.
- Service-level targets, volunteers and auctions, and Canada's denied-boarding rules.
- How the limit moves over the booking horizon, and why overbooking and spill are different things.

DRAWS ON Belobaba, Flight overbooking: models and practice (MIT 16.75J, 2006); Belobaba, O&D control: what have we learned? (2002); Rothstein (1985); Talluri and van Ryzin (2004); Air Passenger Protection Regulations (SOR/2019-150); handbook

A network planner counts seats: 182 on the A321XLR, 298 on the 787-9. Revenue management sells something else. It sells reservations, and a reservation is a promise that the passenger may or may not use. A share of booked passengers never turns up at the gate, and if the airline stopped selling at the physical capacity it would fly every one of those seats empty, on exactly the flights where other customers had been turned away. Overbooking is the response: authorize more bookings than there are seats, so that the expected no-shows are replaced in advance. The price is the occasional flight on which more passengers show up than the cabin holds, and the airline must deny boarding to some of them.

This chapter derives how far to push. It treats the authorization level as an economic decision that weighs the cost of an empty

seat against the cost of a denied boarding, solves it exactly for the binomial show-up model, and works it through on Laurentian's YUL–SVQ flight. For a planner the result matters in three places. The saleable capacity that revenue management works with (Chapters 18 and 20) is the authorization level, not the seat count. A spill model of booking demand (Chapter 9) should cap bookings at the authorization level and passengers at the cabin. And every late gauge change creates an oversale that no overbooking model planned for, at a cost now fixed partly by regulation.

19.1 The overbooking problem

Fix one departure and one cabin. Booked passengers fall into two groups at departure time: those who present themselves for the flight and those who do not. Belobaba's lecture records that US domestic no-show rates could reach 15 to 20 percent of final pre-departure bookings on peak holiday days, with averages that had fallen to 10 to 15 percent as fare penalties spread (Belobaba 2006m). These are historical figures from the early 2000s, but the order of magnitude explains why the problem matters: effective overbooking can produce revenue gains as large as fare-class seat allocation (Belobaba 2006m), and Belobaba's review of the MIT research puts the gain from effective leg-based fare-class control and overbooking together at 4 to 6 percent of system revenue (Belobaba 2002).

DEFINITION 19.1 Overbooking vocabulary

For one departure with physical capacity C (seats in the cabin):

- the **authorization level** AU is the maximum number of confirmed bookings the airline accepts, and the **overbooking factor** is AU/C;
- BKD $\le$ AU is the number of confirmed bookings held at departure;
- a **no-show** is a booked passenger who neither cancels nor presents for the flight; a **show-up** is one who presents; the **show-up rate** is SUR $=$ SU/BKD and the **no-show rate** is NSR $= 1 -$ SUR;
- a **go-show** presents without a booking and a **standby** waits for a free seat; both board only if seats remain;
- a **denied boarding** (DB) is a show-up with a confirmed booking who cannot be carried;
- a **spoiled seat** is a seat that departs empty on a flight that was closed for sale, that is, with BKD $=$ AU.

The lecture ties the quantities together with simple accounting (Belobaba 2006m). Write PAX for passengers boarded, GS and SB for go-shows and standbys. Then

$$\begin{aligned} \text{PAX} &= \min(\text{BKD} - \text{NS} + \text{GS} + \text{SB},\ C), \\ \text{BKD} &= \text{NS} + \text{SU}, \qquad \text{DB} = \text{SU} + \text{GS} + \text{SB} - \text{PAX}, \\ \text{SP} &= C - \text{PAX} \quad \text{only when BKD} = \text{AU}. \end{aligned}$$

The last line deserves attention. Empty seats on a flight that never reached its authorization level reflect weak demand, not the overbooking decision; only a flight that was closed for sale and still departed with empty seats lost revenue because of too few bookings. Go-shows and standbys fill seats only after confirmed passengers, so in the models below we set them to zero, or equivalently fold them into a net show-up rate.

Two kinds of model error are possible, and they have very different visibility. Too low an authorization level spoils seats; nobody complains, and the loss shows up only as a load factor that is lower than it could have been. Too high a level denies boardings at the gate, in front of passengers, staff and, in Canada, a regulator. Belobaba notes that analysts who set limits by judgment tend to focus on avoiding denied boardings (Belobaba 2006m), and the difference in

visibility is one reason. The models in this chapter put both errors on the same scale.

Overbooking and spill are different (Network Planning Handbook 2026). Spill (Chapter 9) is demand turned away while the flight is on sale, because no seat is available at an acceptable fare. A denied boarding happens at the gate, to a passenger holding a confirmed booking. Overbooking reduces spill, since each extra authorized booking can go to a customer who would otherwise have been turned away, and creates the risk of denied boardings in return.

19.2 How many passengers show up?

Passengers fail to show for many reasons: a refundable ticket that cost nothing to abandon, a duplicate booking held as insurance, illness, a meeting that moved. At a hub there is a structural cause as well: a passenger who misses a connection at YUL appears as a no-show on the outbound leg. No-show rates therefore vary by fare product, by connection status, by group or individual booking, and by day and time, which is why they are best estimated segment by segment. Cancellations, which happen before departure and free the seat for resale, are a separate process that we meet again in Section 19.8.

The simplest model treats each booking as an independent coin flip.

DEFINITION 19.2 Binomial show-up model

If b bookings are held at departure and each shows up independently with probability q, the number of show-ups is $N(b) \sim \text{Bin}(b, q)$:

$$\Pr(N(b) = n) = \binom{b}{n} q^n (1-q)^{b-n}, \qquad n = 0, 1, \dots, b.$$

Independence is a strong assumption. Weather, a late inbound bank or a school group that cancels together move many show-ups in the same direction, so the show-up rate itself varies from one de-

parture to the next. The lecture's probabilistic model captures this variation of the rate across departures and ignores the coin-flip randomness of individual passengers (Belobaba 2006m). The two are easy to combine.

RESULT 19.1 Moments of the show-up count

Under the binomial model, $\mathbb{E}[N(b)] = bq$ and $\mathrm{Var}[N(b)] = bq(1-q)$, and for large b

$$\Pr(N(b) \geq C) \approx 1 - \Phi\left(\frac{C - \frac{1}{2} - bq}{\sqrt{bq(1-q)}}\right).$$

If instead the show-up rate Q varies across departures with mean q and variance σ_q^2, and show-ups are binomial given Q, then $\mathbb{E}[N(b)] = bq$ and

$$\mathrm{Var}[N(b)] = bq(1-q) + b(b-1)\,\sigma_q^2. \tag{19.1}$$

Proof. The binomial moments are standard. The normal approximation follows from the central limit theorem, with the continuity correction $\frac{1}{2}$ because $N(b) \geq C$ is the same event as $N(b) > C - \frac{1}{2}$. For the mixed model use the law of total variance, $\mathrm{Var}[N] = \mathbb{E}[\mathrm{Var}(N \mid Q)] + \mathrm{Var}(\mathbb{E}[N \mid Q])$. Given Q the count is binomial, so the first term is $b\,\mathbb{E}[Q(1-Q)]$ and the second is $\mathrm{Var}(bQ) = b^2\sigma_q^2$. Since $\mathbb{E}[Q^2] = q^2 + \sigma_q^2$, we have $\mathbb{E}[Q(1-Q)] = q(1-q) - \sigma_q^2$, and the total is $bq(1-q) - b\sigma_q^2 + b^2\sigma_q^2$, which is (19.1). ∎

The second term grows with b^2, the first only with b. On a large aircraft, uncertainty about the rate dominates the coin-flip noise. Take the Seville flight of Section 19.5: 193 bookings with $q = 0.93$ give a binomial standard deviation of $\sqrt{193 \times 0.93 \times 0.07} = 3.54$ show-ups. A show-up rate that varies by only 2.5 percentage points from departure to departure raises the variance from 12.6 to 35.7 and the standard deviation to 5.98. Figure 19.1 shows what this does to the tail beyond the cabin. A beta distribution for Q makes the mixture a **beta-binomial**, which is the overdispersed model used for comparison below.

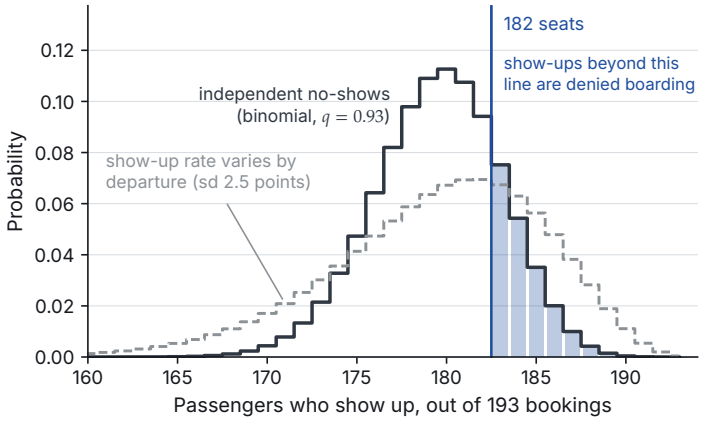


Figure 19.1 Distribution of show-ups for 193 bookings on the 182-seat A321XLR. With independent no-shows ($q = 0.93$, solid) the distribution is tight; when the show-up rate itself varies across departures with a standard deviation of 2.5 percentage points (dashed, beta-binomial), both tails fatten. The shaded bars are show-ups beyond the cabin under independence: every one is a denied boarding.

19.3 Rules of thumb

The lecture traces the evolution of airline practice through four models: judgment, a deterministic rule, a probabilistic (risk) rule and the cost-based model (Belobaba 2006m). The first three are worth knowing because they are still how people talk about overbooking, and because their weaknesses motivate the fourth.

Judgment. An analyst sets the overbooking factor from experience, typically at $1 + \text{NSR}$ or below, so that $\text{AU} = C(1 + \text{NSR})$.

The deterministic rule. Assume the flight books to its limit and exactly the average share shows up. Setting expected show-ups equal to capacity, $\text{AU}\,(1 - \text{NSR}) = C$, gives

$$\text{AU} = \frac{C}{1 - \text{NSR}} = \frac{C}{q}.$$

The probabilistic rule. Treat the show-up rate R of the coming departure as normal with mean q and standard deviation σ, and choose the authorization level so that show-ups exceed capacity with probability α only. Show-ups exceed capacity when $R > C/\text{AU}$, so the rule requires $C/\text{AU} = q + z_\alpha \sigma$ with $z_\alpha = \Phi^{-1}(1 - \alpha)$:

$$\text{AU} = \frac{C}{q + z_\alpha\,\sigma} = \frac{C}{1 - \text{NSR} + z_\alpha\,\sigma}. \tag{19.2}$$

The larger the uncertainty σ, the larger the denominator and the lower the limit.

WORKED EXAMPLE 19.1 Three rules for one flight

The lecture's example has $C = 100$ seats, a mean no-show rate of 20 percent and, for the risk rule, a standard deviation of the rate of $\sigma = 0.05$ with a 95 percent confidence level ($z = 1.645$) (Belobaba 2006m).

- Judgment: AU $= 100 \times 1.20 = 120$.
- Deterministic: AU $= 100/0.80 = 125$.
- Probabilistic: AU $= 100/(0.80+1.645\times0.05)$, which is 113.35, rounded to 113.

Evaluate each with the normal show-up-rate model behind the risk rule, using the loss function L of Chapter 9 (the formulas are derived in Section 19.4). At 125 bookings, expected show-ups equal the 100 seats, so denied boardings occur on half of all departures; on average 2.49 passengers are denied and 2.49 seats spoil. At 120, expected show-ups are 96, the probability of a denied boarding falls to 25 percent, and the flight averages 0.91 denied boardings and 4.91 spoiled seats. At 113, expected show-ups are 90.4: denied boardings almost vanish (0.10 per departure), but 9.70 seats leave empty on a flight that was closed for sale.

Lowering the confidence level to 90 percent ($z = 1.282$) raises the limit to 115.7; raising it to 99 percent ($z = 2.326$) lowers it to 109.1. Each rule controls something, but none of them asks whether the 7.2 extra empty seats of the risk rule are a reasonable price for avoiding the 2.4 denied boardings of the deterministic rule. That needs costs.

The lecture lists four practical extensions of the risk rule (Belobaba 2006m). Lower the confidence level when denied boardings are cheap. Add to the numerator the number of volunteers the station can usually find, since a tolerated voluntary denied boarding is not a service failure. Add the seats expected to be empty in a premium cabin into which economy passengers can be upgraded, which turns the premium cabin into a buffer (and allows it to be overbooked in turn). And remove firm group bookings, which show up far more reliably, overbooking only the remaining capacity.

19.4 The cost-based authorization level

Suppose the flight books to its limit b and show-ups are $N(b)$. Let c_{SP} be the cost of a spoiled seat (the contribution of the passenger who could have occupied it) and c_{DB} the cost of a denied boarding (compensation, meals and hotel, a seat on another carrier, and lost goodwill). The expected cost of authorizing b bookings is

$$K(b) = c_{\text{DB}}\,\mathbb{E}\big[[N(b) - C]^+\big] + c_{\text{SP}}\,\mathbb{E}\big[[C - N(b)]^+\big]. \tag{19.3}$$

This is the same problem as revenue maximization. If each boarded passenger contributes r and no-shows are refunded, contribution is $r\mathbb{E}[\min(N,C)] - c_{\text{DB}}\mathbb{E}[[N-C]^+]$. Since $\min(N,C) = C - [C-N]^+$, maximizing it is minimizing (19.3) with $c_{\text{SP}} = r$.

What is c_{SP} in practice? The lecture offers three candidates: the average revenue per seat on the leg, the highest fare (a closed flight loses late high-fare demand), or the lowest fare (one more authorized booking would have sold one more discount seat), and notes that the choice is as hard as estimating the denied-boarding cost (Belobaba 2006m). On a flight controlled by revenue management, the marginal seat sells at the lowest open fare (Chapter 9), which argues for the low end.

The decision is whether to accept booking number $b+1$. The answer is remarkably clean.

RESULT 19.2 The optimal authorization level

Under the binomial show-up model the cost of one more authorized booking is

$$K(b+1) - K(b) = q\Big[(c_{\text{DB}} + c_{\text{SP}})\,\Pr(N(b) \geq C) - c_{\text{SP}}\Big]. \tag{19.4}$$

K is convex in b, and the cost-minimizing authorization level is the smallest b such that

$$\Pr(N(b) \geq C) \;\geq\; \rho = \frac{c_{\text{SP}}}{c_{\text{SP}} + c_{\text{DB}}}. \tag{19.5}$$

Proof. Couple the two situations: $N(b+1) = N(b) + X$, where $X \sim \text{Bin}(1,q)$ is the show-up indicator of the extra booking, independent of $N(b)$. If $X = 0$ nothing changes. If $X = 1$, the extra passenger either fills an empty seat, when $N(b) < C$, which reduces spoilage by

one, or is denied boarding, when $N(b) \geq C$, which adds one denied boarding. Hence

$$\Delta K = q[c_{\mathrm{DB}} \Pr(N(b) \geq C) - c_{\mathrm{SP}} \Pr(N(b) < C)],$$

which rearranges to (19.4). The same coupling shows that $N(b)$ is stochastically increasing in b, so $\Pr(N(b) \geq C)$ and therefore the increment rise with b: the cost is discretely convex. Its minimum is at the first b whose increment is non-negative, which is (19.5). ■

Three features of the rule deserve a planner's attention. First, the show-up probability of the extra booking cancels: it multiplies both terms, because the extra passenger matters only if they show up, and then the only question is whether the cabin is already full. This relies on the independence of the binomial model. The show-up rate still matters through the distribution of $N(b)$, which is why the limit moves with q (Section 19.5). When show-ups are correlated, as in the mixed model of Result 19.1, an extra booking that shows up signals a high rate, and the exact condition becomes $\Pr(N(b) \geq C \mid \text{it shows up}) \geq \rho$, which is met at the same or a slightly lower limit. Second, the rule is the newsvendor of Chapter 9 seen from the other side. There the airline bought seats until the chance of filling the last one equalled its cost-to-contribution ratio; here it sells reservations until the chance that the last one finds the cabin full equals $c_{\mathrm{SP}}/(c_{\mathrm{SP}} + c_{\mathrm{DB}})$. Third, demand uncertainty does not move the limit. Suppose bookings are $\min(D, b)$, with random demand D independent of show-ups, and the airline maximizes expected contribution, so that an empty seat costs c_{SP} whether or not the flight closed. Raising the limit from b to $b+1$ matters only on departures where $D > b$, so the increment (19.4) is multiplied by $\Pr(D > b) > 0$ and its sign, hence the optimal limit, is unchanged. Demand determines how often the limit binds, not where it should be.

The normal approximation of Result 19.1 turns the rule into a formula.

RESULT 19.3 Normal approximation

Let $z = \Phi^{-1}(1-\rho)$ and $s = \sqrt{q(1-q)}$. The rule (19.5) holds approximately when $b \geq \hat{b}$, where $\sqrt{\hat{b}}$ is the positive root of $qx^2 + zsx - (C - \frac{1}{2}) = 0$:

$$\hat{b} = \left(\frac{-zs + \sqrt{z^2 s^2 + 4q\,(C - \frac{1}{2})}}{2q} \right)^2 \approx \frac{C - \frac{1}{2} - z\sqrt{C(1-q)}}{q},$$

and $\mathrm{AU}^\star = \lceil \hat{b} \rceil$.

The proof is a substitution: the condition $(C - \frac{1}{2} - bq)/(s\sqrt{b}) \leq z$ is a quadratic inequality in $x = \sqrt{b}$, and the approximation replaces $bq(1-q)$ by its value $C(1-q)$ at $b = C/q$. The second form reads naturally. Start from the deterministic limit C/q and pull back by z standard deviations of the show-ups; when denied boardings are more expensive than spoiled seats, $\rho < \frac{1}{2}$, $z > 0$, and the optimal limit sits below C/q.

The lecture's own cost-based example uses the normal show-up-rate model instead of the binomial. With AU bookings and a show-up rate $R \sim \mathcal{N}(q, \sigma^2)$, write $\theta = C/\mathrm{AU}$ for the show-up rate that exactly fills the cabin and $u = (\theta - q)/\sigma$. Then $\mathbb{E}[[\mathrm{AU}\,R - C]^+] = \mathrm{AU}\,\mathbb{E}[[R-\theta]^+]$, and the loss-function identity of Chapter 9 gives

$$\mathbb{E}[\mathrm{DB}] = \mathrm{AU}\,\sigma\,L(u), \qquad \mathbb{E}[\mathrm{SP}] = \mathrm{AU}\,\sigma\,(L(u) + u), \tag{19.6}$$

the second because $[\theta - R]^+ = [R-\theta]^+ - (R-\theta)$ and $\mathbb{E}[R-\theta] = -\sigma u$.

WORKED EXAMPLE 19.2 The lecture's cost-based example

Belobaba's example has $C = 120$, a mean no-show rate of 15 percent with a standard deviation of 8 percentage points, and costs of USD 50 per denied boarding and USD 100 per spoiled seat (Belobaba 2006m). These are historical teaching values; note that they make spoilage twice as costly as a denied boarding, an assumption we revisit below. Table 19.1 evaluates (19.6).

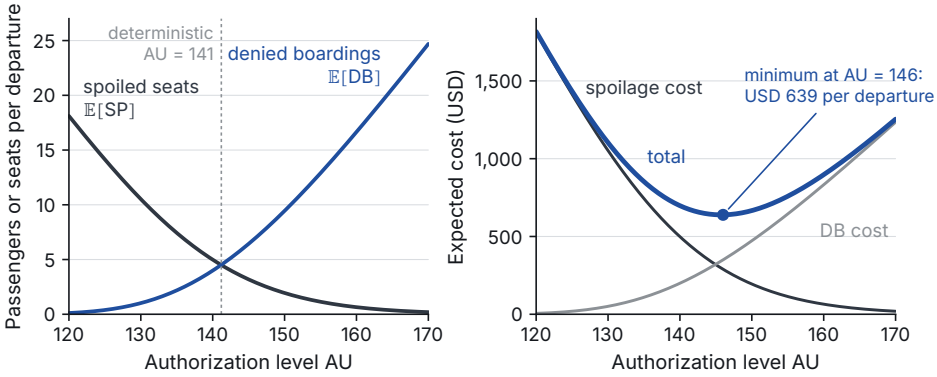


Figure 19.2 The lecture's cost-based example redrawn: 120 seats, mean show-up rate 0.85 with a standard deviation of 0.08, USD 50 per denied boarding and USD 100 per spoiled seat. Left: expected denied boardings and spoiled seats; they cross near the deterministic limit of 141. Right: expected costs; the total is flat near its minimum at 146 bookings. Curves computed with the normal show-up-rate model (19.6).

Table 19.1 Expected denied boardings, spoiled seats and costs (USD) in the lecture's example, computed with the normal show-up-rate model. The lecture's chart shows the same shapes, with a flat minimum of about USD 606 between 143 and 147 bookings.

AU	𝔼[DB]	𝔼[SP]	DB cost	Spoilage cost	Total
120	0.11	18.11	6	1,811	1,817
130	1.02	10.52	51	1,052	1,103
140	3.99	4.99	199	499	698
141	4.43	4.58	221	458	679
145	6.43	3.18	322	318	640
146	6.99	2.89	349	289	639
150	9.44	1.94	472	194	666
160	16.65	0.65	832	65	897
170	24.69	0.19	1,235	19	1,254

The minimum is at 146 bookings, an overbooking factor of 1.22, with an expected cost of USD 639. The continuous version of the first-order condition gives 145.8. The lecture's chart, read off the slide, shows a lower minimum of about USD 606 between 143 and 147 (Belobaba 2006m). The slide does not state how its curves were computed, but the curves reveal the parameters used. For any show-up distribution $\mathbb{E}[\mathrm{DB}] - \mathbb{E}[\mathrm{SP}] = \mathbb{E}[N] - C$, and the slide's two curves satisfy this identity with a mean

show-up rate of 0.851, not 0.85, at every authorization level. A normal rate with that mean and a standard deviation of 0.075 reproduces both of the slide's charts to within about 0.15 passengers and USD 13, with a minimum of about USD 600 at 145 bookings; rounded to two decimals, these parameters are the slide's labels, a no-show rate of 0.15 and a sigma of 0.08. Near the optimum the expected cost is roughly proportional to the standard deviation, so a standard deviation 6 percent below 0.08 explains why the slide's minimum lies about 5 percent below USD 639; the limit itself moves by one booking.

Compare the rules of thumb on the same flight. The deterministic limit, $120/0.85 = 141$, costs USD 679: it sits where expected denied boardings and expected spoiled seats are about equal, which is close to optimal only if the two cost the same (with equal costs the optimum is 140). The 95 percent risk rule, $120/(0.85 + 1.645 \times 0.08) = 122$, costs USD 1,659, 2.6 times the minimum, because it buys an almost certain absence of denied boardings with 16.5 spoiled seats per departure. Because the lecture prices a spoiled seat at twice a denied boarding, the optimum overbooks beyond 141 and accepts 7.0 expected denied boardings against 2.9 empty seats. Raise the denied-boarding cost to USD 200 and the optimum falls to 135.

PITFALL

Reading C/q as "the" overbooking level. The deterministic limit balances expected denied boardings against expected spoiled seats one for one. It is close to optimal only when the two cost the same, and with realistic costs, where a denied boarding costs several times a seat, it overbooks too much.

19.5 Overbooking the Seville flight

Laurentian's YUL–SVQ service uses the 182-seat A321XLR (Chapter 11). For this example treat the 182 seats as one cabin, as the network model of Chapter 20 does, and assume, for illustration, that 93 percent of bookings held at departure show up, independently. Spoilage costs c_{SP} = CAD 550: the restricted (L) fare of that network model, CAD 599, which stands for the lower fares of the ladder in Chapter 17, net of CAD 50 of passenger-variable cost and rounded.

On a flight controlled by revenue management, the seat that an extra booking would fill is a low-fare seat. A denied boarding costs $c_{\mathrm{DB}} = \text{CAD } 1{,}650$ on average, a blend of volunteers who accept a travel credit and a later flight and the occasional involuntary denial that triggers the regulated compensation of Section 19.7 plus rebooking and care. The critical ratio is

$$\rho = \frac{550}{550 + 1{,}650} = 0.25.$$

The rule says: keep accepting bookings while the chance that the show-ups already fill the cabin is below one in four.

WORKED EXAMPLE 19.3 How many bookings for 182 seats?

Table 19.2 evaluates the binomial model around the answer.

Table 19.2 Show-ups and expected overbooking cost per departure on YUL–SVQ ($C = 182$, $q = 0.93$, $c_{\mathrm{SP}} = 550$, $c_{\mathrm{DB}} = 1{,}650$).

Bookings b	$\mathbb{E}[N(b)]$	SD of $N(b)$	$\Pr(N(b) \geq 182)$	Expected cost (CAD)
189	175.8	3.51	0.042	3,493
190	176.7	3.52	0.079	3,068
191	177.6	3.53	0.134	2,719
192	178.6	3.54	0.206	2,480
193	179.5	3.54	0.295	2,391
194	180.4	3.55	0.395	2,483
195	181.4	3.56	0.499	2,779
196	182.3	3.57	0.601	3,290

With 192 bookings the probability of a full cabin is 0.206, below 0.25, so booking number 193 is worth accepting. With 193 it is 0.295, so the 194th is not. The optimal authorization level is $\mathrm{AU}^{\star} = 193$: eleven reservations beyond the seat count, an overbooking factor of 1.06. Result 19.3 gives $\hat{b} = 192.6$, hence the same 193, while the deterministic limit $182/0.93 = 195.7$ would accept two or three bookings more.

At 193 bookings the flight boards 179.0 passengers on average (a load factor of 98.4 percent), denies boarding to 0.46 and spoils 2.97 seats. The expected cost is CAD 758 of denied boardings plus CAD 1,633 of spoilage, CAD 2,391 in all. Without overbooking the same departure would spoil $182 \times 0.07 = 12.7$ seats, CAD 7,007; overbooking saves CAD 4,616 on every departure that books to the limit. Overshooting is costly too: the

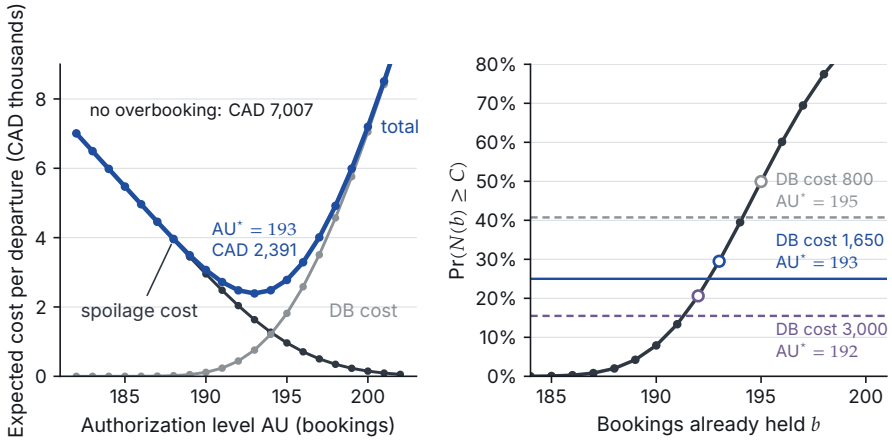


Figure 19.3 Overbooking YUL–SVQ. Left: expected spoilage, denied-boarding and total cost per departure as the authorization level varies; the minimum is at 193 bookings. Right: the marginal rule. The probability that show-ups among b bookings already fill the 182 seats rises steeply with b; the optimal limit is the first b at which it reaches the critical ratio, shown for three denied-boarding costs.

deterministic limit rounded up to 196 costs CAD 3,290, with 1.57 denied boardings per departure.

The decision is robust to the cost estimates. Reading (19.5) backwards, 193 is optimal for any denied-boarding cost between CAD 1,314 and CAD 2,117, since those are the costs at which ρ crosses $\Pr(N(192) \geq 182) = 0.206$ and $\Pr(N(193) \geq 182) = 0.295$. Figure 19.3 shows both the U-shaped cost and the probability rule.

The expected number of denied boardings hides their distribution. Figure 19.4 shows it at the optimal limit. Under independent no-shows, 80 percent of departures booked to 193 deny nobody, 7.5 percent deny one passenger, and 7.1 percent deny three or more. If the show-up rate in fact varies by 2.5 percentage points across departures, the picture changes: only 66 percent of departures deny nobody, 21 percent deny three or more, and the expected number of denied boardings is 1.21 instead of 0.46. An airline that sets its limit with the binomial model when the world is overdispersed pays CAD 4,038 per departure instead of the CAD 3,761 that the correct model would achieve, at its own optimum of 191 bookings.

Table 19.3 collects the sensitivity of the limit to the two inputs that are hardest to know, the denied-boarding cost and the variability of the show-up rate. Cheap volunteers push the limit up; costly invol-

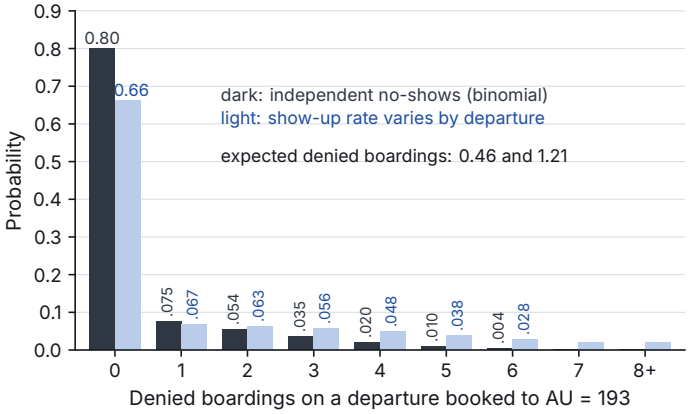


Figure 19.4 Number of denied boardings on a departure booked to the optimal limit of 193, with independent no-shows (dark) and with a show-up rate that varies across departures (light). The means are 0.46 and 1.21 denied boardings.

untary denials and correlated no-shows push it down. Across this range the limit moves between 189 and 195, an overbooking factor between 1.04 and 1.07. The show-up probability itself matters more: with $q = 0.90$ the optimal limit is 199, with $q = 0.95$ it is 189.

Table 19.3 Optimal authorization level on YUL–SVQ for three denied-boarding costs, with independent show-ups (binomial) and with a show-up rate whose standard deviation across departures is 2.5 percentage points (mixed, beta-binomial); $c_{\text{SP}} = 550$, $q = 0.93$.

DB cost (CAD)	ρ	AU, binomial	$\mathbb{E}[\text{DB}]$	$\mathbb{E}[\text{SP}]$	AU, mixed	$\mathbb{E}[\text{DB}]$	$\mathbb{E}[\text{SP}]$
800	0.41	195	1.10	1.75	193	1.21	3.72
1,650	0.25	193	0.46	2.97	191	0.62	4.99
3,000	0.15	192	0.27	3.71	189	0.26	6.49

For the planner, the example says three things. The seats that revenue management can sell on this flight are about 193, not 182, so a spill model of booking demand (Chapter 9) should cap bookings at the authorization level while passengers remain capped at the cabin. A departure that books to its limit carries about ten more passengers than one that stops at 182, which is part of what a gauge decision buys. And the expected denied-boarding rate on departures booked to the limit, 25.6 per 10,000 passengers boarded here, is not the airline's denied-boarding rate: most departures never reach their limit, and most denied boardings are voluntary.

19.6 Service-level criteria

Overbooking models also limit the risk of excessive denied boardings on individual flights, for customer-service reasons (Belobaba 2006m); the lecture's risk rule is one example. Denied boardings are visible, regulated and reported, and a service constraint can take three natural forms:

- the probability of any denied boarding on a departure is at most α;
- denied boardings per 10,000 passengers boarded stay below a target;
- the expected number of denied boardings per departure stays below a target.

Each is easy to evaluate with the show-up distribution. On YUL–SVQ with independent no-shows, the largest limit with at most a 5 percent chance of any denied boarding is 190 bookings; the expected cost is CAD 3,068, CAD 677 more than at the cost-based optimum. A target of 10 denied boardings per 10,000 passengers allows 191 bookings. With overdispersed show-ups the 5 percent rule allows only 187.

Is a service constraint a different philosophy from cost minimization? Not really. Every authorization level is the cost-minimizing one for some range of denied-boarding costs.

RESULT 19.4 The denied-boarding cost implied by a limit

Let $p_b = \Pr(N(b) \geq C)$. The authorization level b is the smallest minimizer of (19.3) if and only if

$$c_{\mathrm{SP}} \frac{1 - p_b}{p_b} \;\leq\; c_{\mathrm{DB}} \;<\; c_{\mathrm{SP}} \frac{1 - p_{b-1}}{p_{b-1}}. \tag{19.7}$$

Proof. By Result 19.2, b is optimal exactly when $p_{b-1} < \rho \leq p_b$. Substitute $\rho = c_{\mathrm{SP}}/(c_{\mathrm{SP}} + c_{\mathrm{DB}})$, which is decreasing in c_{DB}, and solve each inequality for c_{DB}. ∎

Apply it to the 5 percent rule on the Seville flight. The rule picks 190 bookings, with $p_{189} = 0.042$ and $p_{190} = 0.079$, so it behaves as if each denied boarding cost between CAD 6,396 and CAD 12,423, four to eight times the blended cost assumed in Example 19.3. A service target is therefore a statement about the cost of a denied boarding, usually an implicit and very high one. Making the implied cost explicit is the most useful conversation an analyst can have about over-

booking policy: either the airline believes a denied boarding costs CAD 6,000 or more, including the long-run goodwill effect, or the service target is costing it CAD 677 per full departure.

19.7 Volunteers and involuntary denied boardings

A denied boarding is not a single event with a single cost. The passenger may agree to give up the seat in exchange for compensation, or may be refused against their will.

> **DEFINITION 19.3 Voluntary and involuntary denied boarding**
>
> A **voluntary denied boarding** occurs when a passenger holding a confirmed booking agrees to give up the seat, usually in exchange for compensation and a later flight. An **involuntary denied boarding** occurs when the airline refuses boarding to a confirmed passenger who has not agreed to it.

The economics of the two are opposite. An involuntary denial costs regulated compensation and goodwill; a voluntary one costs whatever the volunteer asks, and a passenger who gladly sells a seat for a travel credit may leave happier than they arrived. The lecture urges a comprehensive voluntary program: trained station staff, identifying likely volunteers at check-in, and enough "soft" compensation to keep passengers satisfied (Belobaba 2006m). It reports that in 2004 US airlines denied boarding involuntarily to 0.62 passengers per 10,000 and that 95 percent of US denied boardings were voluntary, with carriers ranging from almost zero to about 1.6 involuntary denials per 10,000 (Belobaba 2006m). These are historical figures, but they show what a volunteer program does: it turns most of the expensive events into cheap ones, lowers the effective c_{DB} and so raises the optimal limit.

Simon (1968) proposed the economist's solution in 1968: instead of choosing whom to bump, buy the seats back in an auction, in which each passenger states in a sealed bid the smallest sum for which they would wait for the next flight, and the lowest bidders are paid and rebooked. Each passenger at the gate has such a reservation value, the smallest compensation for which they would take a later flight. If the airline needs k volunteers and raises its offer in steps until k passengers accept, it pays roughly the k-th lowest reservation value to each, and every passenger who stays values the trip more than anyone who leaves. A gate agent who raises the offer un-

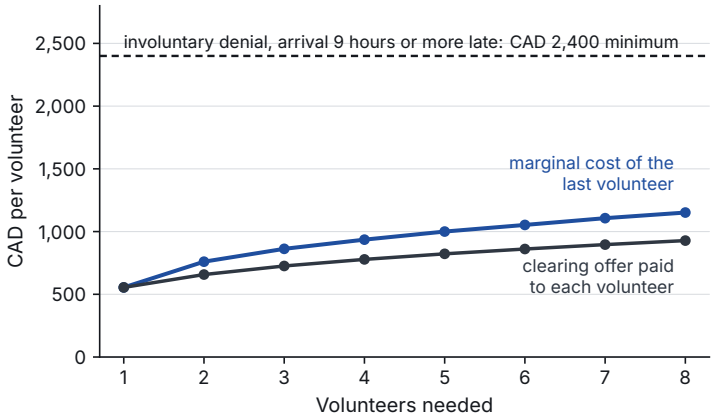


Figure 19.5 An ascending offer for volunteers at the YUL gate, with illustrative reservation values (lognormal, median CAD 2,500). The clearing offer paid to each volunteer rises slowly with the number needed; the marginal cost of the last volunteer rises faster. Both stay far below the CAD 2,400 minimum that an involuntary denial with a long delay would cost.

til enough passengers accept is running an ascending version of Simon's auction.

WORKED EXAMPLE 19.4 Buying back seats at the gate

Suppose three passengers too many show up for the evening YUL–SVQ departure. With three to four weekly frequencies, the next Laurentian flight to Seville is days away, so the alternative offered to a volunteer is a reroute through another hub or a later departure, and few passengers will give up the seat cheaply. Assume, for illustration, that reservation values are lognormal with a median of CAD 2,500 and a log standard deviation of 0.6, and that the agent raises the offer in steps of CAD 100.

Among 185 show-ups, about 0.7 passengers on average would accept CAD 500, 3.1 would accept CAD 700, and 11.7 would accept CAD 1,000. Simulating the auction, the expected clearing offer is CAD 555 when one volunteer is needed, CAD 726 each when three are needed (CAD 2,177 in all) and CAD 861 each for six (CAD 5,166 in all). The marginal cost of the third volunteer, the increase in the total bill, is CAD 862; that of the sixth is CAD 1,053. Figure 19.5 plots the curves.

Two lessons follow. Volunteers are much cheaper than involuntary denials: three involuntary denials on this flight would cost at least CAD 7,200 in regulated compensation alone if the passengers arrived nine hours or more late, before rebooking, care and goodwill. And the cost of denied boardings is convex: the more volunteers a flight needs, the more each one costs. Result 19.2 extends directly, with c_{DB} replaced by the expected marginal cost of one more denied boarding given the show-ups already in hand. The blended CAD 1,650 used for this flight is consistent with mostly voluntary denials at about CAD 700 to 900 plus rebooking, and an occasional involuntary one.

IN PRACTICE **Denied boarding rules for a Canadian carrier in 2026**

Canada's Air Passenger Protection Regulations set out what a carrier must do when it denies boarding; the consolidated text is current to 30 September 2026 and was last amended on 8 September 2022 (Government of Canada 2019). When a denial of boarding is within the carrier's control and not required for safety, which covers commercial overbooking, the carrier must inform affected passengers, follow the boarding rules, provide care, offer alternate travel or a refund, and pay minimum compensation (s. 12(4)). Before denying boarding it must ask all passengers whether they are willing to give up their seat (s. 15(1)); it may not remove a passenger already on board unless safety requires it (s. 15(2)); a benefit accepted by a volunteer must be confirmed in writing before departure (s. 15(3)). If denial is still necessary, boarding priority goes, in order, to unaccompanied minors, persons with a disability with their support person or animal, passengers travelling with family members, and passengers previously denied boarding on the same ticket (s. 15(4)). Passengers must receive food, drink and access to communication, and hotel accommodation with transport if the wait is overnight (s. 16).

Minimum compensation for a denied boarding depends on the delay in arrival at the destination on the original ticket: CAD 900 for less than 6 hours, CAD 1,800 for 6 to 9 hours and CAD 2,400 for 9 hours or more (s. 20(1)), paid no later than 48 hours after the denial (s. 20(2)), in money unless the passenger accepts another form of greater value that does not expire (s. 21) (Government

of Canada 2019). The amounts do not depend on carrier size, and the Canadian Transportation Agency presents the same schedule to passengers (Canadian Transportation Agency 2026a). A denial required for safety triggers the information, care and rebooking duties but not this compensation (s. 11(5)).

The regime is in flux. Amendments published for consultation in the Canada Gazette on 21 December 2024 would replace the current categories of disruption with a closed list of exceptional circumstances, so that compensation is owed unless one applies (Canadian Transportation Agency 2024); the Agency still described them as proposed in September 2026 (Canadian Transportation Agency 2026b). Bill C-31, introduced in May 2026, would move regulation-making and complaint handling to the Minister of Transport and raise maximum penalties (McCarthy Tétrault 2026). Check the status before using any of these details. On the return leg, SVQ–YUL departs from the European Union and falls under EU rules: for a flight over 3,500 km a denied passenger is owed EUR 600, which may be halved if the reroute arrives within 4 hours of the original schedule (European Union 2026). A revision of the EU rules, agreed politically in June 2026, keeps the compensation amounts and is to apply about a year after its publication in the Official Journal (European Commission 2026).

For the overbooking model, the regulation sets a floor under c_{DB} for involuntary denials, makes the volunteer search compulsory rather than optional, and ties the cost to the rebooking options the network offers. The last point is the planner's. On a route with several daily frequencies, a denied passenger reaches the destination within hours and the CAD 900 band applies; on a route served three times a week, like Seville, the CAD 2,400 band is the realistic one, which argues for a lower limit on thin long-haul routes than on dense domestic ones. A schedule change has the same effect: removing a frequency raises the cost of every denied boarding on the remaining ones.

> **PITFALL**
>
> Forgetting that a gauge change is an oversale. When a planner replaces a 169-seat 737-8 by a 137-seat A220 on a departure that is already sold, every booking beyond the new authorization level is a potential denied boarding that no overbooking

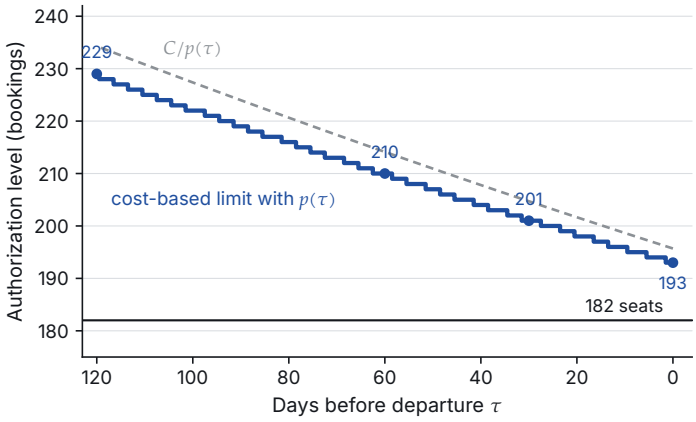


Figure 19.6 The authorization level over the booking horizon on YUL–SVQ. A booking held τ days out survives cancellation with probability $e^{-0.0015\tau}$ and then shows up with probability 0.93. The cost-based limit falls from 229 bookings four months out to 193 at departure; the dashed line is the deterministic limit $C/p(\tau)$.

model chose. The expected cost of such a swap, at regulated compensation rates, belongs in the business case for the schedule change (Chapters 16 and 22).

19.8 Overbooking over the booking horizon

So far the limit was set once, at departure. In reality bookings arrive over months, and a booking made 120 days out is far less likely to fly than one made yesterday, because it can still be cancelled. An airline therefore sets a time-dependent limit $\mathrm{AU}(\tau)$, where τ is the number of days to departure, and revises it at each reading point of the booking curve.

Let $p(\tau)$ be the probability that a booking held τ days before departure is still held at departure and then shows up. In the simplest model, each booking cancels at a constant daily rate h and survives with probability $e^{-h\tau}$, so $p(\tau) = q\,e^{-h\tau}$. A practical heuristic applies Result 19.2 at each reading point with $p(\tau)$ in place of q, as if sales stopped now. Figure 19.6 does this for the Seville flight with $h = 0.0015$ per day (illustrative), so that 83.5 percent of the bookings held 120 days out survive to departure.

The limit falls from 229 bookings at 120 days (25.8 percent above the seat count) to 210 at 60 days, 201 at 30 days, 195 a week out and 193 at departure. The airline can accept far more bookings early because many of them will cancel. The heuristic is biased in a known

direction. Treating the bookings on hand as final ignores the airline's option to resell seats freed by later cancellations, which makes early spoilage look more likely than it is, and it ignores that new sales can still be added, which makes early denied boardings look less likely than they are. Both effects say that the exact dynamic limit lies somewhat below the heuristic early in the horizon. Rothstein (1985) surveys the early operations research on the problem, including its dynamic formulations, and Talluri and van Ryzin (2004b, ch. 4) treat static and dynamic overbooking models and their combination with capacity control.

Overbooking also interacts with the rest of revenue management. On a single leg, the authorization level becomes the virtual capacity on which booking limits are computed: the booking limit of the top class equals the authorized capacity when the flight is overbooked, and lower classes are nested below it (Belobaba 2006l) (Chapter 18). On a network, the capacities fed to the optimization models of Chapter 20 should likewise be authorization levels, so that bid prices are computed on virtual seats; that chapter's examples use physical seats to keep the arithmetic simple. Overbooking is done by cabin, with expected empty premium seats serving as an upgrade buffer for economy, and group bookings, which show up far more reliably, are deducted before overbooking the rest (Belobaba 2006m).

> **PITFALL**
>
> Assuming independent no-shows because the binomial is convenient. Show-up rates move together with weather, late inbound connections and groups. A model that ignores this understates the tail of denied boardings, as Figure 19.4 shows, and sets limits that are too high on large aircraft, where the variance of the rate dominates.

19.9 Summary

- Overbooking authorizes more bookings than seats because a predictable share of booked passengers does not show up; flights closed for sale would otherwise depart with empty seats.
- Show-ups among b bookings are binomial with mean bq and variance $bq(1-q)$; when the show-up rate varies across departures the variance gains a term $b(b-1)\sigma_q^2$, which dominates on large aircraft.

- Judgment, the deterministic rule C/q and the probabilistic rule $C/(q + z\sigma)$ each control one thing; none balances the cost of spoiled seats against the cost of denied boardings.
- The cost-minimizing limit is the smallest b with $\Pr(N(b) \geq C) \geq c_{SP}/(c_{SP} + c_{DB})$. With independent show-ups the extra booking's show-up probability cancels from the marginal condition, and random demand independent of show-ups does not move the limit.
- On Laurentian's 182-seat Seville flight the optimal limit is 193 bookings; it saves CAD 4,616 per full departure compared with no overbooking, and stays optimal for denied-boarding costs between CAD 1,314 and CAD 2,117.
- Every service-level target implies a denied-boarding cost; a 5 percent limit on the probability of any denial behaves as if a denial cost CAD 6,400 or more.
- Volunteers bought back in an ascending auction cost far less than involuntary denials, which in Canada carry CAD 900 to 2,400 of minimum compensation; rebooking options, and therefore the schedule, set which band applies.
- Limits fall over the booking horizon as cancellations become less likely; the authorization level is the virtual capacity on which leg and network revenue management operate.

Table 19.4 Key formulas of the chapter.

Quantity	Formula
Show-up moments	$\mathbb{E}[N(b)] = bq$, $\mathrm{Var}[N(b)] = bq(1-q) + b(b-1)\sigma_q^2$
Deterministic limit	$\mathrm{AU} = C/q$
Risk rule	$\mathrm{AU} = C/(q + z_\alpha \sigma)$
Expected cost	$K(b) = c_{\mathrm{DB}}\mathbb{E}\,[N-C]^+ + c_{\mathrm{SP}}\mathbb{E}\,[C-N]^+$
Marginal booking	$\Delta K = q[(c_{\mathrm{DB}} + c_{\mathrm{SP}})p_b - c_{\mathrm{SP}}]$, $p_b = \Pr(N(b) \geq C)$
Optimal limit	smallest b with $\Pr(N(b) \geq C) \geq c_{\mathrm{SP}}/(c_{\mathrm{SP}} + c_{\mathrm{DB}})$
Normal approximation	$\mathrm{AU}^\star \approx (C - \frac{1}{2} - z\sqrt{C(1-q)})/q$
Rate model	$\mathbb{E}[\mathrm{DB}] = \mathrm{AU}\,\sigma L(u)$, $\mathbb{E}[\mathrm{SP}] = \mathrm{AU}\,\sigma(L(u) + u)$

19.10 Further reading

Belobaba's lecture (Belobaba 2006m) is the source of Examples 19.1 and 19.2 and of the cost-input discussion; it is short and worth reading for its practitioner's view of denied boardings and goodwill. Rothstein (1985) reviews the operations research on overbooking from its beginnings, and Talluri and van Ryzin (2004b, ch. 4) give the modern treatment of static and dynamic models, including overbooking jointly with capacity control. Simon (1968) is the two-page proposal behind volunteer auctions. Belobaba, Odoni and Barnhart (2015) place overbooking within the revenue management process of a network airline. For the regulations, read the Air Passenger Protection Regulations themselves (Government of Canada 2019) and check the Canadian Transportation Agency's site for amendments (Canadian Transportation Agency 2026b).

CHAPTER

20

Network Revenue Management

IN THIS CHAPTER

- Why controlling each flight leg on its own cannot maximize network revenue, on the lecture's Nice–Frankfurt example and on Laurentian's YUL bank.
- The family of origin-destination controls: value buckets and virtual nesting, displacement-adjusted virtual nesting (DAVN), and bid prices.
- The deterministic linear program over legs and fare products, its dual, and the proof that its dual prices are bid prices.
- Laurentian's four-leg, sixteen-product network solved: allocations, bid prices and the acceptance test.
- Uncertainty: probabilistic and randomized bid prices, and a booking simulation comparing first come first served, leg/class EMSR, DAVN and bid prices.
- What the PODS simulations found, and why the gains from network control are 1 to 2 percent of revenue.
- Bid prices as the shadow value of capacity: which legs are scarce, and what an extra seat or a larger aircraft is worth.

DRAWS ON Belobaba, Network revenue management: origin-destination control (MIT 16.75J, 2006); Belobaba, O&D control: what have we learned? (2002); handbook; Talluri and van Ryzin (1998, 2004); Williamson (1992)

Laurentian's evening bank at Montreal is a small machine for turning feeder flights into long-haul revenue. A passenger from Quebec City to Paris occupies a seat on two aircraft the network planner scheduled, the A220 from YQB and the 787-9 to CDG. When the last seat on the Paris flight is requested by a Montrealer and a Quebec City passenger at nearly the same moment, the revenue management system has to choose, and the right choice depends on more than the two fares: it depends on whether the Quebec City flight is

full, on what the YQB seat could otherwise earn, and on what the Paris seat could earn later. Single-leg methods (Chapter 18) cannot see any of this.

This chapter builds the network view. It starts from the two ways in which leg-based control loses money at a hub, introduces the controls that airlines developed in response, and then derives the deterministic linear program whose dual prices, the bid prices, have become the common currency of network revenue management. We solve Laurentian's four-leg network, test the bid-price rule, and compare the controls by simulation. The last section turns the bid prices around and reads them as a planner should: as the marginal value of a seat on each leg, which is where Chapter 22 picks up the thread.

20.1 Why leg control fails on a network

Most airlines in the 2000s practised **fare class control**: high-yield unrestricted fares sold in the top booking classes, discounts in the lower ones, and each flight leg managed by its own booking limits (Chapter 18). A connecting itinerary could be sold in a class only if that class was open on every leg it used, so the airline could not distinguish one itinerary from another, and a congested leg could block long-haul passengers behind it (Belobaba 2006n). The system was designed to maximize yield, not network revenue.

Belobaba's lecture uses a Lufthansa-style hub at Frankfurt with a feeder from Nice and long-haul legs to Hong Kong and New York (Belobaba 2006n). Table 20.1 gives the one-way fares.

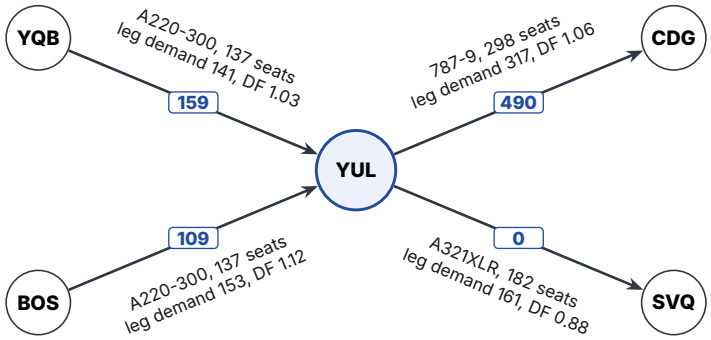
Table 20.1 One-way fares (USD) by booking class in the lecture's Frankfurt hub example (adapted from Belobaba, 2006; historical teaching values).

Class	NCE–FRA	NCE–HKG via FRA	NCE–JFK via FRA	FRA–JFK
Y	450	1,415	950	920
B	380	975	710	670
M	225	770	550	515
Q	165	590	425	385
V	135	499	325	315

Leg control fails in two opposite ways. **Short haul blocks long haul**: when the Nice–Frankfurt leg has closed classes M, Q and V to protect seats for its own high-yield locals, a Nice–Hong Kong passenger in M at USD 770 is refused, although that passenger is worth twice the USD 380 of the local B passenger who is still accepted and the Hong Kong leg has empty seats. **Connecting passengers displace locals**: when both Nice–Frankfurt and Frankfurt–New York are full, a Nice–New York passenger in B pays USD 710 for two seats that local B passengers would have bought for USD 380 + 670 = 1,050. The lecture distils the network problem into two strategies: give more availability to high-revenue long-haul passengers regardless of their yield, and prevent long-haul passengers from displacing high-yield locals on full flights. The first matters more, because both legs of a connection are seldom full at the same time (Belobaba 2006n).

DEFINITION 20.1 O&D fare products and the network

A **leg** $l \in \mathcal{L}$ is a single flight departure with capacity C_l. An **origin-destination fare product** (ODF) $j \in \mathcal{J}$ is an itinerary from an origin to a destination, consisting of one or more legs, combined with a fare product; it has fare f_j and random demand D_j with mean μ_j. The **incidence** a_{lj} equals 1 if product j uses leg l and 0 otherwise.



16 products: 8 O&D markets (4 local, 4 connecting via YUL) × 2 fares (H, L).
DF: mean leg demand divided by seats.
Boxed: bid price π_l of each leg from the deterministic LP (CAD).

Figure 20.1 Laurentian's bank at YUL: two feeder legs operated with A220-300s and two long-haul legs, sixteen fare products in eight O&D markets. Leg demand is the sum of mean demands of all products that use the leg; the boxed numbers are the bid prices computed in Section 20.6.

WORKED EXAMPLE 20.1 Leg control on Laurentian's bank

Laurentian's four-leg network (Figure 20.1) carries local passengers on each leg and connecting passengers from Quebec City and Boston to Paris and Seville, each with a flexible (H) and a restricted (L) fare. The network model aggregates each market's fares into one high and one low product, a simplification of the full fare ladder of Chapter 17. Three legs are oversubscribed on average: YQB–YUL by 3 percent, BOS–YUL by 12 percent and YUL–CDG by 6 percent; YUL–SVQ has 182 seats for 161 expected passengers. A quarter of all O&D demand connects.

Under leg/class control each leg sees two classes. On YQB–YUL, class L contains the local YQB–YUL L passenger at CAD 159, the YQB–CDG L passenger at CAD 649 and the YQB–SVQ L passenger at CAD 699, and the leg cannot tell them apart. Both failure modes appear.

Short haul blocks long haul. When YQB–YUL closes class L to protect seats for its CAD 329 flexible locals, it refuses the YQB–SVQ L passenger at CAD 699 as well, although the Seville leg usually has empty seats and the network would gain CAD 699 for a feeder seat whose next-best use is worth about CAD 159.

Connecting passengers displace locals. When BOS–YUL and YUL–CDG are both full, a BOS–CDG L passenger at CAD 599 occupies two seats that local L passengers would buy for CAD 189 + 549 = 738, and that late-booking flexible locals would buy for CAD 389 + 1,149 = 1,538.

Which effect dominates is an empirical question about the network, and answering it needs a model of the whole bank. Section 20.6 gives one.

20.2 Origin-destination control: value buckets

DEFINITION 20.2 O&D control

Origin-destination (O&D) control is the capability to respond to requests for different O&D itineraries on the same flight with different seat availability, based on the network revenue value of each request rather than on its fare restrictions or yield (Belobaba 2006n, 2002).

The first implementation abandons the fixed link between fare product and booking class. Each ODF is mapped, on each leg it uses, to a **value bucket** according to its revenue, and the leg's inventory controls the buckets with nested booking limits exactly as it controlled fare classes. Buckets can be built inside the existing classes, by republishing fare codes according to value, or as **virtual classes** hidden behind the published ones, which requires a new inventory structure and seamless availability links with the distribution systems (Belobaba 2006n). Virtual nesting of this kind was developed at American Airlines (Smith, Leimkuhler, and Darrow 1992). Table 20.2 shows the lecture's mapping of the Nice–Frankfurt leg into ten virtual classes by total itinerary fare.

Table 20.2 Virtual classes on the Nice–Frankfurt leg, mapped by total itinerary fare (adapted from Belobaba, 2006).

Virtual class	Revenue range (USD)	ODFs mapped
1	1,200 and above	Y NCE–HKG
2	900–1,199	B NCE–HKG, Y NCE–JFK
3	750–899	M NCE–HKG
4	600–749	B NCE–JFK
5	500–599	Q NCE–HKG, M NCE–JFK
6	430–499	V NCE–HKG, Y NCE–FRA
7	340–429	B NCE–FRA, Q NCE–JFK
8	200–339	V NCE–JFK, M NCE–FRA
9	150–199	Q NCE–FRA
10	0–149	V NCE–FRA

The mapping solves the first failure mode: a Nice–Hong Kong V passenger at USD 499 now ranks above a local B passenger at USD 380 on the feeder. It does so with the existing leg-based forecasting and optimization, which made it attractive. But it is **greedy**: ranking by total fare treats the long-haul passenger as if the second seat were free, and favours connecting traffic even when the downstream leg is full. Re-mapping ODFs between buckets also disturbs the bucket-level booking histories on which forecasts depend (Belobaba 2006n). The cure for greed is to charge each ODF for the seats it uses elsewhere.

20.3 Displacement and DAVN

DEFINITION 20.3 Displacement cost and network value

The **displacement cost** of a seat on leg l is the expected revenue the network loses when that seat is sold, that is, the revenue of the passenger it displaces; denote it π_l. The **network value** of ODF j on leg l is its fare minus the displacement costs of the other legs it uses,

$$v_{jl} = f_j - \sum_{l' \neq l} a_{l'j}\, \pi_{l'}. \tag{20.1}$$

Displacement-adjusted virtual nesting (DAVN) maps each ODF to the virtual bucket of each leg by its network value instead of its total fare, and keeps leg-based EMSR booking limits on the buckets. Displacement costs come from a network optimization, typically the

linear program of Section 20.5, or from leg-based EMSR estimates. When demand is high, displacement costs rise, connecting passengers fall to lower buckets and locals get more seats; displacement costs must be revised often enough to track demand, but every revision re-maps ODFs and disturbs the bucket forecasts (Belobaba 2006n).

On Laurentian's network, with the displacement costs computed in Section 20.6 (CAD 159 on YQB–YUL, 109 on BOS–YUL, 490 on YUL–CDG and 0 on YUL–SVQ), the YQB–YUL leg ranks its six ODFs as follows. By total fare the order would be YQB–SVQ H (1,399), YQB–CDG H (1,299), YQB–SVQ L (699), YQB–CDG L (649), YQB–YUL H (329), YQB–YUL L (159). After subtracting the Paris leg's displacement cost of 490, YQB–CDG H is worth 809 and YQB–CDG L only 159 to the Quebec City leg: the restricted Paris connection drops from fourth place to last, tied with the local L fare and below the local H. The Seville connections keep their full fares, since their second leg displaces nobody. On YUL–CDG the same logic ranks the local YUL–CDG L passenger (549) above both restricted connections, which are worth only 490 to the Paris leg once their feeder seats are charged for ($649 - 159$ and $599 - 109$).

Figure 20.2 shows the comparison for all sixteen products: each fare against the displacement cost of the seats it occupies.

The lecture makes a further observation that turns DAVN into something simpler (Belobaba 2006n). Under value buckets, ODF j is accepted on leg l if its network value exceeds the value of the last seat on the leg, which is the leg's own displacement cost π_l:

$$f_j - \sum_{l' \neq l} a_{l'j}\pi_{l'} > \pi_l \iff f_j > \sum_{l} a_{lj}\,\pi_l .$$

The right-hand side involves only one number per leg. Value buckets with displacement and bid prices are two mechanisms for the same decision rule.

20.4 Bid-price control

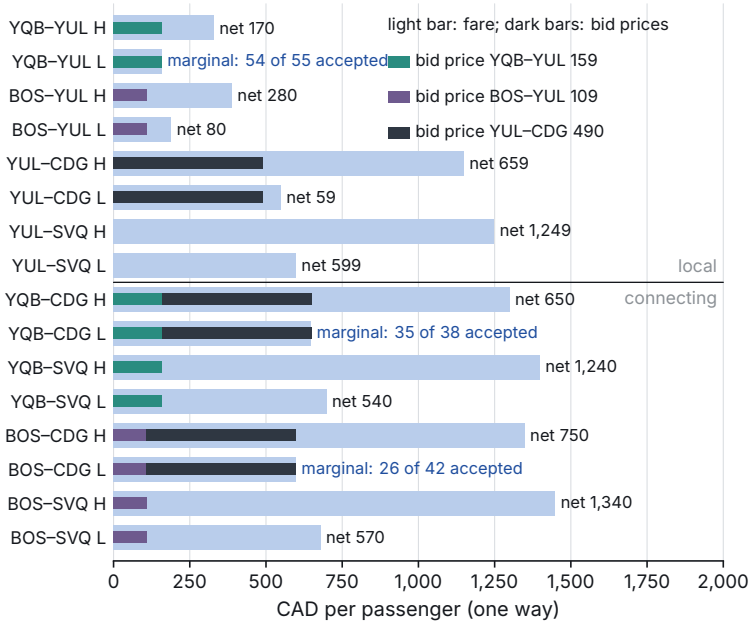


Figure 20.2 The bid-price test on Laurentian's network. Each light bar is a product's fare; the dark bars inside it are the displacement costs (bid prices) of the legs it uses, stacked. A product is worth accepting when its fare exceeds the stack; three restricted products sit exactly at the margin, and the deterministic LP accepts only part of their demand.

DEFINITION 20.4 Bid price and bid-price control

The **bid price** $\pi_l \geq 0$ of leg l is the minimum revenue the airline requires for one more seat on that leg, an estimate of the opportunity value of the next seat. Under **bid-price control**, a request for ODF j is accepted if seats are available and

$$f_j \;\geq\; \sum_{l \in \mathcal{L}} a_{lj}\, \pi_l, \tag{20.2}$$

that is, if the fare covers the sum of the bid prices along the itinerary.

The mechanism is much simpler than virtual buckets: the inventory stores one number per leg and compares the itinerary's fare with the itinerary's bid price when an availability request arrives. The price of simplicity is that bid prices must be recomputed frequently. A bid price says whether a product is worth selling, not how

many units of it to sell, so between revisions every product above its bid price is sold without limit (Belobaba 2006n). Bid prices can come from network optimization or from leg-based heuristics.

WORKED EXAMPLE 20.2 Bid prices on a four-city line

The lecture's example has a line A–B–C–D with leg bid prices of USD 35 on A–B, 240 on B–C and 160 on C–D (Belobaba 2006n). Itinerary bid prices add up along the path: 240 for B–C, 35 + 240 = 275 for A–C, and 35 + 240 + 160 = 435 for A–D.

Table 20.3 Availability under bid-price control in the lecture's example: fare in USD, then whether the class is open (adapted from Belobaba, 2006).

Request	Bid price	Y	M	B	Q
B–C	240	440 yes	315 yes	223 no	177 no
A–C	275	519 yes	374 yes	292 yes	201 no
A–D	435	582 yes	399 no	322 no	249 no

Class B is open for A–C ($292 > 275$) but closed for the shorter B–C ($223 < 240$), and class M is open on both shorter itineraries but closed for A–D. Availability no longer belongs to a class; it belongs to a fare on an itinerary, compared with the displacement it causes.

Leg-based heuristics produce bid prices without a network model. The lecture's EMSR heuristic bid price (HBP) controls local requests with ordinary leg booking limits and accepts a two-leg connection only if its fare exceeds

$$\max(\text{EMSR}_1, \text{EMSR}_2) + d \, \min(\text{EMSR}_1, \text{EMSR}_2), \qquad 0 < d < 1,$$

where EMSR_l is the expected marginal seat revenue of the last available seat on leg l (Belobaba 2006n). With last-seat values of 240 and 160 and $d = 0.5$, the connection must pay 320 rather than 400: the heuristic discounts the less congested leg because both legs are rarely full at once. Such heuristics reuse the airline's leg-based data and processes, which made them a low-risk first step toward network control.

20.5 The deterministic LP and its dual

Where do network bid prices come from? Replace each random demand by its mean and ask how a planner with perfect knowledge of those means would allocate seats. The answer is a linear program.

$$\begin{aligned} V(C) = \max_{x} \ & \sum_{j \in \mathcal{J}} f_j x_j \\ \text{s.t.} \ & \sum_{j \in \mathcal{J}} a_{lj} x_j \le C_l, \quad l \in \mathcal{L}, \\ & 0 \le x_j \le \mu_j, \quad j \in \mathcal{J}. \end{aligned} \tag{20.3}$$

The variable x_j is the number of seats allocated to product j. This **deterministic linear program** (DLP) ignores the randomness of demand and the order in which requests arrive, which is why its allocations make poor booking limits; its value lies in its dual.

Attach a multiplier $\pi_l \ge 0$ to each capacity constraint and $\beta_j \ge 0$ to each demand bound. The Lagrangian is

$$\sum_j f_j x_j + \sum_l \pi_l \Big(C_l - \sum_j a_{lj} x_j \Big) + \sum_j \beta_j (\mu_j - x_j).$$

Its supremum over $x \ge 0$ is finite only if the coefficient of every x_j is non-positive, that is, $f_j - \sum_l a_{lj}\pi_l - \beta_j \le 0$, and then equals $\sum_l C_l \pi_l + \sum_j \mu_j \beta_j$. Minimizing this bound gives the dual.

$$\begin{aligned} \min_{\pi, \beta} \ & \sum_{l \in \mathcal{L}} C_l \, \pi_l + \sum_{j \in \mathcal{J}} \mu_j \, \beta_j \\ \text{s.t.} \ & \sum_{l} a_{lj} \, \pi_l + \beta_j \ge f_j, \quad j \in \mathcal{J}, \\ & \pi_l \ge 0, \quad \beta_j \ge 0. \end{aligned} \tag{20.4}$$

The dual has a clean economic reading. It splits the network's revenue into a price π_l for every seat on every leg and a rent β_j for every unit of demand. Each product's fare must be covered by the prices of the seats it uses plus its rent, and the dual looks for the cheapest such accounting. By strong duality the two programs have the same optimal value.

RESULT 20.1 Dual prices are bid prices

Let x solve (20.3) and (π, β) solve (20.4). Then for every product j and leg l:

1. if $f_j > \sum_l a_{lj}\pi_l$, the product is fully accepted: $x_j = \mu_j$;
2. if $f_j < \sum_l a_{lj}\pi_l$, the product is rejected: $x_j = 0$;
3. if $\pi_l > 0$, leg l is full: $\sum_j a_{lj}x_j = C_l$.

Bid-price control (20.2) with the dual prices therefore reproduces the LP allocation for every product whose fare differs from its itinerary bid price. Products with $f_j = \sum_l a_{lj}\pi_l$ are **marginal**: the LP may accept part of their demand, and the bid price alone does not say how much.

Proof. Complementary slackness for the pair of programs states that at optimal solutions

$$x_j\Big(\sum_l a_{lj}\pi_l + \beta_j - f_j\Big) = 0, \quad \beta_j(\mu_j - x_j) = 0, \quad \pi_l\Big(C_l - \sum_j a_{lj}x_j\Big) = 0.$$

(1) If $f_j > \sum_l a_{lj}\pi_l$, dual feasibility forces $\beta_j \geq f_j - \sum_l a_{lj}\pi_l > 0$, and the second condition gives $x_j = \mu_j$. (2) If $x_j > 0$, the first condition gives $f_j = \sum_l a_{lj}\pi_l + \beta_j \geq \sum_l a_{lj}\pi_l$; so $f_j < \sum_l a_{lj}\pi_l$ forces $x_j = 0$. (3) is the third condition. ∎

The bid price is also the marginal value of capacity, which is what makes it interesting to a planner.

RESULT 20.2 Bid prices are marginal values of capacity

The optimal value $V(C)$ of (20.3) is a concave, piecewise-linear function of the capacity vector C. For any optimal dual π at C and any other capacity vector C',

$$V(C') \;\leq\; V(C) + \sum_l \pi_l\,(C'_l - C_l). \tag{20.5}$$

If the optimal π is unique, V is differentiable at C and $\partial V / \partial C_l = \pi_l$.

Proof. The dual feasible set does not depend on C. By strong duality $V(C')$ is the minimum over that set of $\sum_l C'_l\pi'_l + \sum_j \mu_j\beta'_j$, a minimum of affine functions of C', hence concave and piecewise linear. Evalu-

ating the minimum at the particular point (π, β) that is optimal for C gives

$$V(C') \le \sum_l C'_l \pi_l + \sum_j \mu_j \beta_j = V(C) + \sum_l \pi_l (C'_l - C_l),$$

which is (20.5). The inequality says that π is a supergradient of V at C. Conversely, the minimum is attained at one of the finitely many vertices of the dual feasible set, so near C the function V is the minimum of the affine pieces given by the vertices that are optimal at C. If all optimal duals share the same π, these pieces all have slope π, and V is differentiable at C with gradient π. ∎

Equation (20.5) has a sobering corollary: in the deterministic LP, multiplying a bid price by a large change in capacity can only overstate the value of extra seats and understate the loss from removing them. We return to this in Section 20.9.

Finally, the DLP bounds what any control can achieve.

RESULT 20.3 The deterministic LP is an upper bound

Let the demand D_j for each product be random with mean μ_j, and suppose that a refused request is lost rather than diverted to another product. For any booking policy, including one that knows the realized demand in advance, the expected revenue is at most $V(C)$.

Proof. Let X_j be the number of units of product j the policy sells. On every realization, $\sum_j a_{lj} X_j \le C_l$ and $0 \le X_j \le D_j$. Taking expectations, $\mathbb{E}[X]$ satisfies the capacity constraints and $0 \le \mathbb{E}[X_j] \le \mu_j$, so it is feasible for (20.3), and expected revenue $\sum_j f_j \mathbb{E}[X_j]$ cannot exceed $V(C)$. ∎

The gap between the bound and what policies achieve measures the cost of uncertainty, and the next sections put numbers on it. Talluri and van Ryzin (1998) analyze bid-price controls in depth: they are not optimal in general, because the value of a seat on one leg depends on how many seats remain on the others, but bid prices from the DLP become asymptotically optimal when demand and capacity grow large together.

20.6 Laurentian's network solved

WORKED EXAMPLE 20.3 The bank at YUL as a linear program

The data are Laurentian's sixteen products (four local and four connecting markets, each with an H and an L fare) with their mean demands, and the four leg capacities of Figure 20.1 (physical seats; an airline that overbooks would use its authorization levels, Chapter 19). Solving (20.3) with an LP solver (`scipy.optimize.linprog`, HiGHS method) gives an optimal revenue of CAD 365,949 per bank. Solving the dual (20.4) separately returns the same value and the bid prices

$$\pi_{\mathrm{YQB}} = 159, \qquad \pi_{\mathrm{BOS}} = 109, \qquad \pi_{\mathrm{CDG}} = 490, \qquad \pi_{\mathrm{SVQ}} = 0,$$

in CAD per seat; minimizing and maximizing each π_l over the optimal dual face shows that they are unique. Table 20.4 lists the allocation.

The three oversubscribed legs are exactly full (137, 137 and 298 passengers); YUL–SVQ carries its 161 expected passengers and has a bid price of zero. Twenty requests are refused: one YQB–YUL L, three YQB–CDG L and sixteen BOS–CDG L.

The bid prices form a chain that can be read straight off the marginal products. On YQB–YUL the marginal passenger is the local restricted fare, so a seat there is worth its CAD 159. On YUL–CDG the marginal passenger is the YQB–CDG L connection: it pays 649 and its feeder seat is worth 159, so a Paris seat is worth $649 - 159 = 490$. On BOS–YUL the marginal passenger is the BOS–CDG L connection: it pays 599 and its Paris seat is worth 490, so a Boston seat is worth only 109, less than the Boston local's restricted fare of 189. That last number has a precise meaning: every local Boston passenger is already accepted, and the best use of one more BOS–YUL seat is to let one more BOS–CDG L connect, which earns 599 but costs a Paris seat worth 490.

Now test the bid-price rule. YUL–CDG L at 549 clears the Paris bid price of 490 and is accepted in full, although it pays less than either restricted Paris connection; YQB–SVQ L at 699 clears $159 + 0$ with CAD 540 to spare. The three marginal products have fares exactly equal to their path bid prices, and the LP accepts 54 of 55, 35 of 38 and 26 of 42 of them. Verifying complemen-

tary slackness for every product and leg confirms Result 20.1 numerically.

This is the answer to Example 20.1. With today's means, the network wants all of its YQB–SVQ connections, because the Seville leg has room, and refuses part of its restricted Paris connections, because the Paris leg is full and a restricted connection earns no more than the bid prices of the two seats it occupies. Both failure modes of leg control are corrected by the same four numbers.

Table 20.4 Solution of the deterministic LP for Laurentian's bank (CAD): seats allocated to each product, the sum of the bid prices along its path, and the dual rent β_j, which equals the fare minus the path bid price.

Product	Fare	Mean demand	Allocated	Path bid price	Rent
YQB–YUL H	329	18	18	159	170
YQB–YUL L	159	55	54	159	0
BOS–YUL H	389	20	20	109	280
BOS–YUL L	189	50	50	109	80
YUL–CDG H	1,149	45	45	490	659
YUL–CDG L	549	170	170	490	59
YUL–SVQ H	1,249	22	22	0	1,249
YUL–SVQ L	599	90	90	0	599
YQB–CDG H	1,299	10	10	649	650
YQB–CDG L	649	38	35	649	0
YQB–SVQ H	1,399	4	4	159	1,240
YQB–SVQ L	699	16	16	159	540
BOS–CDG H	1,349	12	12	599	750
BOS–CDG L	599	42	26	599	0
BOS–SVQ H	1,449	5	5	109	1,340
BOS–SVQ L	679	24	24	109	570

PITFALL

Reading a zero bid price as a free seat. The DLP prices YUL–SVQ at zero because 161 expected passengers fit into 182 seats. But demand varies: with a standard deviation of 26 passengers, the Seville flight fills on about one departure in five, and on those departures its last seats are valuable. Deterministic bid prices are blind to variability; Section 20.7 repairs this.

20.7 Uncertainty: probabilistic bid prices and simulation

The DLP replaces random demand by its mean. Two refinements bring the randomness back while keeping a tractable optimization, and a third, simulation, tells us how the resulting controls actually perform.

The first refinement keeps the network structure but values each allocated seat at its expected revenue. If product j receives x_j seats, it sells $\min(D_j, x_j)$ of them, and by Chapter 9 the derivative of $\mathbb{E}[\min(D_j, x_j)]$ is $\Pr(D_j > x_j)$: the x-th seat given to product j earns $f_j \Pr(D_j > x)$ in expectation, its **expected marginal seat revenue** (Chapter 18).

RESULT 20.4 A probabilistic network program

Consider

$$\max_{x \geq 0} \sum_j f_j \, \mathbb{E}[\min(D_j, x_j)] \quad \text{s.t.} \quad \sum_j a_{lj} x_j \leq C_l, \quad l \in \mathcal{L}.$$

The objective is concave. At an optimum with capacity multipliers $\pi_l \geq 0$, every product that receives seats satisfies

$$f_j \Pr(D_j > x_j) = \sum_l a_{lj} \pi_l, \tag{20.6}$$

and a product with $f_j \leq \sum_l a_{lj} \pi_l$ receives none: each product gets seats until its expected marginal seat revenue falls to the bid price of its path.

The proof is the Karush-Kuhn-Tucker condition of a concave program: the derivative of the Lagrangian with respect to x_j is $f_j \Pr(D_j > x_j) - \sum_l a_{lj} \pi_l$, which must vanish when $x_j > 0$ and be non-positive when $x_j = 0$. Equation (20.6) is Littlewood's rule of Chapter 18

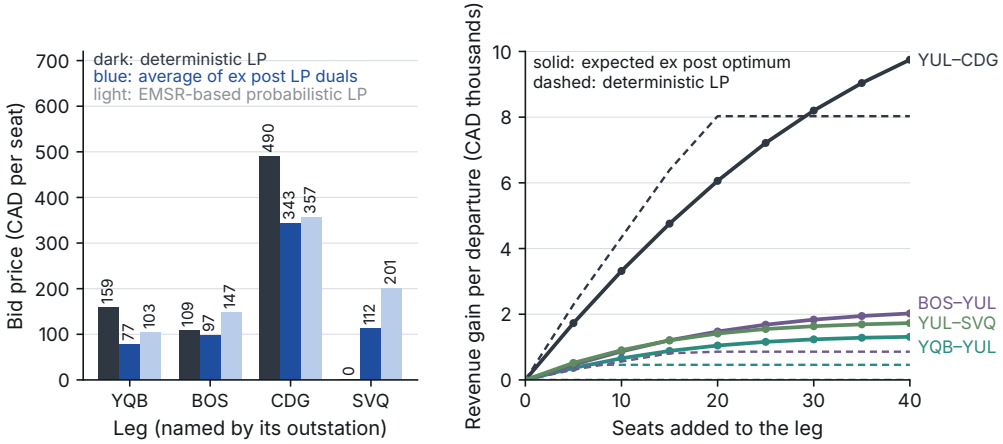


Figure 20.3 Bid prices by leg. Left: three estimates of the value of a seat on each leg: the deterministic LP, the average of ex post LP duals over random demand, and the EMSR-based probabilistic LP. Right: revenue gained by adding seats to one leg, for the deterministic LP (dashed) and for the expected ex post optimum under random demand (solid). The deterministic curves are steep and then flat; under uncertainty the value of capacity declines gradually.

with the network's bid prices in the place of the lower fare. For Laurentian, with normal demands, the multipliers are CAD 103, 147, 357 and 201 for YQB–YUL, BOS–YUL, YUL–CDG and YUL–SVQ. Seville now has a substantial bid price, and in fact too large a one: the program gives every product its own partition of seats with its own safety margin, while real inventories are nested and share one margin. This probabilistic nonlinear program and its nested variants are among the network formulations that Williamson (1992) compared by simulation; Curry (1990) combined an O&D linear program with nested protection on each leg.

The second refinement is the **randomized linear program** of Talluri and van Ryzin (1999): draw many demand scenarios, solve the LP for each with the realized demand in place of the mean, and average the dual prices. For Laurentian, 4,000 scenarios give average bid prices of CAD 77, 97, 343 and 112. The Seville leg is binding in 20 percent of scenarios, with an average dual of CAD 560 when it is, and the average of CAD 112 reflects both facts. YQB–YUL, binding in 57 percent of scenarios, falls from 159 to 77. Because each scenario's LP is solved with hindsight, the average ex post revenue is itself informative: CAD 358,465 per bank, 2.05 percent below the DLP value, which by Result 20.3 it must be. The gap is the cost of uncertainty even for a clairvoyant.

Figure 20.3 compares the three estimates. All agree that the Paris leg is by far the scarcest, and all disagree about Seville. Two further

families complete the picture conceptually. The exact solution of network revenue management is a dynamic program whose state is the number of seats remaining on every leg; the true opportunity cost of selling product j at time t is the drop in the value function, $V_t(n) - V_t(n - a_j)$, where n is the vector of seats remaining on each leg and a_j the column of incidences of product j, and bid prices approximate this difference by a sum of leg prices (Talluri and van Ryzin 2004b). The state space makes the exact program intractable for real networks, which is why approximations matter: Bertsimas and Popescu (2003) use LP-based approximations of the opportunity cost inside a dynamic acceptance rule, and MIT's probabilistic bid prices (PROBP) come from a nested probabilistic network convergence method (Belobaba 2002). The lecture material names the method without describing it; the idea, as it is generally described, is to prorate each itinerary fare across its legs using the other legs' current bid prices, compute leg EMSR curves on the prorated values, and iterate until the bid prices settle.

How much do these differences matter for revenue? Only a simulation of the booking process can say.

WORKED EXAMPLE 20.4 Five controls on one bank

Simulate 500 departures of Laurentian's bank. Demand for each product is normal with its case mean and standard deviation, rounded and truncated at zero. Requests arrive one at a time over a booking horizon: restricted (L) requests uniformly over its first 80 percent, flexible (H) requests uniformly over its last 60 percent, so that cheap demand tends to book first. Every control sees the same requests in the same order (common random numbers) and is re-optimized at evenly spaced reading points from the remaining capacity and the expected remaining demand:

- first come, first served (FCFS): accept while seats remain on every leg of the itinerary;
- leg/class EMSR-b: each leg protects seats for its H class against its L class, valued at the leg's local fares, and a connection needs its class open on both legs;
- DAVN: displacement costs from the DLP, products nested on each leg by network value (20.1) with EMSR-b protection levels;

- bid prices from the DLP, computed once, or re-solved 8 or 32 times;
- bid prices from the probabilistic program of Result 20.4, re-solved 32 times.

The ex post optimum, the LP solved with each departure's realized demand, bounds what any control could achieve.

Table 20.5 and Figure 20.4 report the results.

Four results stand out. First, leg/class control, DAVN and the bid prices re-solved 32 times all beat first come, first served by 3.3 to 4.4 percent, mostly by protecting seats for flexible passengers who book late: FCFS sells 35.6 YUL–CDG H passengers per bank on average out of 45 demanded, leg/class control 44.0. Second, bid prices computed once and never revised earn exactly the FCFS revenue, departure by departure. No fare lies below its path bid price, and the three marginal products sit exactly at it; since rule (20.2) accepts a fare equal to its bid price, the static rule accepts every request while seats remain, which is first come, first served. With a strict inequality it would refuse all demand for the three marginal products and earn 7.0 percent less than FCFS. Either way, a static bid price cannot accept part of a product's demand. Third, re-solving is what makes bid prices work: eight revisions leave them below leg/class control, thirty-two put them 0.6 percent above it, close to the probabilistic bid prices. Fourth, DAVN does best here, 1.07 percent above leg/class control and within 1.4 percent of the clairvoyant optimum. DAVN's advantage shows in the product mix: it sells 26.3 BOS–CDG L connections per bank, close to the DLP allocation of 26, against 34.6 under leg/class control, and uses the freed Paris seats for 161.0 local YUL–CDG L passengers instead of 153.9. As shares of the ex post optimum, FCFS earns 94.5 percent, leg/class control 97.6, re-solved bid prices 98.2 and DAVN 98.6.

The DLP bound of CAD 365,949 lies 2.1 percent above even the clairvoyant optimum, as Result 20.3 requires; it is a bound, not a target.

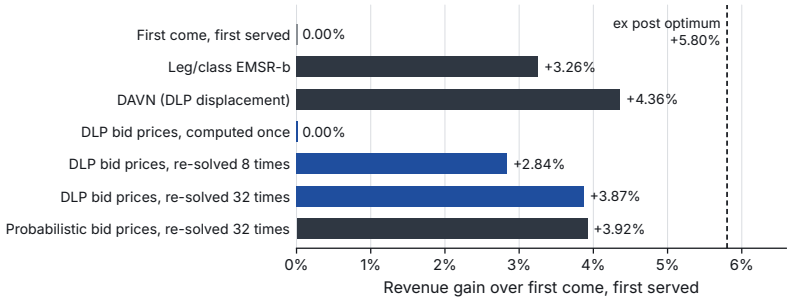


Figure 20.4 Simulated revenue gain over first come, first served for each control on Laurentian's bank; the dashed line is the ex post optimum with perfect knowledge of demand. Bid prices from the deterministic LP (blue) add nothing if never revised and are competitive only when re-solved often.

Table 20.5 Simulated revenue per bank (CAD) of booking controls on Laurentian's network (500 departures), and the gain over first come, first served and over leg/class control. Standard errors of the paired differences are about 0.1 percentage point against FCFS and 0.05 against leg/class control. The ex post optimum is estimated on the same 500 departures (the 4,000 scenarios of the randomized LP give CAD 358,465). The DLP value, CAD 365,949, bounds every row.

Control	Revenue	vs FCFS	vs leg/class
First come, first served	338,844	0.00%	−3.15%
Leg/class EMSR-b	349,877	+3.26%	0.00%
DAVN	353,625	+4.36%	+1.07%
DLP bid prices, once	338,844	0.00%	−3.15%
DLP bid prices, 8 solves	348,469	+2.84%	−0.40%
DLP bid prices, 32 solves	351,964	+3.87%	+0.60%
Probabilistic, 32 solves	352,131	+3.92%	+0.64%
Ex post optimum	358,503	+5.80%	+2.47%

PITFALL

Static bid prices. A bid price says whether a product is worth selling, not how many units to sell. Products whose fares equal their path bid price, which the LP accepts only in part, are accepted without limit until the next revision, so infrequently revised bid prices behave like first come, first served on exactly the products that matter. Either revise often or pair bid prices with limits on the marginal products.

20.8 What the PODS simulations found

The simulation of Example 20.4 has passengers who never switch flights or fare products. MIT's Passenger Origin-Destination Simulator (PODS), developed with a consortium of large airlines, goes much further: airlines forecast demand from their own booking histories and optimize their controls, and simulated passengers choose among airlines, paths and fare products according to schedules, prices, restrictions and availability (Belobaba 2002). Its results in the early 2000s shaped how airlines invested in network revenue management. Table 20.6 transcribes the main comparison; all values are historical simulation results, read from the bar charts of Belobaba's 2002 presentation, which the 2006 lectures reproduce (Belobaba 2006n, 2002).

Table 20.6 Revenue gains of O&D control over leg/class revenue management for airline A in PODS (historical simulation results, adapted from Belobaba, 2002 and 2006; values read from bar charts, accurate to about 0.02 percentage point). The last column is pure bid-price control with deterministic LP bid prices.

Network load factor	HBP	DAVN	PROBP	Bid prices from the DLP
70%	0.08%	0.28%	0.34%	about 0%
78%	0.21%	0.79%	0.88%	0.16%
83%	0.57%	1.36%	1.39%	0.34%
87%	0.94%	1.86%	1.78%	0.48%

Five findings carry over to today's thinking.

Gains are 1 to 2 percent and rise with load factor. Effective O&D control added 1 to 2 percent of network revenue to the 4 to 6 percent already earned by leg-based control and overbooking, with most of the gain at high load factors (Figure 20.5). Gains require substantial connecting traffic, high demand factors on at least some feeder legs and demand variability (Belobaba 2006n). Laurentian's simulated

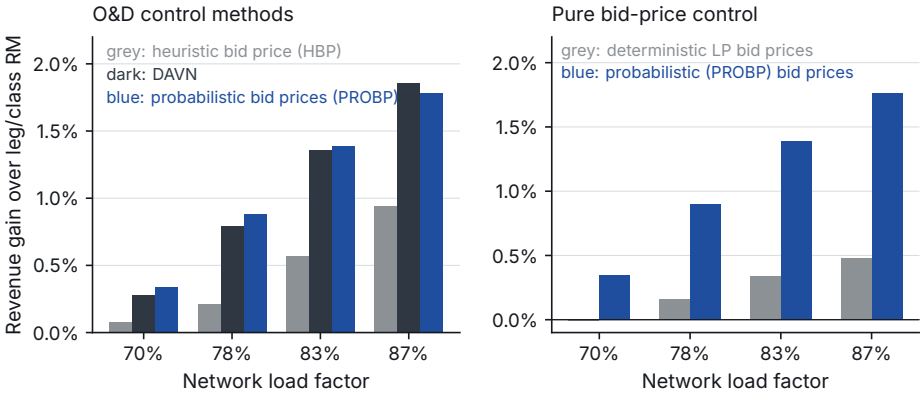


Figure 20.5 Historical PODS simulation results, redrawn (Belobaba, 2002 and 2006). Left: revenue gains of three O&D control methods over leg/class revenue management, by network load factor. Right: pure bid-price control with deterministic LP bid prices and with probabilistic (PROBP) bid prices.

1.07 percent for DAVN, on a network with a quarter of its demand connecting and three of four legs oversubscribed by 3 to 12 percent, sits in the same range.

The mechanism matters less than the optimizer. Value buckets and bid prices gave similar revenue in most simulations (Belobaba 2002). DAVN was robust to the source of its displacement costs: deterministic LP and probabilistic displacement costs gave gains within 0.05 percentage point. Pure bid-price control was not: deterministic LP bid prices gained at most 0.48 percent, probabilistic bid prices up to 1.76 percent, because deterministic LP bid prices are much more volatile and drive accept-or-reject decisions directly (Belobaba 2002); as Section 20.6 shows, an LP dual is built from the fares of the marginal products, so it jumps whenever the marginal product changes. Example 20.4 shows the same sensitivity through the frequency of re-solving.

Forecasting matters as much as optimization. At a 78 percent load factor, replacing booking-curve detruncation and pickup forecasting by projection detruncation and regression forecasting roughly doubled the gains (from 0.44 to 0.98 percent for HBP, 0.80 to 1.60 for DAVN, 0.90 to 1.54 for PROBP; the slides do not explain why the baseline HBP gain here differs from the 0.21 percent of Table 20.6). The forecasts were not demonstrably more accurate; more aggressive detruncation produced larger forecasts and therefore more protection for high-value passengers (Belobaba 2002).

Not adopting is costly. When airline A adopted O&D control at an 83 percent load factor and its competitor B kept basic leg/class

control, A gained 0.57 to 1.39 percent depending on the method and B lost 0.12 to 0.72 percent. The game was not zero-sum, since A's gains exceeded B's losses, and other simulations showed both airlines gaining when both adopted (Belobaba 2002).

Gains erode in the real world. When travel agents and websites booked two local segments to bypass an unavailable connection and priced them at the through fare, DAVN's gain at an 83 percent load factor fell from 1.37 percent to 1.00 percent with a 25 percent probability of such abuse and to 0.56 percent at 50 percent. Removing advance purchase or Saturday-night restrictions cost 8 to 14 percent of revenue under every method; O&D control softened the loss but could not replace fare fences (Belobaba 2002). For alliances, sharing displacement costs or bid prices between partners raised their joint gains, which uncoordinated controls could even turn into losses for one partner (Belobaba 2002).

20.9 Bid prices as the value of capacity

The handbook gives the planner's reading of this chapter in one line: it defines a bid price as an estimate of what scarce capacity is worth in its best alternative use, and states the simplified network rule of accepting an itinerary when its net revenue exceeds the sum of the bid prices on its path (Network Planning Handbook 2026). It adds the caveat that matters most: bid prices move with the remaining capacity, the time left, demand and the alternatives open to passengers, and they are neither fares nor accounting costs (Network Planning Handbook 2026). Read this way, Laurentian's bid prices answer three planning questions.

Which legs are scarce? The Paris leg, by a wide margin: every estimate puts its seats at CAD 340 to 490, against roughly 80 to 160 for the feeders. The feeders' bid prices are low not because they are empty, but because the passengers at their margin pay little. Seville is the instructive case: worthless in the deterministic LP, worth about CAD 90 to 110 a seat once the fifth of departures on which it fills is taken into account.

What is one more seat worth? On the margin, the bid price. Result 20.2 guarantees that the DLP revenue rises by at most π_l per seat added, and the right panel of Figure 20.3 shows how quickly that marginal value can collapse.

What is a larger aircraft worth? Here the bid price is the wrong tool, and the reason is worth working through.

Table 20.7 Revenue value per bank (CAD) of extra seats on Laurentian's network under three methods: the naive bid price times the number of seats, the deterministic LP re-solved with the new capacity, and the expected gain of the LP solved with realized demand, averaged over random demand scenarios (standard errors below CAD 130).

Capacity change	Naive	DLP re-solved	Ex post
YQB–YUL, 137 to 169	5,088	459	1,257
BOS–YUL, 137 to 169	3,488	859	1,833
YUL–CDG, 30 more	14,700	8,030	8,085
YUL–SVQ, 10 more	0	0	905

WORKED EXAMPLE 20.5 Valuing an upgauge with bid prices

Suppose the planner considers replacing the A220-300 (137 seats) on a feeder by a 737-8 (169 seats), and wants the revenue side of the case. Table 20.7 compares three valuations.

The naive valuation, the YQB–YUL bid price of 159 times 32 seats, says CAD 5,088 per bank. Re-solving the DLP with 169 seats says CAD 459: the first seat admits the one refused YQB–YUL L passenger (159), the next three admit the three refused YQB–CDG L connections, each worth only $649 - 549 = 100$ because it now displaces a local Paris passenger, and after that the leg's bid price is zero. The extra 28 seats are worthless in the deterministic world. Under random demand they are not: on departures with high Quebec City demand they carry passengers, and the expected ex post value is CAD 1,257.

The Boston upgauge shows the network effect in its purest form. Re-solving the DLP, the 32 extra seats let all 42 BOS–CDG L connections through, 16 more than before, and the Paris leg makes room by refusing 15 local YUL–CDG L passengers and one YQB–CDG L, whose feeder seat goes to the last YQB–YUL L passenger. The network gains the difference between the fares, not the fares: $16 \times 599 - 15 \times 549 - 649 + 159 = 859$.

Now compare with cost. With Laurentian's illustrative cost model (Chapter 3), a 737-8 costs about CAD 1,400 more per departure than an A220-300 on the 0.9-hour YQB–YUL sector and about CAD 1,900 more on the 1.4-hour BOS–YUL sector, plus CAD 20 to 30 per extra passenger. Even the expected ex post values, which assume revenue management uses the extra seats perfectly, fall short of these costs. Neither upgauge pays

on this bank, and the naive bid-price valuation would have approved both by a wide margin.

The lesson generalizes. A bid price is a derivative, valid for a few seats; capacity decisions are discrete and large, and by Result 20.2 the derivative can only overstate the deterministic LP's value of adding capacity and understate its loss from removing it. The deterministic LP errs in both directions: it overvalues the first seats added to a leg whose mean demand just exceeds capacity (YUL–CDG: about CAD 400 a seat for the first twenty seats, then nothing) and undervalues legs whose mean demand fits (YUL–SVQ). Value capacity changes by re-solving the network model under random demand, and ask the revenue management team for exactly that. The handbook's questions for revenue management are the right ones: which cabins and departures constrain demand, what incremental capacity is worth, where rejected demand would be recaptured, and how fares and availability would change after a schedule adjustment (Network Planning Handbook 2026). Chapter 22 turns these questions into a working procedure for valuing schedule and gauge changes with revenue management outputs, and Chapter 14 shows how fleet assignment uses spill-adjusted values of the same kind.

IN PRACTICE **Network controls in 2026**

The vocabulary of this chapter has outlived the distribution technology it was built for. IATA describes the legacy constraint as a maximum of 26 booking classes, one letter each, and defines the newer layers: dynamic pricing that sets prices from the shopping context, continuous pricing with an unlimited number of price points, and dynamic bundling of the offer (International Air Transport Association 2021). In an offer-based world the network question remains: what does a request for this itinerary displace? In the theory of network pricing, the optimal price for a request maximizes the expected margin of its fare over the opportunity cost of the seats it consumes, which is the itinerary bid price (Talluri and van Ryzin 2004b); continuous pricing changes how the fare is set, not why the opportunity cost matters. Airlines rarely say which network optimizer they run. For Air Canada, a vendor case study confirms the use of a PROS real-time dynamic pricing product (PROS 2025); the public sources reviewed for this book

do not say which forecasting and O&D optimization engines sit behind it. A planner should therefore ask the revenue management team which bid prices, displacement costs or opportunity costs its system produces, at what level (leg, cabin, departure, day before departure), and how often they are recomputed.

PITFALL

Treating bid prices as costs. A bid price is an opportunity value that moves with remaining capacity, time to departure and demand, not an accounting cost of carrying a passenger. Averaging bid prices over a season and adding them to a route's cost base double counts the revenue they already represent.

20.10 Summary

- Leg/class control fails on a network in two ways: congested feeders block valuable connections, and connecting passengers displace more local revenue than they pay when both legs are full.
- O&D control gives different itineraries different availability on the same flight according to their network value; value buckets by total fare are greedy, and displacement-adjusted virtual nesting charges each product for the seats it uses elsewhere.
- Bid-price control accepts a request when its fare covers the sum of the leg bid prices along the itinerary; value buckets with displacement implement the same rule with more machinery.
- The deterministic LP over legs and fare products has dual prices that are bid prices: by complementary slackness, products above their path bid price are fully accepted, products below it rejected, and products at it are marginal.
- The LP value is concave in capacity, bid prices are its supergradients, and the LP value bounds the expected revenue of every policy, even a clairvoyant one.
- On Laurentian's bank the bid prices are CAD 159, 109, 490 and 0; the network refuses part of its restricted Paris connections and accepts all Seville connections.
- In the simulation, bid prices that are never revised add nothing over first come, first served; re-solved often, they beat leg/class control, and DAVN does best, about 1 percent above leg/class control, in line with the 1 to 2 percent found by PODS.

- For planners, bid prices identify scarce legs and value small changes; value large capacity changes by re-solving the network model under random demand, never by multiplying a bid price by a seat count.

Table 20.8 Key formulas of the chapter.

Quantity	Formula
Network value on leg l	$v_{jl} = f_j - \sum_{l' \neq l} a_{l'j}\pi_{l'}$
Bid-price rule	accept j if $f_j \geq \sum_l a_{lj}\pi_l$
Deterministic LP	$\max \sum_j f_j x_j$ s.t. $\sum_j a_{lj}x_j \leq C_l$, $0 \leq x_j \leq \mu_j$
Dual	$\min \sum_l C_l\pi_l + \sum_j \mu_j\beta_j$ s.t. $\sum_l a_{lj}\pi_l + \beta_j \geq f_j$
Marginal value	$V(C') \leq V(C) + \sum_l \pi_l(C'_l - C_l)$
Upper bound	$\mathbb{E}[\text{revenue}] \leq V(C)$ for every policy
Probabilistic program	$f_j \Pr(D_j > x_j) = \sum_l a_{lj}\pi_l$
Heuristic bid price	$\max(\text{EMSR}_1, \text{EMSR}_2) + d\,\min(\text{EMSR}_1, \text{EMSR}_2)$

20.11 Further reading

Belobaba's lecture on O&D control (Belobaba 2006n) is the source of Tables 20.1 and 20.2 and Example 20.2, and his review of a decade of PODS research (Belobaba 2002) is the source of Table 20.6; both are short and worth reading in full. Smith, Leimkuhler and Darrow (1992) describe virtual nesting at American Airlines. Curry (1990) and Williamson (1992) are the classic studies of network formulations and their revenue impacts. Talluri and van Ryzin analyze bid-price controls (1998) and propose the randomized LP (1999); their book (2004b, ch. 3) is the standard reference for network capacity control, including the dynamic programming view. Bertsimas and Popescu (2003) develop approximate dynamic programming rules built on LP opportunity costs. Belobaba, Odoni and Barnhart (2015) place network revenue management within airline practice.

CHAPTER

21

Willingness to Pay, Sell-Up and Competition

IN THIS CHAPTER

- Why the revenue management of Chapters 18 and 20 breaks when fences weaken: its forecasts assume that every booking class has its own demand.
- Sell-up and buy-down, the exponential sell-up model and its FRAT5 parameter, applied to Laurentian's YUL–SVQ fare ladder.
- Q-forecasting: recovering demand by willingness to pay from bookings that availability has censored.
- The spiral-down effect, demonstrated with a simulation of a two-class flight.
- Fare adjustment, the marginal revenue transformation that lets standard optimizers handle buy-down, and its generalization to choice-based revenue management with efficient sets.
- Dynamic pricing as intensity control, and what a continuous price adds to a fare ladder.
- Willingness to pay in a competitive market: what the PODS simulations found, and why a rival's availability shapes your own sell-up.

DRAWS ON Belobaba, Willingness to pay and competitive revenue management (MIT 16.75J, 2006); Belobaba, Pricing challenges: ePODS and reality (2006); Cooper, Homem-de-Mello and Kleywegt (2006); Fiig, Isler, Hopperstad and Belobaba (2010); Talluri and van Ryzin (2004); Gallego and van Ryzin (1994)

Chapters 18 and 20 rested on one assumption: each booking class has its own demand, which does not depend on what else is for sale. Fences make the assumption roughly true. A traveller who must stay over a Saturday night to use the cheap fare and cannot, buys the expensive one; a leisure traveller who cannot afford the expensive

one does not ask for it. Laurentian's YUL–SVQ ladder (Section 17.7) still separates its five classes by advance purchase, change and refund conditions and bags, but the differences between Y and B, or between M and Q, are small. A customer who finds Q closed may pay for M; one who finds M open may never look at B. Demand for a class then depends on which classes are open, and a system that forecasts class by class misreads its own history.

For a network planner the consequences are concrete. The revenue that a new route or a larger aircraft will earn depends on the fare mix revenue management can achieve, which depends on customers' willingness to pay and on the fences (Chapter 17). The value of capacity depends on how customers respond to availability: the buy-down that extra seats can trigger is the reason Chapter 22 warns against valuing seats at average fares. And the demand data that reach the planner have passed through revenue management's controls: forecasts corrupted by spiral down understate high-fare demand and overstate low-fare demand.

The chapter starts with willingness to pay and the exponential sell-up model, then shows how to forecast demand by willingness to pay, why ignoring it makes revenue management spiral down, and how fare adjustment and choice-based optimization repair the optimizer. It then treats prices as a continuous control and closes with competition. The PODS passenger model behind many of the results is described in Chapter 6, the history of fences and simplified fares in Chapter 17, and continuous pricing in current practice in Chapter 23.

21.1 Willingness to pay

The demand curves of Chapter 4 aggregate many individual decisions. Revenue management needs the individual view, because it decides which fares a particular customer can see.

DEFINITION 21.1 Willingness to pay

A customer's **willingness to pay** W for a product is the highest fare at which she buys it rather than take her best alternative: another airline, another day, another mode or no trip at all. Among the customers who consider a flight in a given booking period, W has a distribution. Its survival function $S(f) = \Pr(W \geq f)$ is the share who would buy at fare f, and if N customers consider the flight, expected demand at fare f is $N\,S(f)$.

Three facts about this distribution organize the chapter. First, its shape fixes the price elasticity: with $D(f) = N\,S(f)$, the elasticity of demand at fare f is $\eta(f) = fS'(f)/S(f)$. The logit models of Chapter 5 express the same information through utility coefficients, and the willingness to pay for an attribute is a ratio of coefficients. Second, Laurentian's ladder is a discrete reading of it. Of the 166 customers per departure who would buy at the V fare of CAD 419, 12 would pay the Y fare of CAD 1,249, and the corners of the ladder in Figure 17.5 trace a curve that halves every CAD 219. Third, the distribution changes over the booking period, because the customers who book late are different people. A chart of the PODS research in Belobaba's 2006 lecture shows willingness to pay relative to the lowest fare rising across 16 booking periods, from about 1.2 to 3.0 in a medium scenario, from 1.8 to 5.0 in a high one and from 1.05 to 1.5 in a low one, most steeply between the eighth and the eleventh period (historical simulation settings; values read from the chart) (Belobaba 2006a).

The lecture lists what moves willingness to pay: time to departure, because business travellers book late; peak periods; limited competition and low capacity relative to demand; and availability, the airline's own and its rivals' (Belobaba 2006a). The last item is the troublesome one. If an airline's past controls were weak, its history never shows the high fares its customers would have paid; if a rival kept cheap seats open, high-fare bookings could not be observed at all. Willingness to pay is a property of customers, but what the data reveal about it is a property of past availability.

The same lecture distinguishes two reasons for buying a fare. A booking is **product-oriented** if the next lower class was open when it was made: the customer chose the higher class for what it includes. It is **price-oriented** if the next lower class was closed: she paid the higher fare because the cheaper one was unavailable (Belobaba 2006a). Branded fares (Chapter 17) make both common. Flex bought while Standard was open is a product choice; Standard bought because Basic was closed is a price choice. The two need different forecasts, and the rest of the chapter is about the second.

21.2 Sell-up and buy-down

When the fare a customer wants is closed, she has four options: pay more on the same flight, take another of the airline's flights, go to a competitor, or not travel. Chapter 9 defined recapture and spill; two more terms complete the vocabulary.

DEFINITION 21.2 Sell-up and buy-down

Sell-up occurs when a customer whose preferred fare product is closed buys a more expensive product of the same airline instead. **Buy-down** occurs when a customer who would buy a more expensive product if the cheaper one were closed buys the cheaper one because it is open. Under **perfect fences** buy-down is zero; with **no fences**, every customer buys the cheapest open fare she is willing to pay.

Classical forecasts assume the first world and much of the market lives between the two. In the no-fence limit the demand for a class is the number of customers willing to pay its fare while it is the lowest open class, and the class forecasts of Chapter 18 lose their meaning. A model of sell-up is then a model of the willingness-to-pay distribution above the lowest fare f_n, and one family is standard in the PODS research (Cléaz-Savoyen 2005; Fiig et al. 2010).

RESULT 21.1 The exponential sell-up model

Suppose that, among customers willing to pay the lowest fare f_n, the excess $W - f_n$ is exponential with mean f_n/γ. Then for $f \geq f_n$ the **sell-up probability** is

$$p(f) = \Pr(W \geq f \mid W \geq f_n) = e^{-\gamma(f/f_n - 1)}, \qquad \gamma = \frac{\ln 2}{\text{FRAT5} - 1}, \tag{21.1}$$

where FRAT5 is the fare ratio f/f_n at which half of the customers sell up. Moreover:

1. sell-up is memoryless: for $f' \geq f \geq f_n$, $\Pr(W \geq f' \mid W \geq f) = e^{-\gamma(f'-f)/f_n}$;
2. the price elasticity of demand at fare f is $\eta(f) = -\gamma f/f_n$;
3. without a capacity constraint, revenue per customer $f\,p(f)$ is maximized at $f^* = \max(f_n, f_n/\gamma)$.

Proof. An exponential excess with mean f_n/γ has survival function $\Pr(W - f_n \geq u) = e^{-\gamma u/f_n}$; put $u = f - f_n$. Setting $p = 1/2$ at $f = \text{FRAT5} \cdot f_n$ gives $\gamma(\text{FRAT5} - 1) = \ln 2$. Part 1 is the ratio $p(f')/p(f)$. Part 2: $p'(f) = -(\gamma/f_n)\,p(f)$, so $\eta = fp'/p = -\gamma f/f_n$. Part 3: $\mathrm{d}[fp(f)]/\mathrm{d}f = p(f)(1 - \gamma f/f_n)$, which vanishes at $f = f_n/\gamma$ and is negative beyond it; if $f_n/\gamma < f_n$ the derivative is negative on the whole range $f \geq f_n$, and the lowest fare is best. ∎

The name FRAT5 comes from the PODS literature (Cléaz-Savoyen 2005). The PODS willingness-to-pay function of equation (6.5) is exactly (21.1) with the base fare as f_n and the multiplier m as FRAT5: the leisure value $m = 1.2$ means that half of leisure customers would pay 20 percent more than the Q fare. Laurentian's ladder implies FRAT5 $= 1 + 219/419 = 1.52$ on average, $\gamma = 1.33$. With that value the model reproduces the 12 customers who would pay Y and predicts 34, 66 and 110 at B, M and Q against the ladder's 30, 58 and 102: the ladder's curve is close to exponential, not exactly so. The exponential form is a convenience, not a law, but its single parameter can be estimated from the closure patterns of Section 21.3, and the memoryless property makes the arithmetic of sell-up between any two classes the same.

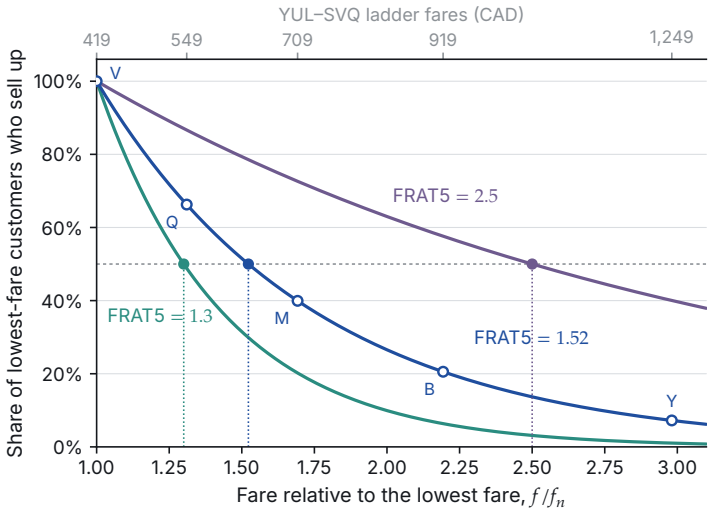


Figure 21.1 Exponential sell-up curves (21.1): the share of the customers willing to pay the lowest fare f_n who would also pay fare f, for three values of FRAT5, the fare ratio at which half of them sell up (dots on the dotted line). FRAT5 $= 1.3$ describes an early, leisure booking period, FRAT5 $= 2.5$ a late, business one, and FRAT5 $= 1.52$ the average implied by Laurentian's ladder, whose five classes are marked by open circles; the top axis gives the ladder fares.

WORKED EXAMPLE 21.1 **Sell-up on the Seville ladder**

Take Laurentian's YUL–SVQ economy ladder and two booking periods, an early one dominated by leisure customers with FRAT5 $= 1.3$ ($\gamma = 2.31$) and a late one with FRAT5 $= 2.5$ ($\gamma = 0.462$). Suppose no fences separate the classes, so that every customer buys the cheapest open fare she is willing to pay. Table 21.1 gives the sell-up probabilities $p_j = p(f_j)$ and the expected revenue per customer, $f_j p_j$, when class j is the lowest open class and seats are plentiful.

Table 21.1 Sell-up probabilities and expected revenue per customer (CAD) on Laurentian's YUL–SVQ ladder with no fences, for an early ($\mathrm{FRAT5} = 1.3$) and a late ($\mathrm{FRAT5} = 2.5$) booking period, when the class in the row is the lowest open class.

Class	Fare	f_j/f_n	p_j, early	$f_j p_j$, early	p_j, late	$f_j p_j$, late
Y	1,249	2.981	0.010	12.8	0.400	500.1
B	919	2.193	0.063	58.3	0.576	529.5
M	709	1.692	0.202	143.3	0.726	514.9
Q	549	1.310	0.488	268.1	0.866	475.7
V	419	1.000	1.000	419.0	1.000	419.0

Early, opening V maximizes revenue per customer: only 49 percent of V customers would pay the Q fare, so each step up the ladder loses more customers than it gains in fare. The static optimum of Result 21.1 is $f_n/\gamma = 181$, below the floor, so the lowest fare is best. Late, the best lowest open class is B, at CAD 529.5 per customer, and opening Q or V loses revenue even on an empty flight: the static optimum is $f_n/\gamma = 907$, close to the B fare of 919. Memorylessness is visible in the late column: the share of V customers who would pay Q, 0.866, is also the share of B customers who would pay CAD 1,049, CAD 130 more than B.

The same fares, on the same flight, call for opposite availability at different times. A system that opens V late because V is the class with the most bookings in its history is selling at 419 to customers who would pay 919.

PITFALL

Reading class bookings as class demand. When fences are weak, the absence of Y bookings says nothing about how many customers would pay the Y fare if nothing cheaper were open; it says only that something cheaper was open. Sell-up must be estimated from departures on which the cheaper classes were closed.

21.3 Q-forecasting

How can an airline learn willingness to pay from data that availability has filtered? Belobaba's lecture describes the approach: transform every historical booking into the maximum demand potential at the lowest fare, then convert that potential back into forecasts by fare level using estimates of sell-up (Belobaba 2006a). The method is known in the PODS research as **Q-forecasting**, after Q, the lowest class of the PODS fare structures (Cléaz-Savoyen 2005). Its statistical core is a thinning argument.

Fix one booking period of a future departure, say 35 to 21 days before departure, and let the number of customers in that period who are willing to pay the lowest fare f_n be Poisson with mean λ. On past departure k, the lowest open class during the period was $j(k)$. With no fences, each customer books exactly when $W \geq f_{j(k)}$, which happens with probability $p_{j(k)}$ independently of the others, so the bookings B_k observed on departure k are Poisson with mean $\lambda\, p_{j(k)}$.

RESULT 21.2 **The Q-forecast**

Given bookings $B_1, \dots, B_K$ observed on K past departures with lowest open classes $j(1), \dots, j(K)$ and known sell-up probabilities, the maximum likelihood estimate of the demand at the lowest fare is

$$\hat{\lambda} = \frac{\sum_k B_k}{\sum_k p_{j(k)}}, \tag{21.2}$$

which is unbiased with variance $\lambda / \sum_k p_{j(k)}$. The implied forecast of customers willing to pay at least f_j is $\hat{\lambda} p_j$, and the **partitioned forecast** for class j, the customers whose willingness to pay lies between f_j and the next higher fare, is $d_j = \hat{\lambda}\,(p_j - p_{j-1})$ with $p_0 = 0$.

Proof. The log-likelihood is $\ell(\lambda) = \sum_k [B_k \ln(\lambda p_{j(k)}) - \lambda p_{j(k)} - \ln B_k!]$, whose derivative $\sum_k B_k/\lambda - \sum_k p_{j(k)}$ vanishes at (21.2); the second derivative is negative. Since $\mathbb{E}[B_k] = \lambda p_{j(k)}$ and $\mathrm{Var}(B_k) = \lambda p_{j(k)}$ with independent departures, $\mathbb{E}[\hat{\lambda}] = \lambda$ and $\mathrm{Var}(\hat{\lambda}) = \lambda \sum_k p_{j(k)} / (\sum_k p_{j(k)})^2$. ∎

The estimator scales each departure's bookings up by the probability that a lowest-fare customer would have bought at the fare that was open, and weights departures by how informative they are: a departure on which only Y was open contributes little, because few customers buy there. Sell-up itself is estimated from the same data, because different departures close classes at different times: the ratio of bookings when M was the lowest open class to bookings when V was identifies p_M. The limit is plain. As the lecture puts it, "we can't estimate what we've never seen" (Belobaba 2006a): if Y was never the lowest open class in a booking period, sell-up to Y there can only be assumed.

In the hybrid forecasting tested in PODS, product-oriented bookings are forecast with standard time-series methods by class, and price-oriented demand is forecast by willingness to pay as above; the two are combined before optimization (Belobaba 2006a).

WORKED EXAMPLE 21.2 A Q-forecast for one booking period

Laurentian's history for one booking period on YUL–SVQ covers eight departures. The lowest open class and the bookings recorded were: V and 14, V and 11, Q and 9, Q and 8, M and 6, V and 13, Q and 10, M and 5. Take the ladder's average sell-up, FRAT5 = 1.52, so that $p_M = 0.399$, $p_Q = 0.663$ and $p_V = 1$.

The bookings sum to 76 and the sell-up probabilities to $3 \times 1 + 3 \times 0.663 + 2 \times 0.399 = 5.787$, so by (21.2)

$$\hat{\lambda} = 76/5.787 = 13.13,$$

with a standard error of $\sqrt{13.13/5.787} = 1.51$: about 13 customers in this period would book at the V fare if it were open. With $p = (0.072, 0.205, 0.399, 0.663, 1)$ for Y, B, M, Q and V, the partitioned forecasts are $d = (0.95, 1.75, 2.55, 3.46, 4.43)$: about one customer per departure in this period would pay the Y fare, and 2.7 would pay B or more.

Compare what a class-by-class forecaster sees. Averaged over the eight departures, bookings were 0 in Y and B, 1.38 in M, 3.38 in Q and 4.75 in V. It concludes that nobody in this period will pay more than 709. If it also unconstrains each class on the departures on which the class was closed, as the methods of Chapter 9 do, it estimates V demand from the three departures with V open at 12.7 per departure: it assigns to V nearly all the customers who would pay more. The lecture shows the same pattern in a PODS network: at the first booking period, standard forecasts summed to about 133 passengers per leg, of which about 81 were in the two lowest classes, against 87 passengers finally carried; forecasts by willingness to pay started near 102, with about 18 instead of 7 in the highest class (values read from the lecture charts) (Belobaba 2006a).

21.4 The spiral-down effect

Put the misreading of Example 21.2 inside the loop of a revenue management system (Figure 18.1). The system protects seats for the high fare according to its forecast of high-fare demand. Protection is too low, so the cheap fare stays open longer, and some customers who would have paid the high fare buy the cheap one. The system records fewer high-fare bookings, forecasts less high-fare demand and protects even fewer seats. Revenue falls, load factor rises, and the history

confirms the forecast at every step. This is the **spiral-down effect**, introduced with the ePODS experiments in Chapter 17: removing all fences left the structure with the highest load factor and 45 percent less revenue, and erased the advantage of leg revenue management over first come, first served (Figure 17.3).

Cooper, Homem-de-Mello and Kleywegt (2006) formalize the mechanism. They study an airline that sets protection levels with Littlewood's rule from past high-fare sales while customers buy the low fare whenever it is available, and show that protection levels, high-fare sales and revenue decrease systematically over successive rounds of forecasting and optimization. The simplest version of their setting can be analyzed in a few lines.

Consider a flight with C seats and two fares $f_1 > f_2$ with no fences between them. D_L customers are willing to pay f_2 but not f_1; D_H are willing to pay f_1. As in Littlewood's model (Section 18.4), the first group books first. Revenue management protects y seats for the high fare, so the low fare has booking limit $b = C - y$, and a high-fare customer who arrives while the low fare is open buys it.

RESULT 21.3 Why the spiral has no floor

In this model:

1. on every departure on which the cabin does not fill, observed high-fare sales equal $D_H - B$, where B, the number of buy-downs, is the smaller of D_H and $[C - y - D_L]^+$;
2. with $y = 0$ the high fare is never sold, so a forecaster that sets y by Littlewood's rule from observed high-fare sales never leaves $y = 0$.

Proof. Low-fare customers buy $\min(D_L, b)$ seats. When the high-fare customers arrive, $r = [b - D_L]^+$ low-fare seats remain. The first $B = \min(D_H, r)$ of them buy the low fare, and the rest buy the high fare while seats remain, so high-fare sales are $h = \min([D_H - r]^+, y)$. If $D_H < r$, then $B = D_H$ and $h = 0 = D_H - B$. Otherwise $B = r$, the low fare sells $\min(D_L, b) + r = b$ seats, and the cabin fills exactly when $h = y$; if it does not fill, $h < y$, so $h = D_H - r = D_H - B$. For part 2, $y = 0$ gives $b = C$: the low fare is open until the cabin is full, every high-fare customer who finds a seat buys it, and $h = 0$. Littlewood's rule applied to a forecast with zero mean and variance protects nothing. ∎

The first part says that the data are biased by exactly the buy-down; the second that the bias is self-reinforcing to the end. Nothing in the loop can recover, because a policy that never sells the high fare never generates the evidence that it could.

WORKED EXAMPLE 21.3 Spiral down on a two-class flight

Take a stylized version of the Seville cabin: $C = 168$ economy seats, the Y fare of CAD 1,249 and the Q fare of CAD 549 with no fences between them, $D_L \sim \mathcal{N}(115, 25^2)$ customers willing to pay only the Q fare and $D_H \sim \mathcal{N}(45, 12^2)$ willing to pay Y, rounded to whole customers (illustrative values, chosen so that mean demand, 160 customers, nearly fills the cabin). The system starts with a forecast of high-fare demand inherited from a fenced past, mean 45 and standard deviation 12, and Littlewood's rule $\Pr(D_H > y) = 549/1{,}249$:

$$y_0 = 45 + 12\,\Phi^{-1}(1 - 0.440) = 45 + 12 \times 0.152 = 46.8.$$

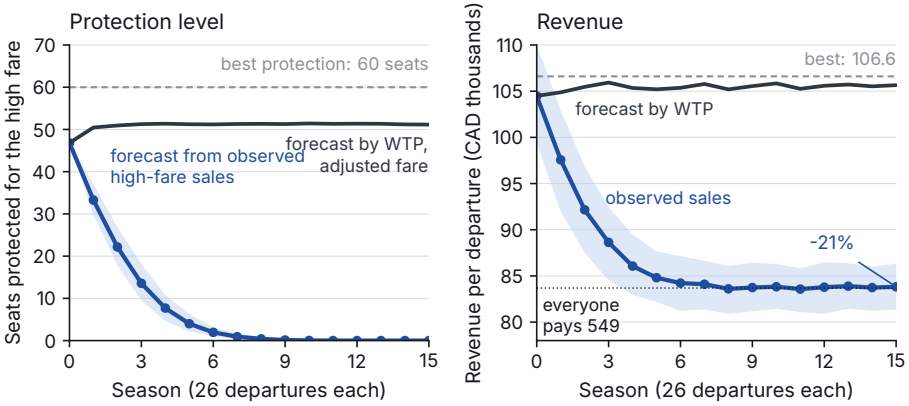


Figure 21.2 The spiral-down effect on the two-class flight of Example 21.3, season by season (26 departures per season, 300 simulated histories). Left: seats protected for the high fare. Right: revenue per departure. A system that forecasts high-fare demand from observed high-fare sales (blue, with the 10th to 90th percentile band) spirals to zero protection and sells every ticket at the low fare; one that forecasts by willingness to pay and uses the adjusted low fare (black) stays close to the best protection level.

With 47 seats protected, the Q fare has a booking limit of 121. Leisure demand falls short of it on most departures, and the business customers who arrive while Q is still open buy it: on average 12.2 of them. Expected high-fare sales are 30.8, with a standard deviation of 15.6. The system reads these sales as high-fare demand and re-applies Littlewood's rule: $y_1 = 30.8 + 0.152 \times 15.6 = 33.2$. The booking limit rises to 135, buy-down to 20.2, and the next round gives $y_2 = 22.1$, then 13.4, 7.2 and 3.6.

Figure 21.2 simulates the loop with 26 departures per season and the protection level re-estimated after each season. The average protection level falls from 46.8 to 33.3, 22.2, 13.6 and 7.7 seats over the first four seasons; by season 8, 93 percent of the simulated histories protect less than one seat, and by season 15 all of them do. Expected revenue per departure falls from CAD 104,500 to CAD 83,800, the revenue when every customer pays CAD 549 (152 passengers on average, not a full cabin).

The best protection level for this flight, found by evaluating expected revenue over a grid, is 60 seats, with expected revenue of CAD 106,603; zero protection earns 21.5 percent less. A system that forecasts by willingness to pay avoids the spiral. In the simulation it counts every booking made in the late period as high-willingness demand, whatever fare was paid, uncon-

strains leisure demand with the EM algorithm of Result 9.4, and applies Littlewood's rule to the adjusted low fare of Section 21.5. It settles at about 51 seats and earns CAD 105,500 per departure, 1 percent below the best policy and 26 percent more than the spiralled system.

With more leisure demand, $D_L \sim \mathcal{N}(140, 25^2)$, so that the cabin is oversubscribed, the spiral is slower but reaches the same end: the average protection level is 28.7 after four seasons and 5.7 after fifteen. Result 21.3 explains why: buy-down biases every season's data, whatever the demand level.

PITFALL

Celebrating a rising load factor. A load factor that climbs while yield and revenue fall is the signature of spiral down, not of a strong market. Belobaba notes that unadjusted systems produced high load factors without higher yields, and that many airlines fell back on manual overrides and rules (Belobaba 2006a, 2006o).

21.5 Fare adjustment

Q-forecasting repairs the forecast. The optimizer needs repair too: even with a correct forecast of how many customers would pay each fare, an optimizer built for independent demand assumes that a customer willing to pay Y will buy Y while Q is open. Fiig, Isler, Hopperstad and Belobaba (2010) showed that the problem with buy-down can be transformed into an equivalent problem with independent demand, so that existing optimizers can solve it. The transformation is easiest to see in the dynamic model of Section 18.6.

Modify that model in one respect. In each of the T short periods, counted by the number remaining, a customer arrives with probability $\lambda(t)$. The airline chooses which classes to open, and with no fences only the lowest open class matters: if it is j, the customer buys class j with probability Q_j, where Q_j is nondecreasing in j (opening a cheaper class never reduces sales), and pays f_j. Write $\mathrm{TR}_j = f_j Q_j$ for the expected revenue per arriving customer, and let $j = 0$ mean that all classes are closed, with $Q_0 = \mathrm{TR}_0 = 0$. With $\Delta V_{t-1}(x)$ the marginal value of a seat (Definition 18.4), the Bellman equation becomes

$$V_t(x) = V_{t-1}(x) + \lambda(t) \max_{0 \le j \le n} [\mathrm{TR}_j - Q_j \, \Delta V_{t-1}(x)]. \qquad (21.3)$$

Opening class j earns TR_j in expectation but uses a seat with probability Q_j, which costs $\Delta V_{t-1}(x)$.

DEFINITION 21.3 Adjusted fare

For classes ordered from the highest fare, the **adjusted demand** of class j is $q_j = Q_j - Q_{j-1}$ and its **adjusted fare** is

$$f'_j = \frac{\mathrm{TR}_j - \mathrm{TR}_{j-1}}{Q_j - Q_{j-1}},$$

the marginal revenue per additional sale from opening class j when class $j-1$ is already open. Replacing (f_j, Q_j) by (f'_j, q_j) is the **marginal revenue transformation**.

RESULT 21.4 The marginal revenue transformation

If the adjusted fares decrease along the ladder, $f'_1 \ge f'_2 \ge \dots \ge f'_n$, then in every period it is optimal to open exactly the classes whose adjusted fare is at least $\Delta V_{t-1}(x)$, and $V_t(x)$ is the value function of the independent-demand model (18.7) with fares f'_j and request probabilities $\lambda_j(t) = \lambda(t)\, q_j$. All the results of Result 18.4, including nested protection levels, apply to the adjusted fares.

Proof. Telescoping the differences,

$$\mathrm{TR}_j - Q_j \Delta = \sum_{i=1}^{j} [(\mathrm{TR}_i - \mathrm{TR}_{i-1}) - \Delta(Q_i - Q_{i-1})] = \sum_{i=1}^{j} q_i \, (f'_i - \Delta).$$

Each q_i is nonnegative. When the f'_i decrease, the terms are nonnegative up to the largest i with $f'_i \ge \Delta$ and negative afterwards, so the sum is maximized by stopping there, and its maximum is $\sum_i q_i \left[f'_i - \Delta\right]^+$. Substituting into (21.3) gives $V_t(x) = V_{t-1}(x) + \sum_i \lambda(t) q_i \left[f'_i - \Delta V_{t-1}(x)\right]^+$, which is (18.7) with fares f'_i and probabilities $\lambda(t) q_i$. Both recursions start from $V_0 \equiv 0$, so they produce the same value function and the same decisions. ∎

Three remarks make the result usable. A negative adjusted fare means a class that should never be opened, whatever the bid price: the revenue from the sales it adds is less than the revenue it loses through buy-down. If the adjusted fares do not decrease, some classes lie below the upper concave envelope of the points (Q_j, TR_j); they are never optimal, and dropping them before transforming restores the condition, the efficient frontier of Section 21.6. And with a continuum of fares and exponential sell-up, the adjusted fare has a closed form: with $Q = p(f)$ from (21.1),

$$\frac{\mathrm{d}(fQ)}{\mathrm{d}Q} = f + \frac{p(f)}{p'(f)} = f - \frac{f_n}{\gamma},$$

so fare adjustment subtracts a constant, the mean excess willingness to pay, from every fare. The discrete formula is the slope of the chord between neighbouring classes and approximates $f - f_n/\gamma$ between their fares.

Why does this cure the spiral? The adjusted fares charge each cheaper class for the revenue it loses through buy-down, so the optimizer closes cheap classes earlier and protects more for high-willingness customers. The Q-forecast supplies demand inputs that do not depend on past availability. Together they break both links of the loop in Section 21.4. In the two-class flight of Example 21.3, with segment means 115 and 45, the adjusted Q fare is

$$f_2' = \frac{549 \times 160 - 1{,}249 \times 45}{115} = 275.1,$$

half the published fare. Littlewood's rule with f_2' in place of f_2 protects $45 + 12\,\Phi^{-1}(1 - 275.1/1{,}249) = 54.3$ seats, and 54 seats earn 0.45 percent less than the best protection of 60 (this static model is not the dynamic one of Result 21.4, so the transformation is a good heuristic here, not an exact solution).

WORKED EXAMPLE 21.4 Adjusted fares on the Seville ladder

Use the sell-up probabilities of Table 21.1. In the early period (FRAT5 = 1.3), for Y to V,

$$Q = (0.010,\ 0.063,\ 0.202,\ 0.488,\ 1),$$
$$\mathrm{TR} = (12.8,\ 58.3,\ 143.3,\ 268.1,\ 419.0).$$

The adjusted fare of V is

$$f_{\mathrm{V}}' = \frac{419.0 - 268.1}{1 - 0.488} = 295,$$

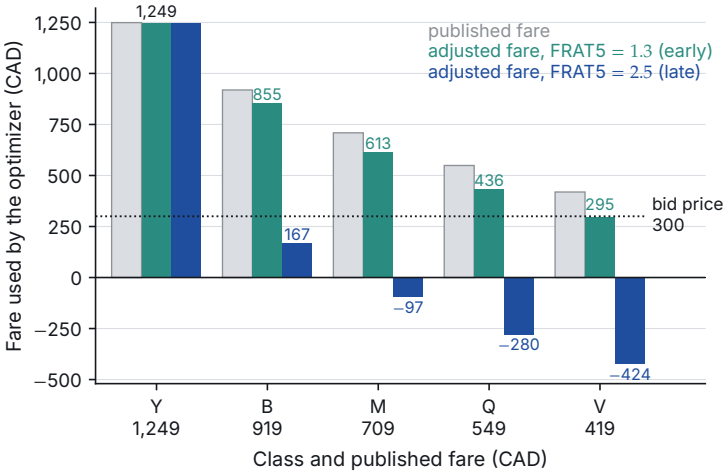


Figure 21.3 Published fares and adjusted fares on Laurentian's YUL–SVQ ladder with no fences, for an early booking period (FRAT5 = 1.3) and a late one (FRAT5 = 2.5). A class is opened when its adjusted fare exceeds the marginal value of a seat; with a bid price of 300 (dotted), the early period opens classes down to Q and the late period only Y. Negative adjusted fares mark classes that should never be opened.

and the full sets of adjusted fares, early and late (FRAT5 = 2.5), are

$$f'_{\text{early}} = (1{,}249,\ 855,\ 613,\ 436,\ 295),$$

$$f'_{\text{late}} = (1{,}249,\ 167,\ -97,\ -280,\ -424).$$

Both decrease, as Result 21.4 requires. Late in the booking period M, Q and V should never be open (Figure 21.3).

Suppose the marginal value of a seat is CAD 300. An optimizer that compares published fares with it opens everything down to V, since even 419 exceeds 300, and earns $\text{TR}_{\text{V}} - 300\,Q_{\text{V}} = 119.0$ per arriving customer. With adjusted fares, the early period opens down to Q, whose adjusted fare is 436 while V's is 295, and earns $268.1 - 300 \times 0.488 = 121.6$: a small gain, because early customers are price sensitive anyway. The late period opens only Y and earns $500.1 - 300 \times 0.400 = 379.9$ per customer, three times as much. Even on an empty flight, with a marginal value of zero, the late period should stop at B: CAD 529.5 per customer instead of 419, 26 percent more.

The continuous approximation $f - f_n/\gamma$ subtracts 181 early and 907 late; between the B and Y fares, for example, it gives 1,084 –

181 = 903 against the exact 855, close enough to see why the late period's adjustments are so large: late customers would pay, on average, 907 more than the lowest fare.

Real structures are mixed: some fences still work, and some demand is product-oriented. Fiig and colleagues apply the transformation to the price-oriented part of demand in structures that mix fenced and unfenced products, leaving product-oriented demand with its own fares (Fiig et al. 2010). Because the output is an ordinary set of fares and demands, it can feed the network optimizers of Chapter 20 as well: a path and class is then available when its adjusted fare exceeds the sum of the bid prices along the path. In PODS, adding a willingness-to-pay component to forecasting lifted the revenue gain of every method over standard EMSR-b to between 2.6 and 4.0 percent in a simplified fare structure (Section 21.8) (Belobaba 2006a).

21.6 Choice-based revenue management

Fare adjustment assumes a particular choice model: every customer buys the cheapest open fare she is willing to pay. Talluri and van Ryzin (2004a) solved the single-leg problem for any discrete choice model, and their solution shows what fare adjustment really is.

In each period a customer arrives with probability λ. The airline offers a set S of products from the n it could sell. Given S, the customer buys product $j \in S$ with probability $P_j(S)$ and nothing with probability $P_0(S)$; these can come from a logit model (Chapter 5), from the PODS rules (Chapter 6) or from the sell-up model above. Write $Q(S) = \sum_{j \in S} P_j(S)$ for the probability of a sale and $R(S) = \sum_{j \in S} f_j P_j(S)$ for the expected revenue per arriving customer. The Bellman equation is (21.3) with offer sets in place of lowest open classes:

$$V_t(x) = V_{t-1}(x) + \lambda \max_S [R(S) - Q(S)\,\Delta V_{t-1}(x)].$$

DEFINITION 21.4 Efficient set

An offer set S is **efficient** if no probability mixture of offer sets has a sales probability of at most $Q(S)$ and an expected revenue of at least $R(S)$, with at least one of the two inequalities strict. Plotted in the plane of (Q, R), the efficient sets are the corners of the upper concave envelope of all offer sets and of the empty set at the origin: the **efficient frontier**.

RESULT 21.5 Optimal offer sets are efficient

For any marginal value of a seat $\Delta \geq 0$, the maximum of $R(S) - Q(S)\Delta$ over offer sets is attained at an efficient set. Ordering the efficient sets $S_1, S_2, \ldots$ by increasing Q, their revenues increase and the slopes of the frontier,

$$\rho_k = \frac{R(S_k) - R(S_{k-1})}{Q(S_k) - Q(S_{k-1})},$$

decrease. The optimal set is S_k for $\rho_{k+1} \leq \Delta \leq \rho_k$: as the marginal value of a seat falls, the airline moves along the frontier toward sets that sell more.

Proof. Maximizing $R - \Delta Q$ over a finite set of points in the plane is maximizing a linear function; its maximum over the points equals its maximum over their convex hull, and that maximum is attained at a corner of the hull. Since $\Delta \geq 0$, the objective increases with R and decreases with Q, so the maximizing corner lies on the upper boundary between the origin and the point of highest revenue: an efficient set. Along this concave boundary the slopes decrease, and the line of slope Δ touches it at the corner where the slope changes from above Δ to below it. ∎

Talluri and van Ryzin also show that when the efficient sets are nested, as when each one adds a cheaper product to the previous one, the optimal policy is again a set of nested protection levels, and that under the multinomial logit model every efficient set consists of the k highest-fare products for some k, so the efficient sets are nested by fare (Talluri and van Ryzin 2004a). The fenceless sell-up model of Section 21.5 is a special case: its offer sets are "all classes down to j", its efficient frontier joins the points (Q_j, TR_j), and the slopes ρ_k

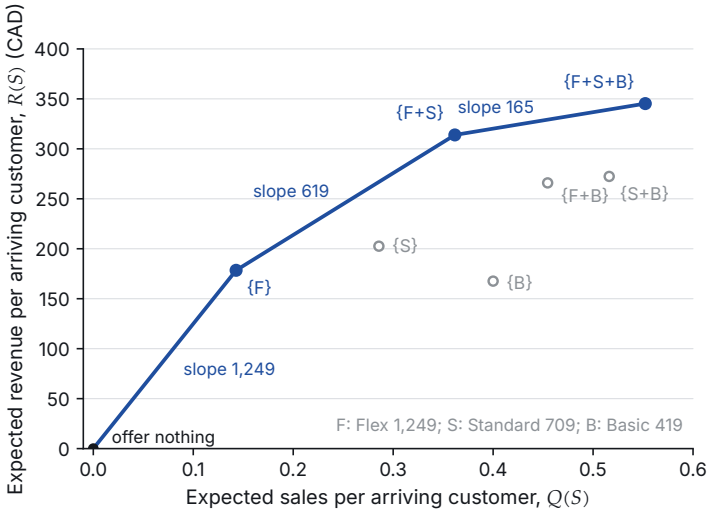


Figure 21.4 Expected revenue against expected sales per arriving customer for the seven offer sets of three Laurentian economy products under a multinomial logit model (Example 21.5). Filled points are efficient sets, joined by the efficient frontier; open points are inefficient. The slopes of the frontier are the marginal revenues that play the role of adjusted fares.

are exactly the adjusted fares of Definition 21.3. Fare adjustment is choice-based revenue management for the simplest choice model.

WORKED EXAMPLE 21.5 Three brands under a logit model

Offer one product from each of Laurentian's brands: Flex at CAD 1,249, Standard at 709 and Basic at 419, the Y, M and V levels of the ladder. Suppose an arriving customer chooses by multinomial logit with attraction weights 0.25 for Flex, 0.60 for Standard, 1.00 for Basic and 1.5 for not buying (illustrative). With all three open, she buys Flex with probability $0.25/3.35 = 0.0746$, Standard 0.1791, Basic 0.2985, and nothing 0.4478, so $Q = 0.5522$ and $R = 345.3$.

Figure 21.4 plots the seven offer sets. Three are efficient: Flex alone ($Q = 0.1429$, $R = 178.4$), Flex and Standard (0.3617, 313.9) and all three (0.5522, 345.3), with slopes 1,249, 619 and 165. Offering Standard and Basic without Flex is inefficient: with $Q = 0.5161$ it earns 272.4, while the mixture of the two neighbouring efficient sets with the same sales probability earns 339.3.

The slope of 165 is the marginal value of adding Basic. Basic sells 0.2985 per customer at 419, worth 125.1, but it takes customers from the other two brands: Flex falls from 0.1064 to 0.0746 and Standard from 0.2553 to 0.1791 per customer, a loss of 93.7. Net, Basic adds 31.4 of revenue for 0.1905 more sales: 165 per seat. The policy follows from Result 21.5: offer all three while the marginal value of a seat is below 165; Flex and Standard between 165 and 619; Flex alone between 619 and 1,249; nothing above 1,249. A system that compared Basic's published fare of 419 with a bid price of 300 would sell Basic on a flight where it destroys value.

The price of generality is estimation. Choice-based optimization needs choice probabilities for offer sets that were rarely or never offered, estimated from bookings made under the airline's own availability decisions. Availability and price are set by revenue management in response to demand, so the endogeneity problems of Chapter 7 apply in full.

21.7 Dynamic pricing

A fare ladder is a menu of a few prices. With continuous pricing, the airline can quote any price (Chapter 23). The theory of that case was laid down by Gallego and van Ryzin (1994), and it answers a question a planner should care about: how much of revenue management is the choice of price, and how much the timing.

A seller has x seats and t days until departure. Customers arrive as a Poisson process whose intensity depends on the price currently quoted, $\lambda(p)$, decreasing in p. The seller chooses the price at every moment to maximize expected revenue. Let $V(x,t)$ be the maximum expected revenue and $\Delta V(x,t) = V(x,t) - V(x-1,t)$ the opportunity cost of a seat. Over a short interval $\mathrm{d}t$, a sale occurs with probability $\lambda(p)\,\mathrm{d}t$, earns p and costs ΔV, which gives the Hamilton-Jacobi-Bellman equation

$$\frac{\partial V(x,t)}{\partial t} = \max_{p}\ \lambda(p)\,[p - \Delta V(x,t)], \qquad V(x,0) = V(0,t) = 0.$$

With exponential sell-up above the lowest fare,

$$\lambda(p) = \lambda_0\, e^{-\gamma(p/f_n - 1)}, \qquad p \ge f_n,$$

where λ_0 is the arrival rate of customers willing to pay f_n. This is the form $\lambda(p) = a\,e^{-\alpha p}$ with $\alpha = \gamma/f_n$ and $a = \lambda_0 e^{\gamma}$, and the model can be solved exactly.

RESULT 21.6 Optimal dynamic price with exponential willingness to pay

With $\lambda(p) = a\,e^{-\alpha p}$, let $\lambda^* = a/e$ be the sales rate at the price $1/\alpha$. Then

$$V(x,t) = \frac{1}{\alpha}\ln\sum_{i=0}^{x}\frac{(\lambda^* t)^i}{i!},$$

and the optimal price is the opportunity cost of a seat plus a constant markup,

$$p^*(x,t) = \Delta V(x,t) + \frac{1}{\alpha}. \tag{21.4}$$

For a given time to go, the optimal price is lower the more seats remain; for a given number of seats, it is higher the more time remains.

Proof. The bracket $ae^{-\alpha p}(p-\Delta)$ has derivative $ae^{-\alpha p}(1-\alpha(p-\Delta))$, which vanishes at $p = \Delta + 1/\alpha$, the maximum, with value $(\lambda^*/\alpha)\,e^{-\alpha\Delta}$. Write $S_x(t) = \sum_{i=0}^{x}(\lambda^* t)^i/i!$, so that $\mathrm{d}S_x/\mathrm{d}t = \lambda^* S_{x-1}$. For $V = \alpha^{-1}\ln S_x$, the left side of the equation is $\alpha^{-1}\lambda^* S_{x-1}/S_x$; on the right, $e^{-\alpha\Delta V} = S_{x-1}/S_x$, so the right side is the same. The boundary conditions hold because $S_x(0) = S_0(t) = 1$. The monotonicity statements are proved for general demand functions by Gallego and van Ryzin (1994); here they follow from $p^* = \alpha^{-1}(1 + \ln(S_x/S_{x-1}))$, which decreases in x and increases in t. ∎

Equation (21.4) is the Lerner rule of industrial organization with a dynamic marginal cost. The markup $1/\alpha$ is the static monopoly price, the price that would maximize revenue with unlimited seats, at which the elasticity of demand is -1; the opportunity cost ΔV is the bid price of Chapter 20, which here rises when seats are scarce relative to the time left. With $\gamma \le 1$, that is FRAT5 ≥ 1.69, the optimal price never falls below the lowest fare, so the floor at f_n never binds. Gallego and van Ryzin also show that the revenue of the deterministic version of the problem, with demand replaced by its mean, bounds the expected revenue of every policy, and that a single well-chosen fixed price is asymptotically optimal as the numbers of seats and customers grow (Gallego and van Ryzin 1994). Much of the

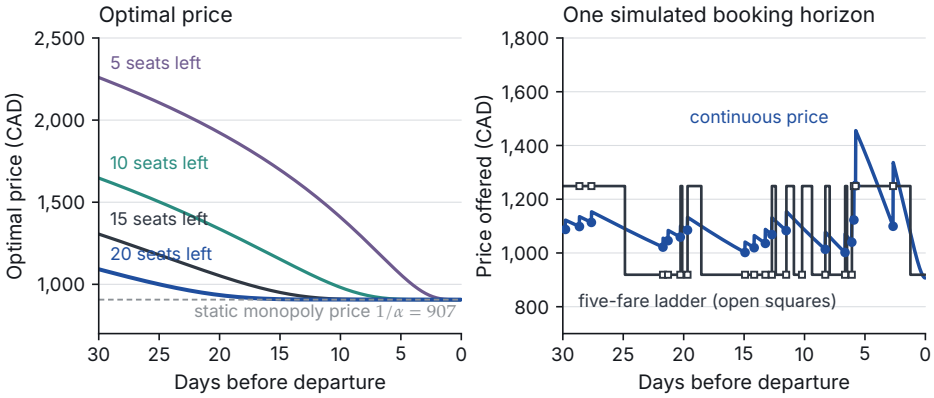


Figure 21.5 Dynamic pricing with exponential willingness to pay in the last 30 days before a YUL–SVQ departure (Example 21.6). Left: the optimal price (21.4) against days to departure for several numbers of seats left; all curves approach the static monopoly price of CAD 907 as departure nears. Right: one simulated booking horizon with the same customers facing the continuous optimal price and the optimal policy restricted to the five ladder fares; dots and squares mark sales.

value of dynamic pricing is the response to uncertainty: raising the price after a run of early sales, lowering it after a lull.

WORKED EXAMPLE 21.6 The last thirty days to Seville

Thirty days before departure Laurentian has 20 economy seats left on YUL–SVQ. Customers willing to pay at least the V fare of CAD 419 arrive at 1.2 a day, 36 in expectation, with the late-period sell-up of FRAT5 $= 2.5$, so $\gamma = 0.462$, $\alpha = \gamma/419 = 0.001103$ per dollar and $1/\alpha = 907$ (illustrative).

By Result 21.6 the expected revenue of optimal dynamic pricing is $\$V(20, 30) = \$$ CAD 18,376. The first quote is $p^*(20, 30) =$ 1,092: the static price of 907 plus an opportunity cost of 185. With 10 seats left and 30 days to go the quote would be 1,647; with 20 seats and 10 days, 907. Figure 21.5 shows these paths and one simulated horizon.

Three benchmarks put the number in context. The deterministic problem sells 20 seats at the price that makes expected demand equal capacity, $419\,[1 + \ln(36/20)/0.462] = 952$, for CAD 19,039: an upper bound. Charging 952 throughout earns CAD 17,348 in expectation, and the best fixed price, 1,130, earns CAD 18,056. The optimal policy restricted to the five ladder fares, computed by dynamic programming, earns CAD 18,203. Con-

tinuous dynamic pricing therefore adds 1.8 percent to the best fixed price, of which the five-fare ladder already captures 0.8 points; the remaining 0.95 percent is the value of being able to quote any price. That is of the same order as the rounding loss that Chapter 23 derives for coarse ladders.

A booking-class system is dynamic pricing with a discrete menu: opening and closing classes moves the price a customer sees, and the adjusted fares of Result 21.4 decide which rung to offer. Wittman and Belobaba (2019) propose a definitional framework with three mechanisms for choosing the price a customer sees: assortment optimization (which fare products to offer, as a booking-class system does), dynamic adjustment of filed fares, and continuous pricing, each of which can in the limit act on every request. The dynamic program does not care which: in every version the quote balances a markup set by willingness to pay against the opportunity cost of the seat.

IN PRACTICE **Willingness to pay in revenue management systems, 2026**

Forecasting willingness to pay is now a selling point of revenue management products rather than a research topic. PROS announced willingness-to-pay forecasting and optimization technology for airline revenue management in 2021 (PROS Holdings 2021) and states that adding willingness-to-pay data to forecasts raises revenue by about 1 to 3 percent, a vendor estimate without a published method (PhocusWire n.d.). A factsheet published by IATA and sourced from the Lufthansa Group defines dynamic pricing, continuous pricing and dynamic bundling, which together describe the move from filed fare ladders to prices computed per request (International Air Transport Association 2021). The best independent evidence on continuous pricing is simulation: in PODS it raised revenue by up to 2.0 percent for a first mover and 1.2 percent when both competitors adopted it (Szymański, Belobaba, and Papen 2021); Chapter 23 reviews this evidence and the vendor claims. For Air Canada, public sources confirm only that its pricing uses a PROS real-time dynamic pricing product (PROS 2025). Whatever the system, the questions of this chapter remain the ones to ask: how willingness to pay is estimated, from which booking

history, conditional on whose availability, and what the optimizer does with it.

21.8 Willingness to pay and competition

Everything so far concerned one airline. In a market with rivals, a customer whose fare is closed can buy a competitor's seat, and a rival's open fare caps what she will pay. The competitive side of revenue management appears elsewhere in this book: matching a rival's fare in Example 17.4, the PODS results on origin-destination control against a rival with basic controls in Chapter 20, and segmented continuous pricing in Chapter 23. Willingness to pay adds two points of its own.

The first is that sell-up is not a property of customers alone. A Laurentian customer who would pay CAD 709 for the nonstop will not pay it while a rival's connection sells at 549, unless the nonstop is worth the difference to her. If her premium for Laurentian is δ, she buys Laurentian's lowest open class j only if $W \geq f_j$ and $f_j - \delta \leq g$, where g is the rival's lowest open fare. With δ independent of W, the effective sell-up probability is

$$p_j^{\text{eff}} = p_j \cdot \Pr(\delta \geq f_j - g). \tag{21.5}$$

Take the late booking period of Example 21.1, with FRAT5 = 2.5. Laurentian's own sell-up probabilities for Y to V are 0.400, 0.576, 0.726, 0.866 and 1, and on its own it should open nothing below B. Now let a competitor's one-stop itinerary sell at $g = 549$, and let the premium for Laurentian's nonstop be normal with mean CAD 150 and standard deviation CAD 100 (illustrative). The probability that the premium covers the fare gap is 0.997 for V, 0.933 for Q, 0.460 for M, 0.014 for B and essentially zero for Y, so by (21.5)

$$p^{\text{eff}} = (0.000, 0.008, 0.334, 0.809, 0.997).$$

Revenue per customer is now highest with Q as the lowest open class, CAD 443.9 per customer against 7.4 for B: the rival's availability turns "open nothing below B" into "open Q".

The data consequence is as important. A history in which the rival was usually open at 549 shows only 33.5 percent of V-level customers selling up to M instead of 72.6 percent, a sell-up curve that would be read as a FRAT5 of about 1.44 rather than 2.5. This is why Belobaba's lecture points toward conditional estimates of willingness

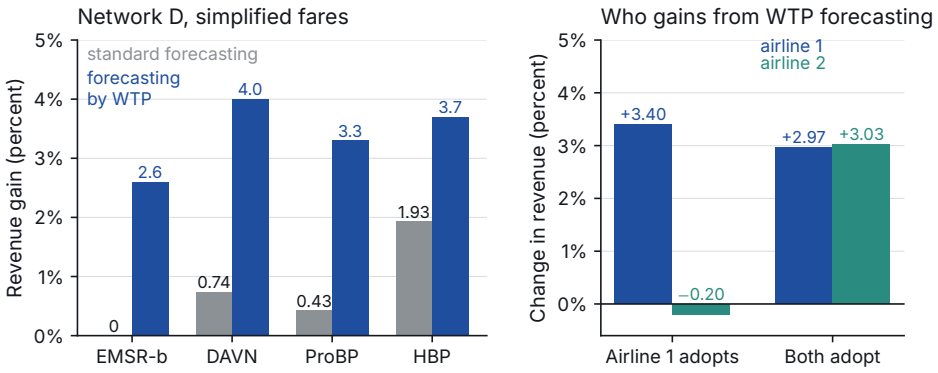


Figure 21.6 PODS simulations of forecasting by willingness to pay in Network D, two competing hub airlines with simplified fares (historical results, 2006; redrawn from Belobaba). Left: revenue gains of four revenue management methods over EMSR-b with standard forecasting, with standard forecasting and with forecasting by willingness to pay. Right: change in each airline's revenue when airline 1 alone adopts forecasting by willingness to pay, and when both do.

to pay, estimated separately for each scenario of class closures, the airline's own and possibly its competitors' (Belobaba 2006a).

The second point is about who gains. Figure 21.6 redraws the PODS results of the same lecture for Network D, two competing hub airlines with 40 spoke cities, 252 legs and 482 O&D markets under a simplified fare structure of six classes with a fare ratio of 4.1 (Belobaba 2006a). With standard forecasting, the network methods gained little over leg-based EMSR-b: 0.74 percent for DAVN, 0.43 for probabilistic bid prices and 1.93 for heuristic bid prices. Hybrid forecasting with a willingness-to-pay component raised the gains over standard EMSR-b to 2.6 percent for EMSR-b itself and 4.0, 3.3 and 3.7 percent for the three network methods. In a larger network with four competing hubs the totals were 1.3, 2.4, 3.0 and 2.4 percent. Forecasting, not optimization, was the binding constraint.

The competitive result is the instructive one. When airline 1 alone adopted hybrid forecasting, its revenue rose 3.4 percent, its yield rose and its load factor fell by 0.75 points, while airline 2 lost only 0.2 percent. When both adopted it, each gained about 3 percent (Belobaba 2006a). Forecasting by willingness to pay mostly recovers revenue that an airline was giving away to its own customers through buy-down, so it is close to a gain for everyone who adopts it. That contrasts with origin-destination control, where a non-adopting rival lost up to 0.7 percent (Chapter 20), and with segmented continuous pricing, whose large first-mover gains in PODS came mostly at rivals' expense and invited responses (Chapter 23). For a planner,

the lesson is to ask what a competitor's revenue management can do before forecasting its response to new capacity: an airline that forecasts by willingness to pay will not chase load factor with cheap seats.

21.9 Summary

- Classical revenue management assumes independent demand by class. When fences weaken, customers buy down, and the demand for a class depends on which classes are open.
- The exponential sell-up model describes willingness to pay above the lowest fare with one parameter, FRAT5, the fare ratio at which half the customers sell up. Laurentian's ladder implies about 1.52; late booking periods have much higher values.
- Q-forecasting scales each departure's bookings by the sell-up probability of the lowest open class to estimate demand at the lowest fare, then partitions it by willingness to pay. It cannot estimate responses to closures never observed.
- Forecasting high-fare demand from observed high-fare sales under buy-down produces spiral down: protection, high-fare sales and revenue fall season after season, and zero protection is an absorbing state.
- Fare adjustment replaces each fare by its marginal revenue, $(\mathrm{TR}_j - \mathrm{TR}_{j-1})/(Q_j - Q_{j-1})$, which turns the problem with buy-down into an independent-demand problem that standard optimizers solve exactly. Negative adjusted fares mark classes never to open.
- Choice-based revenue management generalizes the transformation to any choice model: only efficient offer sets are optimal, and the slopes of the efficient frontier are the adjusted fares.
- With exponential willingness to pay, the optimal dynamic price is the opportunity cost of a seat plus the static monopoly markup. On the Seville example continuous dynamic pricing adds 1.8 percent to the best fixed price; a five-fare ladder captures 0.8 points of it.
- A rival's open fare caps sell-up, so estimates of willingness to pay must be conditional on competitors' availability. In PODS, forecasting by willingness to pay raised revenue about 3 percent for each adopter and hurt a non-adopting rival very little.

Table 21.2 Key formulas of the chapter.

Quantity	Formula
Exponential sell-up	$p(f) = e^{-\gamma(f/f_n - 1)}$, $\gamma = \ln 2/(\text{FRAT5} - 1)$
Elasticity, static optimum	$\eta(f) = -\gamma f/f_n, f^* = \max(f_n, f_n/\gamma)$
Q-forecast	$\hat{\lambda} = \sum_k B_k / \sum_k p_{j(k)}$, $d_j = \hat{\lambda}\,(p_j - p_{j-1})$
Adjusted fare	$f_j' = (\text{TR}_j - \text{TR}_{j-1})/(Q_j - Q_{j-1})$
Continuous adjusted fare	$f - f_n/\gamma$
Efficient frontier slope	$\rho_k = (R(S_k) - R(S_{k-1}))/(Q(S_k) - Q(S_{k-1}))$
Optimal dynamic price	$p^* = \Delta V(x,t) + 1/\alpha$, $\alpha = \gamma/f_n$
Sell-up with a rival	$p_j^{\text{eff}} = p_j \Pr(\delta \geq f_j - g)$

21.10 Further reading

Belobaba's 2006 lecture on willingness to pay and competitive revenue management (Belobaba 2006a) and his ePODS pricing lecture (Belobaba 2006o) are short and contain the PODS results used here. Cooper, Homem-de-Mello and Kleywegt (2006) analyze the spiral-down effect formally. Fiig, Isler, Hopperstad and Belobaba (2010) develop fare adjustment and its use in mixed fare structures and network optimizers. Talluri and van Ryzin (2004a) is the foundational paper on choice-based revenue management; their book (2004b) and Gallego and Topaloglu (2019) treat choice models, assortment and dynamic pricing at textbook depth. Gallego and van Ryzin (1994) remains the clearest treatment of dynamic pricing with a fixed inventory, and Wittman and Belobaba (2019) relate it to airline practice.

CHAPTER

22

Where Planning Meets Revenue Management

IN THIS CHAPTER

- What network planning and revenue management each control, on which horizons, and the information that must pass between them.
- Bid prices as the value of a seat, recomputed for Laurentian's bank, and what their distribution across departures says that their average hides.
- Eight valuations of the same extra seats, and why the right one depends on the control that will sell them.
- RM-aware spill: why the classical spill model misprices capacity on flights that revenue management controls, and when one more seat loses money.
- The network contribution of a new long-haul leg, from O&D fares to contribution, with the displacement it causes on the feeders.
- Consistent forecasts, the reaction of revenue management to a schedule change, and why a first season misleads.
- The questions to ask the revenue management team, and how the two functions work together at a hub carrier.

DRAWS ON Belobaba, Introduction to the airline planning process, Airline demand analysis and spill modeling, and O&D control: what have we learned? (MIT 16.75J, 2006); handbook; Abramovich (2013); Talluri and van Ryzin (2004)

A network planner never sells a seat. Every number a planner uses to judge capacity, from the load factor and average fare of a flight to the spill it suffers and the revenue of a new route, is produced by a revenue management system that decides, request by request, which customers get the seats the plan creates. The plan draws the box; revenue management fills it. A planner who values capacity without asking how it will be filled is valuing a different airline.

This chapter joins the two halves of the book. Chapter 9 valued a seat with an average fare and the probability of filling it; Chapter 12

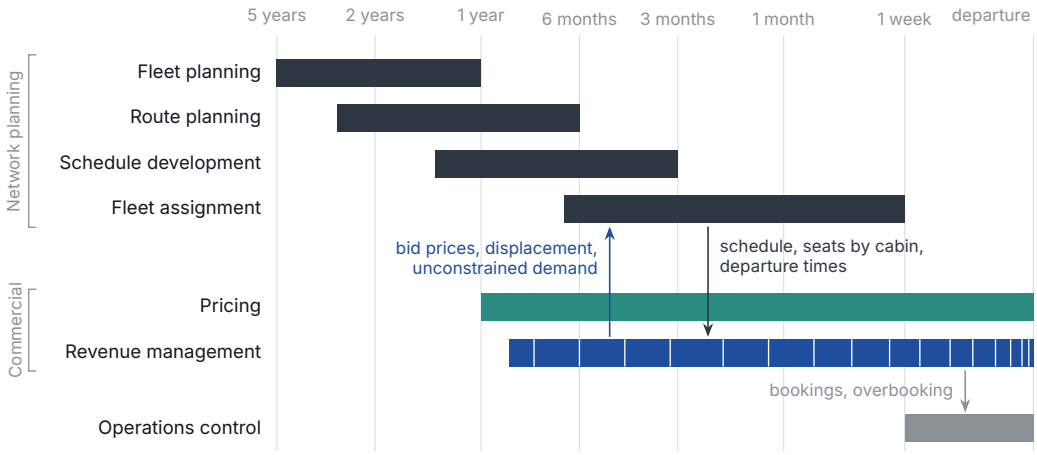


Figure 22.1 What network planning and the commercial functions control, and when. Bars are typical horizons before departure on a roughly logarithmic scale (Figure 1.1 gives the sources); white ticks mark revenue management's re-optimizations, which become more frequent as departure approaches. Arrows are the main flows of information between the functions.

measured network contribution with prorates and displacement; Chapter 14 priced spill into fleet assignment; Chapter 20 computed bid prices for Laurentian's four-leg bank at Montreal and showed that multiplying them by a seat count misleads. Here the same capacity changes are valued all these ways at once, so that the differences can be seen and explained. The running example is the evening bank of Chapter 20, with Quebec City and Boston feeding Paris and Seville, extended in Section 22.5 by a fifth leg to Lisbon.

One claim runs through the chapter: the value of capacity is not a property of a flight. It depends on the demand the flight faces, the network around it and the control that sells it, and changing any of the three can change the answer several times over.

22.1 Two functions, two horizons

Belobaba's introduction to airline planning describes revenue management as the last decision in the chain: given a scheduled flight, its capacity and its prices, how many bookings to accept by fare type, so as to fill each seat with the highest possible revenue (Belobaba 2006c). Everything before it, from the fleet plan to the fleet assignment, creates or reshapes capacity (Chapter 1). Figure 22.1 sets the two families of decisions side by side.

Planning controls what exists: the fleet, the routes, the frequencies and departure times, the connection banks, and through fleet

assignment the number of seats in each cabin of each departure. Its horizons run from years to weeks, and most of its decisions are fixed before the first booking arrives. Revenue management controls who gets what exists: which fare products and itineraries are open at each moment, the protection levels or bid prices that decide this, and the overbooking authorization (Chapters 18 to 20). It works on a booking period of about 300 days and revises its controls continuously (Belobaba 2006l). Pricing, the third partner, sets the fare products and their restrictions (Chapter 17).

Three asymmetries shape the interface. **Information**: the planner decides early, on market forecasts that revenue management will later replace with bookings in hand. In the Seville example of Example 8.4, a forecast that ignores the bookings on hand misses a departure's final demand by 30.5 passengers on average, against 8.7 for a pickup regression made 28 days before departure. **Objective**: revenue management maximizes revenue from capacity that is already there, so the cost of a seat is sunk for it, while the planner must weigh revenue against avoidable cost. **Granularity**: revenue management sees each departure, fare product and day before departure; planning works with seasons, markets and averages. Each asymmetry is a source of numbers that disagree for good reasons.

The information that crosses the interface follows from these roles. Planning hands revenue management a schedule, seats by cabin and departure times, and with them the itineraries that can connect. Revenue management hands back what it knows that planning does not: demand by origin-destination and fare product with the censoring removed, the fare mix it can achieve, and the value of a seat on each leg. The handbook describes the division in the same terms: planning sets much of the capacity and schedule, revenue management controls how the product is sold over time, and a bid price estimates the opportunity value of scarce capacity (Network Planning Handbook 2026).

In 2006 Belobaba described airline practice as sequential: scheduling, pricing and revenue management were performed one after another. He argued that integrated models would let feedback from pricing and revenue management improve the choice of schedule and aircraft (Belobaba 2006c). He also listed why joint optimization is hard: few airlines had consistent databases of demand and cost, models of dynamics and competitive behaviour were immature, and coordination inside the airline was difficult (historical) (Belobaba 2006c). Much of this chapter is about doing that feedback by hand, carefully, while integrated tools mature. Its starting point is a definition that makes the dependence on the control explicit.

DEFINITION 22.1 Value of capacity

Fix a demand model and a booking control u, which may be first come, first served, a revenue management method, or the ex post optimum that knows each departure's demand in advance. Let $R_u(C)$ be the expected network revenue per departure when the legs have capacities $C = (C_l)_{l \in \mathcal{L}}$. The **revenue value** of a change of capacity from C to C' under control u is $R_u(C') - R_u(C)$. Its **contribution value** subtracts the change in avoidable cost, including the passenger-related costs of every leg whose load changes.

A seat has no value of its own. Its value is a change in what the network earns, and that change depends on the demand, the rest of the network and the control that sells the seat.

22.2 Bid prices: the value of a seat, departure by departure

Revenue management already computes a value of capacity: the bid price. Solving the deterministic linear program of Section 20.5 for Laurentian's bank with `scipy.optimize.linprog` reproduces the numbers of Example 20.3: revenue of CAD 365,949 per bank and bid prices of CAD 159 on YQB–YUL, 109 on BOS–YUL, 490 on YUL–CDG and 0 on YUL–SVQ. The sum of the bid prices along a product's path is its **displacement cost**, the revenue the network gives up elsewhere to carry it (Definition 20.3); the dual rent β_j is what the product earns above it.

For a planner these four numbers already answer the question of the sixth feedback loop of Section 1.2: where would capacity be worth more? On the Paris leg, by a factor of three or more. They also screen proposals. A new connection over Montreal to Paris must earn at least the Paris bid price plus the feeder's to be worth carrying at mean demand, and in the deterministic program a feeder seat adds at most CAD 159 or 109.

Two limits matter. A bid price is a derivative, valid for a few seats (Result 20.2). And the deterministic program replaces random demand by its mean, so it describes an average departure that never

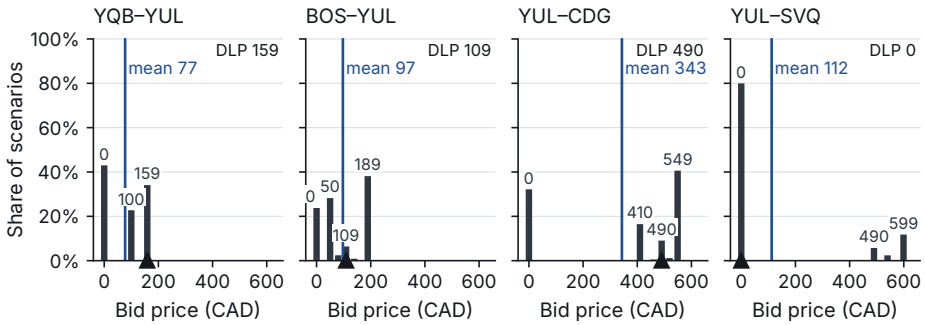


Figure 22.2 Ex post bid prices on Laurentian's bank: for each of the 4,000 demand scenarios of Section 20.7, the dual price of each leg's capacity in the LP solved with realized demand. Bars show how often each value occurs; the vertical line is the mean, the triangle the deterministic LP bid price. Every value is a fare difference.

occurs. To see the value of a seat on actual departures, solve the program again for each of many demand scenarios, with the realized demand in place of the mean. The dual of each scenario's program is the **ex post bid price**: the value of one more seat on that departure to a controller who knew its demand in advance. Section 20.7 averaged these prices over 4,000 scenarios and obtained CAD 77, 97, 343 and 112. Figure 22.2 shows their distributions.

The distributions are lumpy, and each lump has an explanation: a bid price equals the fare of the departure's marginal product minus the bid prices of that product's other legs. On YUL–CDG the bid price is zero on 32 percent of departures (the Paris flight has room), CAD 549 on 41 percent (the marginal passenger is a local restricted traveller), CAD 410 on 17 percent (599 − 189: the marginal passenger is a restricted Boston–Paris connection whose Boston seat could go to a local) and CAD 490 on 9 percent (649 − 159, the deterministic case). YUL–SVQ is worth nothing on 80 percent of departures and CAD 599, the local restricted fare, on 12 percent. YQB–YUL takes essentially three values: zero on 43 percent, the local restricted fare of CAD 159 on 34 percent, and CAD 100 (649 − 549) on 23 percent, when a restricted Quebec City–Paris connection is marginal against a local Paris passenger.

The average of these prices has a precise meaning, which follows from Result 20.2 applied departure by departure.

RESULT 22.1 Mean ex post bid prices bound the value of capacity

Let $V(C;d)$ be the optimal value of the deterministic LP (20.3) with the realized demand d in place of μ, let D be random demand, and let $\pi(D)$ be an optimal dual of each scenario's program. Then the expected ex post value $\bar{V}(C) = \mathbb{E}[V(C;D)]$ is concave in C, and with $\bar{\pi}_l = \mathbb{E}[\pi_l(D)]$, for every capacity vector C',

$$\bar{V}(C') \;\le\; \bar{V}(C) + \sum_{l \in \mathcal{L}} \bar{\pi}_l \,(C'_l - C_l). \tag{22.1}$$

Proof. For each realization d, Result 20.2 applied with d in place of μ gives $V(C';d) \le V(C;d) + \sum_l \pi_l(d)(C'_l - C_l)$, and $V(\cdot\,;d)$ is concave. Taking expectations of both sides gives (22.1), and an average of concave functions is concave. ∎

The mean ex post bid price is therefore a supergradient of the value of capacity to a clairvoyant controller, and like any derivative it overstates the value of large additions. For the 32 extra seats of the Quebec City upgauge it gives about CAD 2,470, against the expected ex post value of CAD 1,257 found in Example 20.5; for 30 extra Paris seats, CAD 10,299 against 8,085.

For planning, the distribution says more than the mean. A leg whose bid price is zero on most departures and high on a few, like Seville, needs seats on its peak departures, which argues for a seasonal or day-of-week gauge pattern rather than a permanent upgauge. A leg whose bid price is high on most departures, like Paris, is short of seats almost always. Ask revenue management for bid prices by departure and by days before departure, not for a seasonal average.

22.3 Eight valuations of the same seats

Example 20.5 valued two capacity changes three ways: the bid price times the seats, the deterministic LP re-solved, and the ex post optimum. Planning and revenue management together have more tools than that, and they disagree. This section values the same changes eight ways.

The first three are leg-based arithmetic. **Bid price times seats** multiplies the deterministic bid price by the number of seats added. The **spill model** of Chapter 9 treats the leg's demand as normal, with the

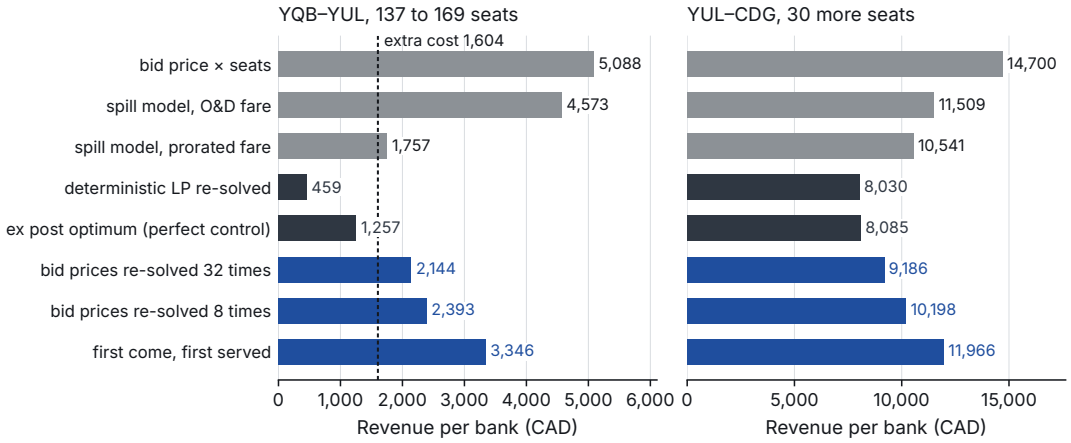


Figure 22.3 The revenue value of the same extra seats under eight methods, for the YQB–YUL upgauge of Example 20.5 (left) and thirty more seats on YUL–CDG (right). Grey: leg-based arithmetic. Dark: the network LP at mean demand and with realized demand. Blue: booking simulations of 2,000 departures under three controls. The dashed line is the extra cost of the larger aircraft, including passenger costs.

sum of the means and variances of the products that use it, and multiplies the extra expected load by an average fare of the leg's passengers: either their full O&D fares or their fares prorated to the leg by straight-rate prorate (Definition 12.2). The next two are network optimizations: the **deterministic LP re-solved** with the new capacity, and the **ex post optimum**, the expected value of the LP solved with each departure's realized demand, which bounds every control (the proof of Result 20.3 applies departure by departure).

The last three are booking simulations with the simulator of Example 20.4, run on 2,000 new departures. Requests arrive one at a time, restricted ones over the first 80 percent of the booking period and flexible ones over the last 60 percent, and every control sees the same requests in the same order. The controls are bid prices from the DLP re-solved 32 times and 8 times from the remaining capacity and the expected remaining demand, and first come, first served. They run from the best control to none: on the base network they earn CAD 352,636, 349,092 and 339,334 per bank, against 359,274 for the ex post optimum on the same departures.

WORKED EXAMPLE 22.1 Thirty more seats to Paris

Suppose Laurentian could offer 328 seats instead of 298 on YUL–CDG. The leg's demand is the sum of six products, with mean 317 and standard deviation 36.1.

Leg arithmetic. The bid price gives 490 × 30 = 14,700 dollars per bank. The spill model says that the flight fills on 70 percent of departures and that expected load rises from 291.1 to 307.4, by 16.3 passengers. At the average O&D fare of the leg's passengers, CAD 706.7, they are worth CAD 11,509; at the average straight-rate prorated fare, CAD 647.3, they are worth 10,541.

The deterministic LP. Re-solving gives CAD 8,030, as in Table 20.7, and the solution shows why. The extra seats admit the 16 restricted Boston–Paris and 3 restricted Quebec City–Paris connections that the program refused (CAD 9,584 and 1,947). But a connection needs a feeder seat, and both feeders are full: 16 Boston locals and 3 Quebec City locals in the restricted fares lose their seats (CAD 3,024 and 477). Each extra Paris passenger earns its fare minus the local fare it displaces, 599 − 189 = 410 or 649 − 159 = 490. The remaining 11 seats stay empty at mean demand, the Paris bid price falls to zero, and the Boston bid price rises from 109 to 189. The ex post optimum adds the departures with high Paris demand and gives 8,085.

The simulated controls. With bid prices re-solved 32 times the extra seats are worth CAD 9,186 per bank; re-solved 8 times, 10,198; under first come, first served, 11,966 (standard errors 160 to 250). Under eight re-solves the extra seats carry 16.4 more Paris passengers on average, and the revenue comes from local Paris passengers (CAD 6,172) and Paris connections (5,496), less the feeder locals that those connections displace (1,370). Under first come, first served the extra seats mostly rescue late-booking flexible Paris passengers who used to find the flight full, 6.4 more per bank: locals bring CAD 9,886 and connections 2,486.

Reading the eight numbers. They range from 8,030 to 14,700, and the three tools a planner would reach for first, the bid price and the two spill models, give three of the four largest values. The spill model with prorated fares lands near the eight-solve simulation, but by coincidence: it ignores both the displacement on the feeders and the control, and the errors offset. The

four control-based numbers are ordered by the quality of the control: the better the control, the less the extra seats are worth.

The last observation is not an accident of this example. A control that loses revenue on the existing seats, by selling them to the wrong passengers or leaving them empty, gains more from extra seats, because the extra seats repair part of its mistakes. Perfect control has no mistakes to repair. In these simulations capacity and control are substitutes, so the value of capacity must be measured under the control that will sell it.

WORKED EXAMPLE 22.2 The Quebec City upgauge under three controls

Example 20.5 valued replacing the A220-300 (137 seats) by a 737-8 (169 seats) on YQB–YUL and found that even the ex post optimum, CAD 1,257 per bank, falls short of the cost. The extra cost is CAD 1,400 per trip with Laurentian's cost model, plus passenger costs of about CAD 204 per bank on all the legs whose loads change: CAD 1,604 in total.

The spill model is the planner's usual first estimate. The leg's demand has mean 141 and standard deviation 19.7, and expected load rises from 131.0 to 140.3 passengers, by 9.33. Priced at the straight-rate prorated average fare of CAD 188.3, they are worth CAD 1,757; at the mileage-prorated average of 119.6, 1,116; at the full O&D fares of the leg's passengers, 4,573, because nearly half of them connect to Europe.

The simulated controls give CAD 2,144 (32 re-solves), 2,393 (8 re-solves) and 3,346 (first come, first served). Under eight re-solves the extra seats sell, per bank, 1.8 more flexible and 3.2 more restricted local Quebec City tickets, 2.9 more Quebec City–Paris and 0.5 more Quebec City–Seville connections, and 2.1 fewer local Paris tickets, displaced by those connections. Split by straight-rate prorate, CAD 1,729 of the gain (72 percent) appears on YQB–YUL and the rest on the other legs: 354 on Paris, 298 on Seville and 12 on Boston.

The decision now depends on the control. Against CAD 1,604, the upgauge loses CAD 347 per bank under perfect control, as Chapter 20 found, but gains CAD 540 under bid prices resolved 32 times, 789 under 8 re-solves and 1,742 under first

come, first served. The business case for a gauge change contains a hidden assumption about revenue management. If the revenue management team is about to improve its control, the case weakens; if the extra seats will be sold by a control as rough as the simulation's, it holds.

PITFALL

Valuing capacity under the wrong control. The ex post optimum assumes a controller who knows demand in advance and understates what extra seats are worth to a real one; first come, first served assumes no control and overstates it. Neither is the airline. Ask revenue management to value a capacity change by simulating the control it actually runs, with the forecasts it will actually have.

22.4 RM-aware spill

The classical spill model of Chapter 9 values a seat at the average fare times the probability of filling it. It assumes that seats are sold first come, first served, so that the spilled customers are a random sample of all customers. Revenue management breaks that assumption in three ways. It spills from the bottom of the fare ladder, so a spilled passenger is worth less than the average fare (Section 9.8). It turns away low-fare requests even on departures that leave with empty seats, so some demand that a spill model counts as carried is in fact refused. And customers respond to what is open: when more seats keep cheap fares open more often, some customers who would have paid more buy down (Definition 21.2).

Abramovich studied these effects with MIT's Passenger Origin-Destination Simulator in a master's thesis supervised by Belobaba (Abramovich 2013). He argues that earlier spill models ignored passenger choice and revenue management, and that revenue management changes both the number and the value of spilled passengers, because it limits low-fare availability even while seats remain. In less restricted fare structures under traditional revenue management, adding capacity could reduce revenue because passengers bought down; capacity added to one flight could also lose revenue on another through recapture; forecasting and optimization that account for willingness to pay reduced these losses. A heuristic spill

cost that accounts for sell-up roughly halved the error of earlier models in restricted fare structures with moderate spill, to about 5 percent (Abramovich 2013).

The two-fare flight of Section 9.8 shows the mechanism in a form that can be solved exactly. The flexible fare is $f_H = 1{,}200$ with demand $D_H \sim \mathcal{N}(40, 12^2)$, the restricted fare $f_L = 600$ with demand $D_L \sim \mathcal{N}(150, 40^2)$, and restricted customers book first. Littlewood's rule (Section 18.4) protects $y = 40$ seats for the flexible fare, since $f_L/f_H = 0.5$ and the median of D_H is 40, so the restricted fare's booking limit is $b = C - y$. With **fenced** fares each customer buys the product meant for them. **Without fences** every flexible customer buys the restricted fare while it is open, and pays f_H only after it closes.

RESULT 22.2 The marginal seat with and without fences

Treat capacity as continuous and hold the protection level y fixed, so that $b = C - y$. With fences, the derivative of expected revenue R with respect to capacity is

$$\frac{\mathrm{d}R}{\mathrm{d}C} = f_L \Pr(D_L > b) + f_H \Pr(D_L \le b,\ D_L + D_H > C).$$

Without fences it is

$$\begin{aligned}\frac{\mathrm{d}R}{\mathrm{d}C} &= f_L \Pr(D_L > b \text{ or } D_L + D_H > C) \\ &\quad - (f_H - f_L) \Pr(D_L \le b < D_L + D_H \le C).\end{aligned} \tag{22.2}$$

The second term can dominate, and one more seat then reduces expected revenue.

Proof. Fix the demands and follow one extra seat, which raises C and b by the same amount. With fences, if $D_L > b$ the restricted fare is at its limit and the seat goes to one more restricted customer, worth f_L; the flexible customers still have y seats. If $D_L \le b$, all restricted customers are carried, the flexible ones have $C - D_L$ seats, and the extra seat carries one more of them, worth f_H, exactly when $D_H > C - D_L$.

Without fences there are four cases. If $D_L > b$, the flexible customers find the restricted fare closed and have y seats as before, and the seat earns f_L. If $D_L + D_H \le b$, everyone buys the restricted fare and the seat is unused. In between, $D_L \le b < D_L + D_H$, the first

$b - D_L$ flexible customers buy down and the rest pay f_H for at most y seats, and raising b lets one more flexible customer buy down. If $D_L + D_H > C$, some flexible customers were spilled, the protected seats stay full, and the extra passenger adds f_L. If $D_L + D_H \le C$, nobody was spilled, so the converted customer simply pays f_L instead of f_H, a change of $-(f_H - f_L)$. Revenue is Lipschitz in C on every realization, so the derivative of its expectation is the expectation of these derivatives, which gives both formulas. ∎

The loss is not inevitable. It comes from holding the protection level fixed while capacity grows, which is what a control that ignores buy-down does.

RESULT 22.3 Re-optimized protection never makes capacity harmful

If the protection level is re-optimized for each capacity, with buy-down anticipated, expected revenue is non-decreasing in capacity, with or without fences.

Proof. Compare capacity C with protection y and capacity $C+1$ with protection $y + 1$. The booking limit b is the same, so the restricted sales and the buy-down are the same on every realization, and the flexible customers who remain have one more seat. Revenue cannot fall. Hence the best revenue at $C + 1$ is at least the revenue with protection $y + 1$, which is at least the best revenue at C. ∎

The airline can always keep an extra seat for the flexible fare. A control that instead hands it to the restricted fare, because its protection was computed as if customers never bought down, is what turns capacity into a loss.

Figure 22.4 plots the marginal seat under the four assumptions. At 150 seats the spill model says CAD 604, the fenced flight earns about CAD 530 (the value of Section 9.8), and the unfenced flight 448 with fixed protection or 454 with protection chosen for buy-down. At 180 seats the four values are 432, 406, 232 and 257. The fenced curve and the spill model meet near mean demand, as Chapter 9 found; the unfenced curves fall far below both.

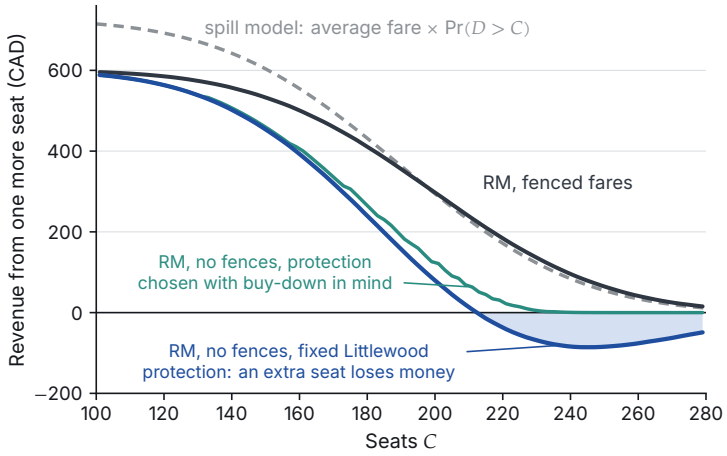


Figure 22.4 The revenue from one more seat on the two-fare flight of Section 9.8. Dashed: the classical spill model. Dark: revenue management with fenced fares. Without fences every flexible customer buys the restricted fare while it is open: with Littlewood's protection of 40 seats held fixed (cobalt), extra seats lose money from about 213 seats; with protection chosen with buy-down in mind (teal) they are worth little but never negative. Simulated with the 400,000 demand draws of Figure 9.6.

WORKED EXAMPLE 22.3 **When one more seat loses money**

Take the unfenced flight with 220 seats. With fixed protection $y = 40$ the restricted fare closes after $b = 180$ sales. In the 400,000 simulated departures:

- the restricted fare reaches its limit ($D_L > 180$) on 22.6 percent of departures;
- restricted demand stays below the limit but flexible customers are spilled ($D_L \le 180$ and $D_L + D_H > 220$) on another 4.0 percent;
- the restricted fare closes after some flexible customers have bought it, and nobody is spilled ($D_L \le 180 < D_L + D_H \le 220$), on 32.8 percent.

By (22.2), one more seat at 220 earns $600 \times (0.226 + 0.040) - 600 \times 0.328 = -37$ dollars, and a direct simulation of $R(221) - R(220)$ gives -39. The spill model values the same seat at CAD 172, and the fenced flight earns CAD 182 from it.

With protection re-optimized for buy-down the seat is worth CAD 24, because the control raises protection as capacity grows, from 60 seats to 61. This does not contradict Result 22.3: the loss of CAD 37 comes from holding protection at 40 seats while capacity grows, not from the extra seat itself. With fixed protection, every extra seat from about 213 seats upward loses money, up to CAD 86 near 245 seats.

The planner learns two things. On flights with weak fences, the classical spill value can overstate the value of capacity several times over, and in the wrong range it gets even the sign wrong. And the remedy for revenue lost to buy-down belongs to revenue management and pricing, through fare adjustment (Section 21.5) or fences (Chapter 17), not to the fleet plan.

PITFALL

Concluding that capacity destroys revenue. Negative marginal revenue is a property of a fare structure with weak fences combined with a control that ignores buy-down. The handbook presents Abramovich's result as a model-specific warning against mechanical capacity arithmetic, not as a claim that more capacity normally destroys revenue (Network Planning Handbook 2026). Ask how the airline's own control handles buy-down before reading anything into it.

In network terms, RM-aware spill means three adjustments to the formulas of Chapter 9. Value spilled passengers at the fares revenue management would actually refuse, which are close to the lowest open fares, not at the average fare. Count separately the demand that revenue management rejects on departures with empty seats: it is lost by choice, and capacity does not recover it. And include the customers' response, buy-down and sell-up, whose size depends on the fences and on whether forecasts and optimizer account for willingness to pay (Chapter 21). The simulations of Section 22.3 make all three adjustments at the network level, which is why they are the numbers to ask for.

22.5 The network contribution of a new leg

Valuing a new leg needs every piece of the book at once: O&D demand and fares, spill on the new flight, displacement on the legs it shares, the response of revenue management, and avoidable cost. Specialize the clean comparison (12.1) of Chapter 12 to ticket revenue and avoidable trip costs. If capacities change from C to C' under control u, the incremental network contribution per bank is

$$\Delta \mathrm{NC} = [R_u(C') - R_u(C)] - \sum_l \left[\Delta \mathrm{TC}_l + c_l^{\mathrm{pax}}\, \Delta \mathbb{E}[Y_l]\right], \qquad (22.3)$$

where $\Delta \mathrm{TC}_l$ is the change in the flight-related trip cost of leg l (the whole cost of a new flight, the difference between types for a gauge change), c_l^{pax} the passenger-related cost per passenger on leg l, and $\Delta \mathbb{E}[Y_l]$ the change in its expected load. Ownership and the alternative uses of the aircraft are compared separately, as in Section 12.5.

Chapter 11 lists Lisbon among the thin transatlantic routes for Laurentian's A321XLRs. Add a daily YUL–LIS flight with 182 seats to the evening bank, connecting from Quebec City and Boston. Following the assumptions of Example 11.5, the Lisbon market is taken to be 10 percent larger than Seville's, with fares in the ratio of 600 to 620; Table 22.1 lists the six new products.

Table 22.1 Illustrative Lisbon products added to Laurentian's bank (fares in CAD, demand per bank), scaled from the case's Seville products and rounded.

Product	Fare	Mean demand	Standard deviation
YUL–LIS H	1,199	24	9
YUL–LIS L	579	99	24
YQB–LIS H	1,349	4	2
YQB–LIS L	679	18	6
BOS–LIS H	1,399	6	2
BOS–LIS L	659	26	8

WORKED EXAMPLE 22.4 Adding Lisbon to the evening bank

The leg view. The Lisbon leg's demand has mean 177 and standard deviation 27.7 against 182 seats, so the flight fills on 43 percent of departures and carries 168.3 passengers on average, spilling 8.7. The block time is 6.63 hours eastbound by the rule of Example 11.5, and the trip cost with Laurentian's cost model is CAD 76,430 plus CAD 50 per passenger. At the straight-

rate prorated average fare of CAD 673.4 the flight earns CAD 113,328 and contributes CAD 28,484 after its trip cost of 84,844. With mileage prorate the average fare is CAD 715.5 and the contribution CAD 35,565. These are flight contributions in the sense of Definition 12.1, and they differ by the choice of prorate alone.

The network at mean demand. Re-solving the deterministic LP with five legs gives CAD 485,886 per bank, a gain of 119,937. The Lisbon passengers pay 129,243 in O&D fares, and all of them are accepted: every Lisbon product clears its old path bid price by a wide margin (the narrowest margin, for YQB–LIS L, is $679 - 159 = 520$). But the 54 connecting Lisbon passengers need seats on two feeders that are already full, and the program makes room by refusing 25 restricted Quebec City locals (CAD 3,975) and 29 restricted Boston locals (CAD 5,481); a small reshuffle of Paris connections adds CAD 150. Displacement costs CAD 9,306. The Boston bid price rises from 109 to 189, because the marginal passenger on that leg is now a local, and the Paris bid price falls to 410.

Random demand and revenue management. With realized demand, the ex post optimum gains CAD 117,607, 2,330 less than the deterministic program. Under bid prices re-solved 8 times the gain is 112,813: imperfect control costs another 4,794. The Lisbon flight then carries 166.6 passengers on average, whose fares bring CAD 120,377, and displacement is CAD 7,563: per bank, 10.8 fewer restricted and 1.8 fewer flexible Quebec City locals, 16.2 and 3.0 fewer Boston locals, and 5.5 fewer restricted Boston–Paris connections, partly offset by 5.9 more local Paris passengers in the seats those connections free. Under first come, first served the gain is only CAD 100,082: without control, the Lisbon connections, which book early in their restricted fares, also take feeder seats from Paris and Seville connections, and the loads of those two flights fall by about five passengers each.

Contribution. The Lisbon trip cost at the simulated load is CAD 84,761, and passenger costs on the other legs rise by CAD 89. By (22.3) the network contribution is $112{,}813 - 84{,}761 - 89 = 27{,}963$ dollars per bank, before ownership. Figure 22.5 shows the bridge from O&D fares to contribution.

Three comparisons show what each partial view misses. The leg view with straight-rate prorate, CAD 28,484, lands within CAD 521

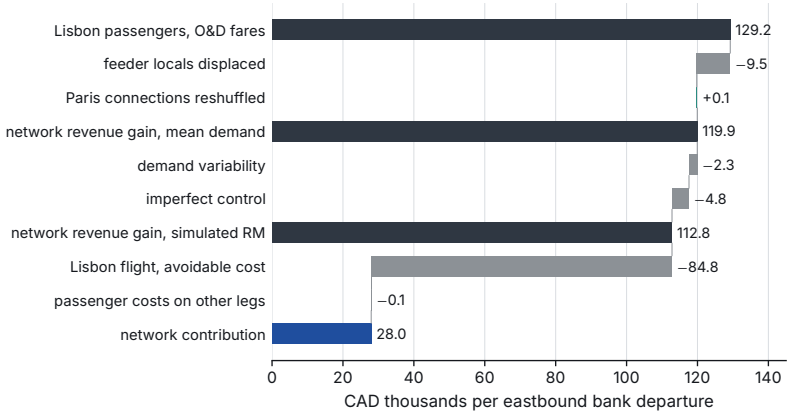


Figure 22.5 From the O&D fares of the Lisbon passengers to the network contribution of the new flight, per bank. The first block is the deterministic LP at mean demand; the second replaces it by the ex post optimum and then by simulated bid-price control; the last subtracts the avoidable cost of the Lisbon flight and the passenger costs on the other legs.

of the network answer, but only because two errors cancel: it charges the flight nothing for the displaced feeder locals, and it credits the feeders with a large share of each connecting fare, CAD 146 of the YQB–LIS L fare. With mileage prorate the leg view overstates the network contribution by CAD 7,602. The deterministic network view, $119{,}937 - 85{,}280 = 34{,}657$ at the mean load, gets the displacement right but ignores variability and the control, which together cost 7,124. And the spill cost that the leg-based fleet assignment coefficient (14.2) would charge the Lisbon flight, $8.71 \times 673.4 = 5{,}868$ dollars, overstates the network cost of that spill: giving the flight unlimited seats raises the ex post value by only CAD 4,123, because when Lisbon spills a connecting passenger the feeder seat goes to a local who would otherwise have been displaced. The leg coefficient is 42 percent too high, which is the network effect that the itinerary-based fleet assignment model of Section 14.7 is built to capture.

Two lessons follow for the planner. The displacement caused by a new long-haul leg falls on the feeders, in passengers that no report of the new flight will show, and the feeders' own reports hide it too: at mean demand the Quebec City and Boston flights carry as many passengers as before and, with straight-rate prorate, even report slightly more revenue (CAD 252 and 318 per bank), because the Lisbon connections that take the restricted locals' seats are credited with prorated shares close to or above the local fares. Only the network view shows the CAD 9,456 of local fares given up. And in

this simulation the new leg is worth more under better control, the opposite of the upgauges of Section 22.3: added demand competes for scarce feeder seats, and allocating scarce seats is what revenue management does. Extra capacity substitutes for control; extra demand calls for more of it.

22.6 Consistent forecasts

Planning and revenue management forecast the same travellers for different decisions, at different levels and on different horizons. Many arguments between the two functions about "the demand" are differences of definition, and the first step is to name the quantities.

> **DEFINITION 22.2 Constrained and unconstrained demand**
>
> The **unconstrained demand** for a product is the number of requests it would receive if it were open throughout the booking period at its fare. **Constrained demand**, or bookings, is what was sold given capacity and the availability that revenue management allowed. The difference is lost in two ways: **spill**, requests refused because seats ran out, and **RM rejections**, requests refused while seats remained because the product was closed to protect seats for others.

With sell-up and buy-down, the unconstrained demand for a product depends on which other products are open, so it is defined relative to an availability policy; price-oriented forecasts avoid the ambiguity by forecasting demand by willingness to pay instead (Section 21.3). Planning needs demand that capacity has not censored. Revenue management's history is censored by both mechanisms, so its raw bookings understate the demand a larger aircraft would face (Section 9.5). Table 22.2 lists the reconciliations a planner should check.

Table 22.2 Demand quantities that planning and revenue management share, and how to keep them consistent.

Quantity	Planning uses it for	Revenue management uses it for	Consistency check
Unconstrained O&D demand	market size, share, spill	product forecasts	same definition of unconstrained, same vintage
Leg demand	gauge, spill, fleet assignment	booking limits, overbooking	build it as $\sum_j a_{lj} D_j$, never by adding leg totals
Fare mix and average fare	revenue per passenger	protection, bid prices	the next passenger's fare is not the average fare
Bookings and loads	post-launch review	history for forecasts	unconstrain before comparing with demand
Recapture, sell-up, buy-down	spill cost, network value	choice-based forecasts	the same customer-choice assumptions

Three rows deserve numbers. **O&D and leg.** At mean demand Laurentian's bank faces 621 O&D travellers, who would make 772 leg boardings, because 151 of them connect and appear on two legs. The handbook's warning is exactly this: O&D travellers and leg boardings are different measures and cannot simply be summed into one hierarchy (Network Planning Handbook 2026); Section 8.5 shows how to reconcile forecasts that must add up. **Horizon.** A gauge decision taken six months ahead faces both day-to-day variation and the error of the forecast itself, while revenue management's forecasts close to departure carry little of the second. If the YQB–YUL forecast is uncertain by 15 percent of its mean, the standard deviation of the demand the decision faces rises from 19.7 to 28.9 passengers, and the 32 extra seats of the upgauge carry 11.1 passengers on average instead of 9.3. **Fares.** The handbook asks planners not to assume that the observed average fare is the fare of the next passenger (Network Planning Handbook 2026); Section 22.4 showed by how much the two can differ.

> **PITFALL**
>
> Summing leg forecasts. Adding the forecasts of the legs of a bank counts every connecting traveller twice: Laurentian's bank has 621 O&D travellers and 772 leg boardings at mean

> demand. Build leg demand from O&D demand, and keep the two in separate hierarchies.

22.7 How revenue management reacts to a schedule change

A schedule change reaches the revenue management system as new capacities, and the system responds at once. Bid prices move: thirty more Paris seats send the Paris bid price from 490 to zero and raise Boston's from 109 to 189 at mean demand (Example 22.1). Availability follows: restricted fares open more often on the enlarged leg and on the connections that use it, and may close on other legs. The effect of the change therefore appears across the network and in the fare mix, which is why the gain from the Quebec City upgauge is split across four legs (Example 22.2).

Changes to flights already on sale add a constraint. A smaller aircraft on a departure that is already sold turns every booking beyond the new authorization level into a potential denied boarding, a cost that belongs in the case for the change (Chapter 19). A larger one gives revenue management seats to sell late, mostly to late-booking customers who pay more.

Revenue management's forecasts also lag the change. They are built from bookings under the old capacity, which were censored by the old booking limits, so the system starts the first season of a larger aircraft with forecasts that describe the smaller one. If unconstraining is weak, lower forecasts reduce protection, reduced protection sells more cheap seats, and the history drifts downward: the spiral of Section 21.4. A new route has no history at all and is sold on forecasts borrowed from similar markets (Section 8.7) while its own demand ramps up (Example 12.5).

The first season after a schedule change is therefore a poor measure of its value. Revenue management is still learning, demand is still ramping up, the market and competitors moved at the same time (Section 21.8), and the leg's own report misses the network effects. A last reason is statistical. In the simulation, the bank's revenue varies from departure to departure with a standard deviation of CAD 18,597 under bid-price control, so a season of 180 departures estimates its mean with a standard error of about CAD 1,390, and the difference between two seasons with a standard error of about CAD 1,960, comparable to the CAD 2,393 that the Quebec City upgauge

is worth. Correlation between departures of the same season makes this worse, not better.

WORKED EXAMPLE 22.5 Reading the first season

Suppose the Quebec City upgauge has flown for a season. The leg's revenue under straight-rate prorate rose from CAD 25,000 to 28,230 per bank (12.9 percent), and a comparable feeder outside the bank, unchanged, rose from 21,000 to 22,260 (6.0 percent). The numbers are illustrative: they are built from a market that grew 6 percent and the leg effect of the simulation, CAD 1,729, rounded to 1,730.

Before and after. The leg gained CAD 3,230, twice the cost of the upgauge. Almost half of it, CAD 1,500, is market growth.

Difference in differences. Subtracting the control's change in levels gives $3{,}230 - 1{,}260 = 1{,}970$. Because the growth is proportional and the two legs differ in size, a proportional counterfactual is better: without the upgauge the leg would have earned $25{,}000 \times 1.06 = 26{,}500$, so the effect is $28{,}230 - 26{,}500 = 1{,}730$. Both readings rest on the parallel-trends assumption discussed in Section 7.8.

The network. Even the right leg estimate misses the network: the simulated network gain is CAD 2,393, of which 664 appears on the Paris, Seville and Boston legs (Example 22.2). The leg's own report understates the gain by 28 percent, and the naive before-and-after comparison overstates it by 35 percent.

The planner learns to read a first season with a control group, at the network level, and with its uncertainty attached. Such designs are used in airline practice: an Air Canada presentation at the 2026 AGIFORS scheduling and strategic planning meeting used difference in differences to measure the effect of the closure of Russian airspace on flights between North America and India (AGIFORS 2026).

22.8 Questions for the revenue management team

The handbook lists the questions a planner should put to revenue management: which cabins and departures constrain demand, what incremental capacity is worth, where rejected demand would be recaptured, and how fares and availability would change after a schedule adjustment, with sensitivity to competitor fares and to

changes in booking curves (Network Planning Handbook 2026). Its first-week table adds where the valuable capacity constraints are, which fare and cabin assumptions planning should use, and how schedule changes affect availability and displacement (Network Planning Handbook 2026). Table 22.3 expands them into a checklist with the reason behind each question.

Table 22.3 Questions for the revenue management team, expanded from the handbook's list.

Question	Why it matters	See
Which legs, cabins and departures constrain demand, and how often?	capacity may be scarce only on peaks	Section 22.2
What are the bid prices or displacement costs, at what level, and how often are they recomputed?	a seasonal average hides the departures that matter	Chapter 20
What is the change worth, re-optimized under random demand and the control actually used?	bid price times seats overstates large changes	Section 22.3
Where would rejected demand be recaptured?	recapture needs seats elsewhere	Section 9.7
Which demand forecast is unconstrained, and how was it unconstrained?	bookings understate demand	Section 22.6
Do the revenue management and planning forecasts agree, and which vintage is each?	different vintages explain many disputes	Section 8.5
Which fares and cabin mix should planning assume?	the next passenger is not the average one	Section 22.4
How much buy-down and sell-up does the system assume, and does it adjust fares for them?	weak fences can make extra seats worthless	Chapter 21
How sensitive are the results to competitor fares and booking curves?	competitors respond to capacity	Section 21.8
How will availability and displacement change on other legs?	the effects fall on other flights	Example 22.4
How long will the forecasts take to adapt, and what is used meanwhile?	the first season is sold on old forecasts	Section 22.7
For changes to flights on sale, how many bookings are affected?	downgauges create oversales	Chapter 19
Can the system reproduce a past change, and what did its review find?	trust is earned on known results	Section 12.9

The horizons of Figure 22.1 imply four points at which the two functions must meet: the seasonal schedule build, in which planning proposes and revenue management supplies demand and the value of capacity; changes to schedules already on sale, in which

revenue management re-optimizes inventory and booked passengers must be re-accommodated; late changes of aircraft type, which revenue management can value best because its forecasts are then sharpest; and the post-implementation review, in which both compare outcomes with the approved forecast. The handbook's plan for a new analyst's second month asks for this collaboration explicitly: work with forecasting and revenue management on demand and displacement, with scheduling and operations on feasibility, and with Finance on the contribution bridge (Network Planning Handbook 2026).

How the teams are organized varies by airline, and public sources say little. For Air Canada they say that the Chief Commercial Officer's portfolio includes both the development and performance of the network and the airline's yield management and pricing strategies (Air Canada 2025), and a vendor case study names a Director of Revenue Optimization (PROS 2025). The handbook's first-week questions are the way to learn the rest.

IN PRACTICE **The interface in 2026**

Commercial network planning systems evaluate schedules by chaining demand and market-share models, spill and recapture models, and cost and revenue models into O&D-based profitability; Lufthansa Systems describes its NetLine/Plan product in these terms (Lufthansa Systems 2026). Research presented at the 2026 AGIFORS scheduling and strategic planning meeting pushes toward joint models: Sabre presented a cabin-level O&D fleet assignment model that uses itinerary demand and cabin fares, Embry-Riddle Aeronautical University a capacity allocation model in which demand responds to the seats offered, and Stanford University a research prototype that links forecasting, scheduling, fleet assignment, maintenance, crew and revenue management through one shared demand model with feedback from flown outcomes (AGIFORS 2026). On the selling side, a factsheet published by IATA and sourced from the Lufthansa Group describes continuous pricing with an unlimited number of price points instead of at most 26 booking classes (International Air Transport Association 2021), so a planner may work with a revenue management system that has no classes at all; a vendor case study reports that Air Canada uses a PROS real-time dynamic pricing product (PROS 2025). The planner's questions do not change: what does a seat

on this departure displace, and how will the system that sells it respond to more or fewer seats?

22.9 Summary

- Planning creates capacity on horizons of years to weeks; revenue management decides who gets it over about 300 days of bookings. The value of capacity depends on the demand, the network and the control that sells it.
- Bid prices identify scarce legs and value small changes. Under random demand they are fare differences that jump from departure to departure; their mean is a supergradient of the expected ex post value and overstates large additions.
- The same extra seats are worth from CAD 8,030 to 14,700 per bank on the Paris leg, and from 459 to 5,088 on the Quebec City feeder, depending on the method. The bid price times the seats is the most optimistic in both cases.
- In the simulations, the better the control, the less extra seats are worth: capacity and control are substitutes. The Quebec City upgauge pays under simulated bid-price control and fails under perfect control.
- On flights that revenue management controls, the classical spill model misprices capacity. Without fences and with protection that ignores buy-down, extra seats can lose money; re-optimized protection prevents it.
- A new long-haul leg displaces locals on full feeders. Lisbon contributes about CAD 28,000 per bank under simulated revenue management after CAD 7,563 of displacement, and its leg-based spill cost is 42 percent too high.
- Planning and revenue management forecasts must agree on what is unconstrained, on O&D versus leg, on vintage, on fares and on customer choice.
- A first season misleads: revenue management is still learning, demand is ramping up, markets move, network effects appear on other legs and one season is statistically noisy. Read it with a control group, at the network level.

Table 22.4 Key formulas of the chapter.

Quantity	Formula
Value of capacity under control u	$R_u(C') - R_u(C)$
Mean ex post bid prices	$\bar{V}(C') \le \bar{V}(C) + \sum_l \bar{\pi}_l (C'_l - C_l)$
Marginal seat, fenced	$f_L \Pr(D_L > b) + f_H \Pr(D_L \le b,\ D_L + D_H > C)$
Marginal seat, no fences	$f_L \Pr(D_L > b$ or $D_L + D_H > C)$ minus $(f_H - f_L) \Pr(D_L \le b < D_L + D_H \le C)$
Network contribution	$\Delta\mathrm{NC} = \Delta R_u - \sum_l (\Delta \mathrm{TC}_l + c_l^{\mathrm{pax}} \Delta \mathbb{E}[Y_l])$
Leg demand	$D_l = \sum_j a_{lj} D_j$

22.10 Further reading

Belobaba's introduction to the planning process (Belobaba 2006c) sets out the sequential practice and the case for integrated planning; his spill lecture (Belobaba 2006b) and his review of PODS research on O&D control (Belobaba 2002) are the library's sources for the two halves joined here, and the handbook's sections on revenue management and on first-week questions (Network Planning Handbook 2026) are the source of Table 22.3. Abramovich (2013) is the essential study of spill under revenue management. Talluri and van Ryzin (2004b, ch. 3) give the theory of network bid prices and their estimation from random demand; Barnhart, Kniker and Lohatepanont (2002) value spill at the network level in fleet assignment, and Lohatepanont and Barnhart (2004) integrate schedule design with fleet assignment. Cooper, Homem-de-Mello and Kleywegt (2006) analyze the spiral-down effect formally, and Fiig and coauthors (2010) the fare adjustment that protects revenue when fences are weak. Belobaba, Odoni and Barnhart (2015) place both functions within the airline business.

PART V

Practice

The last part turns from models to the conditions in which they are used. Chapter 23 describes the industry as it stands in October 2026: pricing technology, planning systems, aircraft supply, carbon rules, the Canadian market and Air Canada’s public position. Chapter 24 assembles the analyst’s working toolkit: the data, the workflow, the checks and the one-page memo that carries a recommendation.

CHAPTER

23

The Industry in 2026

IN THIS CHAPTER

- How airlines price and sell in 2026: continuous pricing, dynamic offers, NDC and orders, and how much of it is in production.
- What finer price points can earn, derived from first principles, and how to read vendor claims about machine learning.
- Planning systems in 2026, and why they still run the model chain of this book.
- Engines, delivery backlogs and the A321XLR: why aircraft time is scarce, and which thin markets a long-range narrowbody opens.
- The carbon rules that enter route economics, and why their cost depends on the route and even the direction of flight.
- The Canadian market and Air Canada's public position in the autumn of 2026.
- What changed since the 2006 teaching material, and what did not.

DRAWS ON Air Canada, Second quarter 2026 MD&A (11 August 2026); Competition Bureau, Cleared for take-off (2025); Lufthansa Systems, NetLine/Ops ++ release (2025); handbook; public sources dated up to 28 October 2026

The methods in this book come mostly from teaching material written in 2003 and 2006. The airline you are joining operates in 2026, and between the two dates the industry changed how it sells seats, which aircraft it can buy, which regulations price its emissions and who it competes with. None of these changes repeals the economics of the earlier chapters, but each changes an input, a constraint or a tool. This chapter is a dated snapshot of those changes, taken at the end of October 2026, just before Air Canada was due to report its third-quarter results on 29 October (Network Planning Handbook 2026). Every fact in it carries a source; facts of this kind age quickly, so recheck them before you rely on one at work.

The chapter has a second purpose. Industry news arrives as press releases, vendor brochures, investor presentations and regulatory filings, each written by someone with an interest in how you read it.

A planner needs the habit of separating what is measured from what is claimed, and of translating each announcement into the question it raises for a schedule. The sections below practise that habit on pricing technology, aircraft supply, carbon policy, the Canadian market and Air Canada itself. Chapter 24 then turns the same discipline on data.

23.1 Pricing and revenue management technology

Revenue management grew up inside a distribution system built for travel agents. Fares were filed in advance with a fare-filing utility, distributed through global distribution systems (GDS), and mapped to booking classes, one letter each, so that an airline could offer at most 26 price points on a flight, shared among its cabins (International Air Transport Association 2021). Revenue management controlled which classes were open (Chapters 18 and 20). Three developments, all visible in 2026, loosen those constraints.

> **DEFINITION 23.1 Dynamic pricing, continuous pricing and dynamic bundling**
>
> In IATA's terminology (International Air Transport Association 2021), **dynamic pricing** sets the price of an offer from the context of the request at the time of shopping (days to departure, length of stay, day of week, competitor activity, remaining seats) without requiring personal data. **Continuous pricing** removes the limit of a fixed number of price points, so that the price can take any value as demand changes. **Dynamic bundling** assembles the content of the offer itself (fare, bags, seats, flexibility) at the time of the request.

Lufthansa Group began selling continuously priced fares in October 2020 (Travel Daily 2020): on most European continental routes of Lufthansa, SWISS and Austrian Airlines, fares can fall between the traditional fixed price points when this gives the customer a lower price, and they are sold only through the airline's NDC interface and its own websites, not through traditional GDS channels (Business Travel News Europe n.d.). The economics of this step are worth deriving, because they explain both why it was worth taking and why its gains are modest.

What a finer price grid can earn Consider one booking request. If the airline quotes a price p, the customer buys with probability $q(p)$,

a smooth decreasing function, and a sale consumes a seat whose opportunity cost is π, the bid price of Chapter 20. The expected contribution of the quote is $R(p) = (p - \pi)\, q(p)$. Continuous pricing can quote the maximizer p^*. A class-based system quotes a price from a fixed ladder; at best, with the fare adjustment methods of Chapter 21, it picks the ladder price with the highest R. The difference between the two is a rounding loss.

RESULT 23.1 The price-grid loss is second order

Let R be twice continuously differentiable with an interior maximum at p^* and $R(p^*) > 0$, and define $\kappa = -R''(p^*)/R(p^*)$. Quoting $p^* + \delta$ loses a share

$$\ell(\delta) = 1 - \frac{R(p^* + \delta)}{R(p^*)} = \tfrac{1}{2}\,\kappa\,\delta^2 + o(\delta^2)$$

of the attainable contribution. If the ladder prices around p^* are h apart, the best ladder price loses at most $\kappa h^2/8$ to second order, and if p^* is equally likely to fall anywhere between two ladder prices, the expected loss is $\kappa h^2/24$. With exponential willingness to pay, $q(p) = e^{-p/\theta}$, the optimum is $p^* = \pi + \theta$, $\kappa = 1/\theta^2$, and the exact loss is $\ell = 1 - (1 + x)e^{-x}$ with $x = \delta/\theta$.

Proof. Expand around the maximum, where $R'(p^*) = 0$: $R(p^* + \delta) = R(p^*) + \frac{1}{2}R''(p^*)\delta^2 + o(\delta^2)$; dividing by $R(p^*)$ gives the first statement. The nearest ladder price is at most $h/2$ away, and the best ladder price does at least as well, so the loss is at most $\frac{1}{2}\kappa(h/2)^2 = \kappa h^2/8$. If the distance δ to the nearest price is uniform on $[-h/2, h/2]$, then $\mathbb{E}[\delta^2] = h^2/12$ and $\mathbb{E}[\ell] \approx \kappa h^2/24$. For the exponential case, $R'(p) = e^{-p/\theta}(1 - (p - \pi)/\theta)$ vanishes at $p^* = \pi + \theta$, and $R(p^* + \delta)/R(p^*) = (1 + \delta/\theta)\, e^{-\delta/\theta}$. Expanding, $(1 + x)e^{-x} = 1 - x^2/2 + x^3/3 - \ldots$, so $\kappa = 1/\theta^2$. ∎

Two features of the result matter. First, the loss is quadratic in the step: halving the distance between price points cuts the loss by three quarters, so the first few extra price points are worth much more than the last ones. Second, the curvature has an economic reading. With exponential willingness to pay the purchase probability has price elasticity $-p/\theta$, so at the optimum the absolute elasticity is $|\eta^*| = p^*/\theta = p^*/(p^* - \pi)$, the Lerner condition of industrial organization, and $\kappa\, p^{*2} = \eta^{*2}$. On a ladder whose steps are a constant share s of the price, the expected loss is therefore close to $(s\,\eta^*)^2/24$: rounding hurts most where demand is most price sensitive.

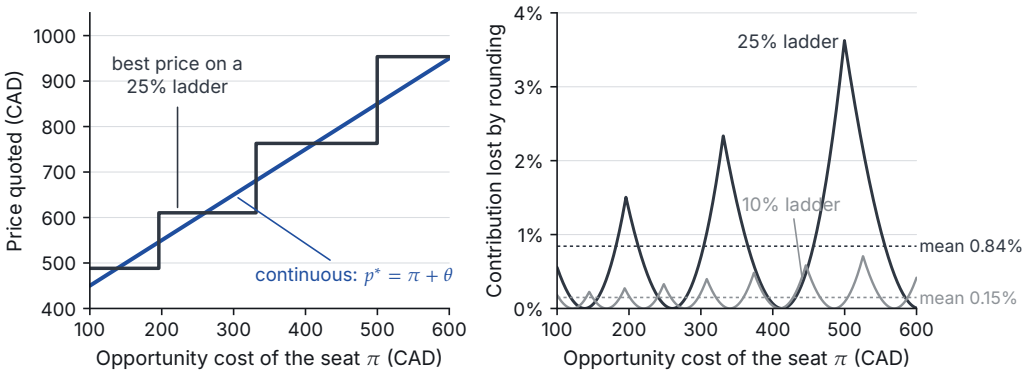


Figure 23.1 Continuous pricing against a fixed ladder of price points, for requests with exponential willingness to pay (scale θ = CAD 350) and a seat opportunity cost π between CAD 100 and 600. Left: the contribution-maximizing price moves continuously with π; a ladder whose steps are 25 percent apart can only follow it in jumps. Right: the share of contribution lost by rounding to the best ladder price, for 25 percent and 10 percent steps, with the averages over the range shown.

WORKED EXAMPLE 23.1 How much is a finer ladder worth?

Take requests with exponential willingness to pay, θ = CAD 350, and a bid price of π = CAD 300, so that continuous pricing quotes p^* = CAD 650. Compare two geometric ladders starting at CAD 200: one with 25 percent steps (CAD 488, 610, 763, 954, ...) and one with 10 percent steps (CAD 571, 628, 691, ...).

On the coarse ladder the neighbours of 650 are 610 and 763. By Result 23.1 the loss at 610 is $1-(1-0.113)e^{0.113} = 0.69$ percent and at 763 it is $1-(1+0.323)e^{-0.323} = 4.2$ percent, so the best ladder price is 610 and the airline gives up 0.69 percent of the attainable contribution on this request. On the fine ladder the best price is 628, with a loss of 0.21 percent.

Averaged over bid prices spread evenly between CAD 100 and 600 (Figure 23.1), the coarse ladder loses 0.84 percent and the fine ladder 0.15 percent; the worst cases are 3.6 and 0.7 percent. A ladder of 26 price points spanning a tenfold range of fares has steps of about 10 percent, close to the fine case.

The planner's lesson is about orders of magnitude. Rounding alone costs a few tenths of a percent once a cabin has many price points in use. Gains larger than that, such as the simulation results and vendor claims discussed next, must come from other sources: better estimates of willingness to pay, segmenta-

tion of requests, competitive effects, or ladders that are coarser in practice than on paper because few classes are open or filed fares are stale.

The simulation evidence agrees with this arithmetic. In the MIT Passenger Origin-Destination Simulator (PODS), continuous pricing raised revenue by up to 2.0 percent for the first airline to adopt it and up to 1.2 percent when both competitors adopted it; the gains were smaller at lower demand-to-capacity ratios and depended on how many fare classes the baseline structure already had (Szymański, Belobaba, and Papen 2021). A segmented version, which quotes separate continuous prices to requests classified as business or leisure with 80 percent accuracy, gave the first mover as much as 17 percent more revenue in PODS, mostly by undercutting rivals on late leisure fares and at their expense; rivals could offset the gain by removing fare restrictions while keeping traditional revenue management (Long and Belobaba 2023). Large gains, in other words, come from competition and segmentation, and they invite responses.

Offers, orders and NDC The second development changes the plumbing. The **New Distribution Capability (NDC)** is an IATA data standard through which an airline sends its own offers, including dynamic prices and bundles, to sellers through programming interfaces instead of through filed fares in a GDS. **ONE Order** goes further: it replaces the passenger name record, the electronic ticket and the electronic miscellaneous document with a single customer order record, so that accounting follows payment for the services ordered rather than tickets (IATA 2026c).

Adoption is real but gradual, and the numbers are worth knowing precisely. NDC transactions were 21.2 percent of transactions settled through the Airlines Reporting Corporation (ARC), the US agency settlement system, in December 2025, against 20.3 percent a year earlier (Airlines Reporting Corporation 2026a), and 21.6 percent in May 2026 against 21.4 percent in May 2025 (Airlines Reporting Corporation 2026b). ARC's orders-based settlement went live for cash payments in January 2026, with card payments still in development (Travel Distribution News 2026). At its October 2026 symposium IATA counted seven airlines live or implementing Orders, more than twenty expected to follow from 2027, and over US$1.7 billion committed by technology providers (IATA 2026e). In October 2022 ATPCO set the goal of 80 percent of sold offers being dy-

namically created by 2026; the latest estimate found for this book, published in July 2023, was 23 percent, up from 6 percent in October 2022 (ATPCO 2023). No 2026 progress figure was found, so whether the goal was met cannot be stated.

Air Canada launched its NDC program on 14 June 2023. In June 2024 it narrowed its distribution cost recovery charge, a fee on bookings made through traditional GDS connections, to the lowest fare in each cabin (classes G, K, A and P), with NDC, direct and group bookings exempt; NDC was then enabled in 20 points of sale covering 96 percent of its ticket volume (Travelweek 2024). Its pricing uses PROS Real Time Dynamic Pricing Advantage, according to the vendor (PROS 2025), and in July 2026 it added basic fares to its premium economy and business cabins in several international markets (Open Jaw 2026a). Which forecasting and optimization engine Air Canada's revenue management runs in 2026 is not public.

Machine learning, and how to read a claim The third development is the one most often oversold. Vendors now sell forecasting and pricing engines built on machine learning, and some report large gains. Table 23.1 sorts the main 2026 claims by the kind of evidence behind them.

Table 23.1 Claims about pricing and revenue management technology in 2026, by type of evidence. Sources: Szymański, Belobaba, and Papen (2021); Long and Belobaba (2023); PhocusWire (n.d.); ePlane AI (2025); PhocusWire (2025); Yeoman (2025).

Claim	Source	Evidence
Continuous pricing: up to +2.0% revenue for a first mover, +1.2% if both adopt	MIT PODS study	peer-reviewed simulation
Segmented continuous pricing: up to +17% for a first mover, at rivals' expense	MIT PODS study	peer-reviewed simulation
Willingness-to-pay data in forecasts: +1 to 3%	PROS	vendor analysis
Lufthansa uplift from 2.3% to 5.2%, net of continuous pricing	PROS and Lufthansa	vendor and customer statement
Per-request continuous price optimizer: up to +3.5%	Sabre	vendor projection
Delta prices about 3% of its domestic network with a generative-AI tool (mid-2025)	Delta	company statement, no revenue result
Reinforcement learning against dynamic programming: winner depends on data volume	Academic papers	academic comparison

The peer-reviewed results are simulations, carefully done, of a stylized competitive market. The vendor numbers are self-reported, usually without a published method or control group, and some come from launch customers. Delta, which tested a generative-AI pricing tool from Fetcherr on about 3 percent of its domestic network by mid-2025 and aimed for 20 percent by the end of that year, disclosed no revenue result (PhocusWire 2025); in a September 2026 letter to a member of the US Congress it described its generative-AI pricing tool, without naming the vendor, as decision support for analysts, fed with route-level demand and market data and no personal information (Delta Air Lines 2026). Academic comparisons of reinforcement learning with dynamic programming report that the better method depends on how much data is available (Yeoman 2025);

no peer-reviewed field evidence of reinforcement learning controlling airline inventory at scale was found for this book.

> **PITFALL**
>
> Treating a vendor's uplift as a forecast for your airline. A percentage gain depends on the baseline it is measured against, on the competitive response, and on the counterfactual. Ask what the baseline system was, how the counterfactual was constructed, over what period, and whether rivals reacted. Result 23.1 gives one sanity bound: price granularity alone is worth tenths of a percent, not whole percentage points.

23.2 Network planning practice and tools

The planning systems of 2026 run the same chain of models this book derives: a market-size and share model, a spill and recapture model, and a cost and revenue model that produce origin-and-destination (O&D) profitability. Lufthansa Systems describes its NetLine/Plan product in exactly these terms, as an engine of demand and market share, spill and recapture, and cost and revenue models that evaluates a scenario in under 30 seconds across up to 250,000 markets, delivered as a cloud service; the vendor also says its platform plans about 20 percent of the world's available seat kilometres (Lufthansa Systems 2026). The chain is the one of Chapters 4, 5, 9 and 12: share models in the QSI or logit family, spill and recapture as in Chapter 9, and contribution with the accounting conventions of Chapter 12. What changed is speed, scale and calibration data, not structure.

At Air Canada, Lufthansa Systems confirms only the operations side: the airline went live with NetLine/Ops ++, its operations control system, by January 2025, and also uses its Lido Flight 4D flight planning tool (Lufthansa Systems 2025a). A June 2026 posting for a schedule optimization analyst, reproduced by a job aggregator (and removed by late October), places a Scheduling and Network Planning team in the Network Planning branch, describes schedule adjustments from 3 to more than 300 days before departure using profitability and yield measures, fleet assignment, the integration of regional flying and slot portfolio management, and lists Sabre Schedule Manager as an asset (ResumeGeni 2026). Such traces are clues, not an organization chart; the handbook's advice is to ask which

system is authoritative for published schedules and who owns each model input (Network Planning Handbook 2026).

The research agenda shows where practice is heading. The tentative program of the 2026 AGIFORS Schedule and Strategic Planning meeting includes an Air Canada presentation on large-scale aircraft routing under uncertainty with simulation-derived risk measures, a cabin-level O&D fleet assignment model from Sabre, integrated network planning and scheduling with bank structures from HEC Montréal, KLM and the University of Twente, Monte Carlo stress tests of future schedules, and capacity allocation with endogenous demand stimulation (AGIFORS 2026). Each item extends a model of Part III (Chapters 13, 14 and 16) rather than replacing it.

IN PRACTICE **Re-cutting a plan in a fuel shock**

The year 2026 tested every planning calendar. After conflict in the Middle East disrupted shipping through the Strait of Hormuz in the first half of the year (Air Canada 2026e), jet fuel prices surged; IATA's June outlook assumed jet fuel at US$152 per barrel for the year, almost 70 percent above 2025, and an industry net margin of 2.0 percent (IATA 2026d). Air Canada suspended its full-year guidance on 30 April 2026 (The Canadian Press 2026) and reinstated it on 11 August with capacity growth cut from 3.5 to 5.5 percent to 2.25 to 3.25 percent, on jet fuel assumed at about CAD 1.38 per litre in the third quarter against CAD 0.90 originally (Air Canada 2026d). WestJet ended several transatlantic and sun routes in September and October 2026, citing fuel costs and capacity alignment (Aviation A2Z 2026). On 22 October the Argus US jet fuel index reached a record US$206 per barrel (Open Jaw 2026b). In planning terms, a fuel shock raises the avoidable cost of every flight in proportion to its fuel burn, so it lowers the contribution of long, thin and low-yield flying first; the re-cut is a fresh ranking of flying by contribution per aircraft-hour, against a shadow price of aircraft time (Section 23.3) that the cuts themselves lower.

23.3 Supply constraints: engines, deliveries and the long-range narrowbody

For most of the history of network planning, a planner could assume that the aircraft in the fleet plan would exist. In 2026 that assumption fails in two ways.

The first is engines. Durability problems with Pratt & Whitney's geared turbofan (GTF) engines grounded about 650 A320neo-family aircraft at the peak; in July 2026 Pratt said the issues should be behind it in 2027 or 2028, while one operator expected two or three more years of disruption, and Pratt's shop capacity for these engines was up 300 percent since 2019 (Hamilton 2026). On the A220, whose PW1500G engines belong to the same family, Airbus and Pratt said in June 2026 that no aircraft would remain grounded for engine reasons by the end of the year (Leeham News 2026). The exposure is direct for Canadian carriers: Air Canada's A220-300s fly PW1500G engines (Air Canada n.d.a) and it chose Pratt & Whitney GTF engines for its A321XLRs (Aviacionline 2022); Air Transat still cited GTF problems in its results for the quarter to 31 July 2026 (Transat A.T. 2026) and in December 2025 pushed its first A321XLR deliveries back to 2027 (FlightGlobal 2025).

The second is deliveries. IATA counted an order backlog of 18,100 aircraft in May 2026, almost 60 percent of the active fleet, and estimated that deliveries had fallen about 5,600 aircraft short of the pre-pandemic trend; airlines absorbed an effective shortage of about 3,170 aircraft through higher utilization (about 890), delayed retirements (about 1,700) and higher load factors (about 580), and the average fleet age rose above 15 years (IATA 2026d). The FAA certified the Boeing 737-10 on 20 October 2026, with 1,555 orders on Boeing's books (Boeing 2026). Air Canada reduced its 787-10 order from 18 to 14 in November 2025, with deliveries scheduled to begin in 2026 (Air Canada 2026e).

The value of a late aircraft Scarcity has a price, and a planner should know how to compute it.

> **DEFINITION 23.2 Shadow price of aircraft time**
>
> The **shadow price** of an aircraft over a period is the change in the network's optimal contribution when one more aircraft of that type is available for the period, all other plans being re-optimized. It is the opportunity cost of using the aircraft for any particular flying, and the standard against which every use of it must be judged.

RESULT 23.2 A missing aircraft costs the marginal line

Suppose K identical aircraft are available for a period and the candidate lines of flying, each needing one aircraft, have contributions $v_1 \geq v_2 \geq \dots \geq v_J > 0$ for the period. The best plan flies the K most valuable lines, with value $V(K) = \sum_{j \leq K} v_j$. The function V is concave, losing one aircraft costs exactly v_K, gaining one adds v_{K+1}, and the shadow price of an aircraft lies between v_{K+1} and v_K. A delivery that slips by d periods therefore costs $d\,v_K$, not d times the average line.

Proof. Any plan with at most K lines is improved or unchanged by swapping a flown line for a more valuable unflown one, so the K most valuable lines are optimal. Then $V(K) - V(K-1) = v_K \geq v_{K+1} = V(K+1) - V(K)$, which is concavity. In the linear relaxation, maximize $\sum_j v_j x_j$ subject to $\sum_j x_j \leq K$ and $0 \leq x_j \leq 1$; its dual minimizes $K\lambda + \sum_j \mu_j$ subject to $\lambda + \mu_j \geq v_j$ and $\lambda, \mu_j \geq 0$, and any $\lambda \in [v_{K+1}, v_K]$ with $\mu_j = \left[v_j - \lambda\right]^+$ is optimal with the same objective value $V(K)$, so λ is the price of one aircraft. ∎

WORKED EXAMPLE 23.2 An A321XLR delivered eight weeks late

Laurentian plans six A321XLRs for the summer and has seven candidate lines of flying for them, with illustrative contributions of CAD 70,000, 58,000, 51,000, 44,000, 37,000, 29,000 and 22,000 per week. With six aircraft it flies the first six lines and earns CAD 289,000 a week; the seventh line is the best use it forgoes.

Suppose an engine shortage delays the sixth aircraft by eight weeks. By Result 23.2 the cost is $8 \times 29{,}000 = $ CAD 232,000: the airline drops its least valuable line, not an average one. Valuing the delay at the average line, CAD 48,167 a week, would give CAD 385,000 and overstate the loss by two thirds. While the fleet is short, the shadow price of an aircraft-week rises from somewhere between CAD 22,000 and 29,000 to between CAD 29,000 and 37,000, so any proposal to add flying in those weeks must beat the higher hurdle.

The lesson generalizes. Engine groundings and late deliveries make the marginal line, not the average, the relevant cost, and they raise the hurdle that new routes must clear. In a real net-

work, lines differ in the aircraft types that can fly them and connect to each other through feed, so the marginal value comes from a fleet assignment model (Chapter 14); the logic is the same.

The long-range narrowbody The aircraft that defines network planning in 2026 is the long-range narrowbody.

> **DEFINITION 23.3 Long-range narrowbody**
>
> A **long-range narrowbody** is a single-aisle aircraft with enough range to fly long-haul sectors, such as transatlantic flights from eastern North America. The Airbus A321XLR, with a published range of 4,700 nautical miles and between about 140 and 240 seats depending on the airline, is the current example (Cirium 2026b).

By the third quarter of 2026, 50 A321XLRs had been delivered and 44 were in revenue service; 28 airlines had committed to 440 aircraft. Cabins range from 138 seats (JetBlue) to 239 (Wizz Air), with Air Canada and Iberia at 182. Cirium reports up to 30 percent lower fuel consumption per seat than the Boeing 757, the aircraft this role used to fall to; it lists the trade-offs as lost belly cargo and the passenger preference for widebodies, and expects Air Canada and five other operators to fly more than 500 round trips a month on routes longer than 5,000 kilometres by the second quarter of 2027 (Cirium 2026b). Boeing's newly certified 737-10 is rated for up to 3,100 nautical miles (Boeing 2026), well short of the XLR. At its December 2024 investor day Air Canada claimed about 30 percent lower fuel burn than its older A321s and about 12 percent better unit revenue (Skies Magazine 2024); these are company claims. The economics of the aircraft are developed in Chapter 11; the next example shows the network consequence.

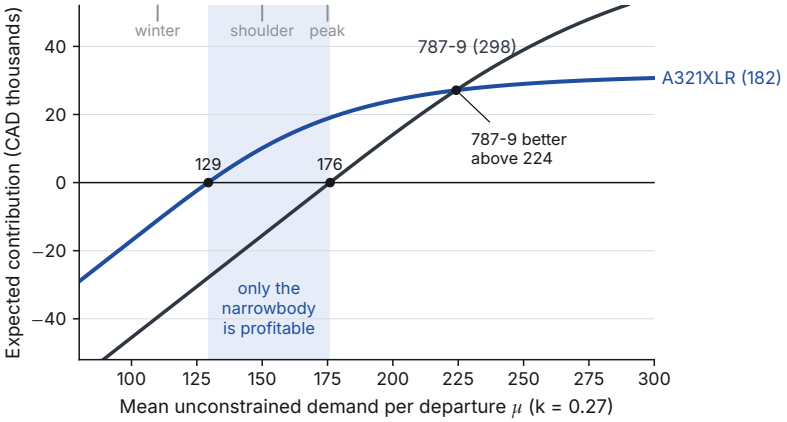


Figure 23.2 Expected contribution per eastbound YUL–SVQ departure for Laurentian's A321XLR (182 seats) and 787-9 (298 seats) as mean demand grows, with demand uncertainty $k = 0.27$, CAD 600 contribution per passenger and illustrative cargo contributions of CAD 3,500 and 10,000. The band marks demand levels at which only the narrowbody is profitable. The ticks at the top mark Laurentian's winter, shoulder and peak mean demands.

WORKED EXAMPLE 23.3 A thin transatlantic market

Laurentian's YUL–SVQ departure carries passengers worth $r = 620 + 30 - 50 =$ CAD 600 each (net fare, ancillary revenue and passenger costs from the running case). Its trip cost excluding passenger costs is $11{,}000 \times 7.0 + 3{,}500 =$ CAD 80,500 on the A321XLR and $15{,}500 \times 7.0 + 7{,}000 =$ CAD 115,500 on the 787-9; per seat, the widebody is cheaper (CAD 388 against 442). Assume belly cargo contributes CAD 3,500 on the narrowbody, the handbook's figure, and CAD 10,000 on the widebody.

With shoulder-season demand $\mathcal{N}(150, 40^2)$, the spill formula of Chapter 9 gives an expected load of 145.2 on the XLR (79.8 percent load factor) and a contribution of $600 \times 145.2 + 3{,}500 - 80{,}500 =$ CAD 10,115 per departure. The 787-9 carries all 150 expected passengers at a 50.3 percent load factor and loses CAD 15,501. At peak, $\mathcal{N}(175, 42^2)$, the XLR earns CAD 19,907 and the 787-9 still loses CAD 512; in winter, $\mathcal{N}(110, 35^2)$, both lose money, the XLR CAD 11,153 and the 787-9 CAD 39,500.

Figure 23.2 generalizes. The XLR breaks even at a mean demand of about 129 passengers per departure, the 787-9 at about 176, and the 787-9 becomes the better aircraft only above about 224. Between 129 and 176 only the narrowbody can serve the

market profitably. Each extra CAD 6,000 of widebody cargo contribution would lower the 787-9 break-even by 10 passengers.

The lower seat cost of the widebody is irrelevant when its seats cannot be filled; what matters is trip cost against the revenue the market can supply. This is how a long-range narrowbody creates markets such as Montreal–Seville that a widebody could not serve. The winter loss is a reminder that such markets are often seasonal: Air Canada's own published Seville schedule pauses between early January and mid-February 2028 (Air Canada 2026g), although nothing in Laurentian's invented numbers explains the real airline's choice.

23.4 Carbon rules that enter route economics

Carbon policy reaches a route through three instruments with different scopes, and the handbook's warning applies: do not assign one carbon charge to every tonne on every route, and do not assume that an EU-bound transatlantic flight is treated like a flight within Europe (Network Planning Handbook 2026).

DEFINITION 23.4 Three carbon instruments

The **Carbon Offsetting and Reduction Scheme for International Aviation** (**CORSIA**) of ICAO requires operators to buy eligible emissions units for the growth of international aviation emissions above a baseline, on routes between participating states. **ReFuelEU Aviation** requires fuel suppliers at airports in the European Union to blend rising minimum shares of **sustainable aviation fuel** (**SAF**) and requires operators to buy most of their fuel there. The **EU Emissions Trading System** (**EU ETS**) requires operators to surrender an allowance for each tonne of CO_2 emitted on covered flights, a scope that in 2026 covers flights within the European Economic Area.

CORSIA CORSIA's pilot phase ran from 2021 to 2023 and its first phase runs from 2024 to 2026; in both, only flights between volunteering states carry offsetting obligations. From 2027 the second phase covers all international flights except those to or from exempt states (least developed countries, small island developing states, landlocked developing countries and states with very small traffic)

unless they volunteer. From 2024 the baseline is 85 percent of 2019 emissions. As of 1 January 2026, 130 states participated, and the 2024 offsetting requirement was 55.6 million tonnes of CO_2, with a sector growth factor of 15.4 percent; IATA expects first-phase demand of 170 to 236 million units against an eligible supply nearing 40 million (IATA 2026b). An operator's requirement in a year is its covered emissions multiplied by the sector growth factor; the requirement is fully sectoral until 2032 and gives a 15 percent weight to the operator's own growth from 2033 (ICAO 2023). Emissions are computed from fuel: 3.16 tonnes of CO_2 per tonne of Jet A or Jet A-1 (Hellenic Civil Aviation Authority n.d.). Canada applies CORSIA by regulation to operators emitting more than 10,000 tonnes of CO_2 a year on international flights (Transport Canada 2026); domestic flights are outside it (Network Planning Handbook 2026).

ReFuelEU Aviation Regulation (EU) 2023/2405 sets minimum SAF shares in the fuel supplied at Union airports of 2 percent from 2025, 6 percent from 2030, 20 percent from 2035, 34 percent from 2040, 42 percent from 2045 and 70 percent from 2050, with a synthetic-fuel sub-mandate that starts at an average of 1.2 percent over 2030 and 2031 and reaches 35 percent in 2050; to stop airlines from carrying cheaper fuel in from elsewhere, operators must uplift at each Union airport at least 90 percent of the fuel their flights from it require (European Parliament and Council of the European Union 2023; EASA 2024). The blending obligation falls on fuel suppliers, who can be expected to pass its cost on in the price of fuel. In EASA's 2025 reference prices, fossil jet fuel averaged EUR 640 per tonne, aviation biofuel EUR 1,925 and synthetic fuel an estimated EUR 7,520 (SAF Investor 2026).

The EU ETS Aviation has been in the EU ETS since 2012, but its scope was temporarily limited to flights within the European Economic Area, with flights to the United Kingdom and Switzerland handled under separate arrangements, a limitation extended one last time until the start of 2027; free allocation fell by 25 percent in 2024 and 50 percent in 2025, and the sector moved to full auctioning in 2026. Reporting of non-CO_2 effects began on 1 January 2025 and extends to all flights to and from the European Economic Area from January 2027 (European Commission n.d.). On 17 July 2026 the European Commission proposed extending the ETS, from 2029, to departures from the European Economic Area to destinations within 5,000 kilometres of Frankfurt, with a reassessment of CORSIA in 2032; the proposal needs the agreement of the Parliament and the

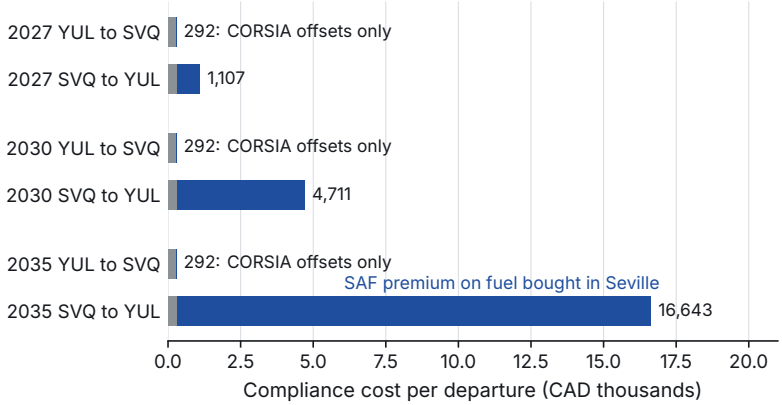


Figure 23.3 Carbon offsetting and SAF-mandate cost per Laurentian YUL–SVQ departure by direction, for 2027, 2030 and 2035 mandate levels, from (23.1). Fuel burn, unit prices and the euro exchange rate are illustrative assumptions, and 2025 reference fuel prices are held constant, which overstates later SAF costs if those prices fall.

member states (Legal 500 2026). If the 5,000 kilometre test were applied as reported, departures to Montreal (about 5,850 kilometres from Frankfurt by great circle) and Halifax (about 5,240) would fall outside it and departures to St. John's (about 4,370) inside; this is this book's calculation, not a reading of the legal text.

Canada has no federal SAF blending mandate. Its Aviation Climate Action Plan sets an aspirational goal of 10 percent SAF use by 2030, British Columbia requires 1 percent renewable jet fuel from 2028 rising to 3 percent in 2030, and SAF suppliers can earn credits under the federal Clean Fuel Regulations (Transport Canada 2026).

The cost per departure For one flight, the three instruments combine into a compliance cost that depends on where the fuel is bought. With F tonnes of fuel, emission factor $e = 3.16$, sector growth factor g (zero for a flight outside CORSIA), unit price p_C, allowance price p_E, SAF share s of which synthetic s_x, and price premiums Δ_b and Δ_x of biofuel and synthetic fuel over fossil fuel,

$$\begin{aligned} K = {} & F\,e\,(g\,p_C + \mathbb{1}_{\mathrm{ETS}}\,p_E) \\ & + \mathbb{1}_{\mathrm{EU}}\,F\,[(s - s_x)\,\Delta_b + s_x\,\Delta_x], \end{aligned} \tag{23.1}$$

where $\mathbb{1}_{\mathrm{ETS}}$ indicates that the ETS covers the flight and $\mathbb{1}_{\mathrm{EU}}$ that the fuel is bought at an EU airport.

WORKED EXAMPLE 23.4 Which direction pays for SAF?

Assume, for illustration only, that Laurentian's A321XLR burns 17 tonnes of fuel eastbound from Montreal and 19 tonnes westbound from Seville. The flights emit $3.16 \times 17 = 53.7$ and $3.16 \times 19 = 60.0$ tonnes of CO_2.

CORSIA. In 2027 the route is covered in both directions. At a sector growth factor of 15.4 percent (the 2024 value, used here for illustration) the offsets are 8.3 and 9.2 tonnes. At US$25 per unit, the low end of the range one market analyst projects for 2027 (Sylvera 2025), and CAD 1.41 per US dollar, Air Canada's 2026 planning rate (Air Canada 2026e), they cost CAD 292 eastbound and CAD 326 westbound.

ReFuelEU. Fuel bought in Montreal carries no mandate; fuel bought in Seville does, and anti-tankering rules prevent the airline from avoiding it. At the 2 percent share, the premium is $0.02 \times (1{,}925 - 640) =$ EUR 25.70 per tonne of fuel, or EUR 488 for 19 tonnes, CAD 781 at an assumed CAD 1.60 per euro. At the 2030 shares, 6 percent including 1.2 percent synthetic, the premium per tonne is $0.048 \times 1{,}285 + 0.012 \times 6{,}880 =$ EUR 144.24, or CAD 4,385 per westbound departure: CAD 24 per seat, CAD 29 per passenger at 150 passengers. At 2035 shares and unchanged prices it would be CAD 16,317.

EU ETS. Under the 2026 rules the route is outside the ETS in both directions. A flight within Europe would pay for every tonne: at the EUR 80 per allowance the Commission uses to value its SAF support (European Commission n.d.), that is $3.16 \times 80 =$ EUR 253 per tonne of fuel, about ten times the cost of the 2 percent SAF share.

The westbound flight thus carries most of the policy cost, and by 2030 that cost is of the same order as the CAD 20 per passenger by which the handbook's Seville candidate falls short of its alternative (Chapter 12). The example ignores interactions between the instruments and holds prices fixed; its point is the structure. Carbon cost is route-specific and direction-specific, it steps up on known dates, and a business case that applies one average price per tonne to the whole network will misrank routes.

Table 23.2 Which carbon instrument reaches which flight. Sources: IATA (2026b); European Parliament and Council of the European Union (2023); European Commission (n.d.); Transport Canada (2026).

Flight	CORSIA (2027 on)	SAF mandate on fuel bought	EU ETS (2026 rules)
Domestic Canada	no	none (British Columbia from 2028)	no
Canada to United States	yes	none	no
Montreal to Seville	yes	none in Montreal	no
Seville to Montreal	yes	ReFuelEU in Seville	no
Madrid to Seville	no (domestic)	ReFuelEU in Madrid	yes

PITFALL

One carbon price for every tonne. CORSIA prices only the growth share of international emissions, the EU ETS prices every tonne but only within its scope, and SAF mandates raise the price of fuel bought at particular airports. Get the approved fuel and carbon assumptions by region and year from the sustainability and finance teams, and carry fuel burned, the SAF premium, compliance payments and any internal shadow price of carbon as distinct lines (Network Planning Handbook 2026).

23.5 The Canadian market in 2026

Structure and the Competition Bureau's study On 19 June 2025 the Competition Bureau published a market study of the domestic airline industry. It found that Air Canada and WestJet together carried 56 to 78 percent of domestic passengers at major airports in 2023; that concentration fell by a tenth between 2019 and 2023 as Porter and Flair grew, but remained high; and that 2023 domestic passenger shares were 34.1 percent for Air Canada, 30.1 for WestJet, 9.8 for Flair and 9.0 for Porter. One additional competitor on a route lowered average fares by about 9 percent. Fees for airports, air navigation, security and fuel tax made up about 30 percent of what passengers pay full-service carriers (Competition Bureau Canada 2025). The full report adds that the 9 percent is about CAD 27.80, that aver-

age concentration remains more than three times the level the Competition Act's merger provisions treat as concentrated, that the average base-fare gap between Air Canada and WestJet on competing top routes widened from CAD 76 in 2019 to CAD 214 in the twelve months to October 2024, and that the share of passengers on routes with four or more airlines rose from 24 percent in 2019 to 58 percent in 2023 (Competition Bureau Canada 2025).

For an economist the shares give a quick check on concentration: the Herfindahl index of the four named carriers alone is $34.1^2 + 30.1^2 + 9.8^2 + 9.0^2 \approx 2{,}246$ on the 0 to 10,000 scale, before adding the 17 percent held by others. The study's ten recommendations include removing the Minister of Transport's power to override Competition Act review of airline mergers and joint ventures, publishing traffic, quality and slot data, and raising foreign ownership limits (Competition Bureau Canada 2025). No formal government response or implementing legislation was found as of October 2026; in March 2026 the limits of 49 percent total and 25 percent for a single foreign investor, and the ban on foreign carriers flying domestic routes, were unchanged (Michell 2026).

Competitors WestJet's ownership changed in October 2025, when Delta (12.7 percent), Korean Air (10 percent) and Air France-KLM (2.3 percent) bought a combined 25 percent stake, with Onex keeping control (WestJet 2025). In 2026 WestJet retrenched: its Copenhagen, Glasgow and Madrid flying ended in September and Calgary–Guadalajara on 18 October, with the airline citing fuel costs and capacity alignment, and Cuba was suspended after fuel shortages on the island (Aviation A2Z 2026). Porter, twenty years old in October 2026, flies 56 Embraer E195-E2s and 29 Dash 8-400s to 33 destinations on 131 routes; its five-year plan prioritizes denser service in eastern Canada, including Montreal's Metropolitan Airport, more US flying and sun routes, while testing western point-to-point flights (Travelweek 2026b). Air Transat reported a net loss of CAD 106.6 million for the quarter to 31 July 2026 on revenue of CAD 792.7 million, with capacity up 6 percent and fuel costs up about CAD 105 million (Transat A.T. 2026). Lynx Air and Canada Jetlines left the market in 2024 (Competition Bureau Canada 2025). In June 2026 the federal government created a repayable liquidity facility offering eligible airlines up to CAD 150 million of support, linked to the rise in jet fuel prices and their fuel use (Department of Finance Canada 2026); Porter and Transat were approved for CAD 150 million each and Flair for CAD 76 million, and WestJet criticized the loans as market-distorting (Daily Hive 2026).

Passenger protection The Air Passenger Protection Regulations, in force since 2019, tie compensation to whether a disruption was within the airline's control, within its control but required for safety, or outside its control (Government of Canada 2019). Amendments pre-published in the Canada Gazette on 21 December 2024 would replace the three categories with a closed list of exceptional circumstances, so that compensation is owed unless one applies; they would also require assistance from a two-hour delay, refunds within 15 days rather than 30, free adjacent seating for children under 14, and raise the maximum corporate penalty from CAD 25,000 to CAD 250,000 (Canadian Transportation Agency 2024). In September 2026 the Canadian Transportation Agency still described the amendments as proposed and reported a backlog of more than 95,000 complaints (Canadian Transportation Agency 2026b). Bill C-31, introduced in May 2026, would move regulation-making to the Minister of Transport, have designated third parties resolve complaints and raise penalties to as much as CAD 1 million (McCarthy Tétrault 2026); it passed second reading on 3 June 2026 (openparliament.ca 2026), and in late October 2026 it was still at committee stage in the House of Commons (Parliament of Canada 2026). For a planner these rules convert reliability into money: the compensation and care exposure of a tight connection or rotation belongs in the business case (Chapters 16 and 19).

Airports In 2024 the Canadian Airports Council counted only three fully slot-coordinated (IATA Level 3) airports in Canada, Toronto Pearson, Vancouver and Billy Bishop Toronto City, and cited Montreal-Trudeau, Calgary and Quebec City as examples of Level 2 airports, where schedules are adjusted by agreement with a facilitator (Canadian Airports Council 2024). From the summer 2027 season a series of slots at a coordinated airport needs at least seven slots in a northern summer season instead of five (IATA, ACI, and WWACG 2025). Montreal-Trudeau operates around the clock with restrictions (Network Planning Handbook 2026): aircraft certified to the Chapter 3 or 4 noise standards and heavier than 45,000 kilograms may not depart between 00:00 and 07:00 or arrive between 01:00 and 07:00, and louder Chapter 2 aircraft face earlier cut-offs (NAV CANADA 2023), with exemptions granted by the airport operator (Boeing 2011). In December 2025 the airport operator proposed banning all takeoffs and landings between 01:00 and 06:00 (CBC News 2025); whether and when that will take effect was not known at the time of writing. The handbook's advice holds: never say casually that YUL closes at night; check the restriction

for the aircraft category and the permission that applies (Network Planning Handbook 2026).

23.6 Air Canada in the autumn of 2026

What follows is a public snapshot for orientation, not an internal briefing (Network Planning Handbook 2026). Air Canada describes itself as the largest provider of scheduled passenger services in the Canadian domestic, transborder and international markets, with regional flying bought from Jazz under a capacity purchase agreement and from PAL Airlines, and with Air Canada Rouge as its leisure carrier (Air Canada 2026e).

Second-quarter results Table 23.3 reproduces the main figures of the second-quarter 2026 MD&A. Chapter 2 uses the same document to work through the identities that link them; here the point is the story they tell.

Table 23.3 Air Canada, second quarter of 2026 against 2025, as printed in the MD&A, with changes as printed (Air Canada 2026e). Statistics include flying bought from Jazz. The printed RPM level does not reproduce the printed ratios exactly: load factor, RPM growth, the load factor change, yield and yield growth are all consistent with RPMs between about 23,561 and 23,577 million, not 23,597.

Measure	Q2 2026	Q2 2025	Change
Operating revenues (CAD millions)	6,266	5,632	+11%
Passenger revenues (CAD millions)	5,576	5,031	+10.9%
Operating income (CAD millions)	−215	418	−633
Adjusted EBITDA margin	11.5%	16.1%	−4.6 pp
Available seat miles (millions)	26,942	26,860	+0.3%
Revenue passenger miles (millions)	23,597	22,796	+3.4%
Passenger load factor	87.5%	84.9%	+2.6 pp
Yield (cents per RPM)	23.7	22.1	+7.2%
PRASM (cents)	20.7	18.7	+10.5%
CASM (cents)	24.1	19.4	+23.9%
Adjusted CASM (cents)	15.5	14.4	+7.4%
Fuel cost per litre (cents)	132.7	88.0	+50.8%
Aircraft in operating fleet at quarter end	357	364	−1.9%

Passenger revenue rose 11 percent on nearly flat capacity, because the load factor rose 2.6 points and yield 7.2 percent; but operating expenses rose 24 percent, driven by fuel (up 49 percent to CAD 1,713 million as the price per litre rose by about half, despite a CAD 205 million hedging gain), higher labour costs and CAD 388 million

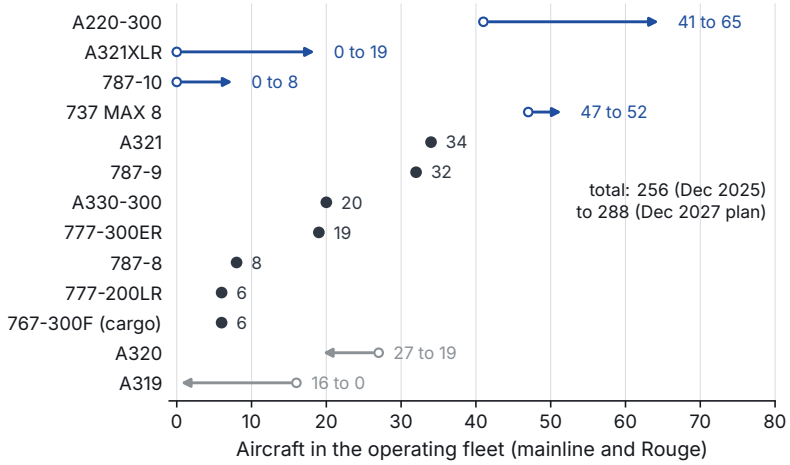


Figure 23.4 Air Canada mainline and Rouge operating fleet by aircraft type, actual on 31 December 2025 against the plan for 31 December 2027, as shown in section 6 of the second-quarter 2026 MD&A (Air Canada 2026e). Brands are combined; the Air Canada Express fleet flown by regional partners (95 aircraft in mid-2026) is planned unchanged and not shown. Planned transfers are subject to approvals.

of pension, labour and legal charges, and the quarter ended with an operating loss. Capacity grew 0.3 percent, below guidance, because weather cut completion late in the quarter. Premium cabins brought about 30 percent of passenger revenue, up 11 percent, and sixth-freedom revenue, from passengers connecting between two foreign countries over Canadian hubs, grew 9 percent, mostly from the United States to Europe (Air Canada 2026e). The regional pattern is a lesson in capacity and price: transborder capacity fell 5.7 percent and unit revenue rose 18.5 percent, while Pacific capacity grew 7.9 percent and unit revenue only 3.3 percent; correlation is not causation, since fuel, fares and a weak Canada–US travel market all moved at once. After the quarter, Air Canada agreed to sell 25 percent of Aeroplan, its loyalty program, for CAD 2.5 billion while keeping control (Air Canada 2026e), and set its 2026 guidance at CAD 2.9 to 3.2 billion of adjusted EBITDA, down from CAD 3.35 to 3.75 billion (Air Canada 2026d).

Fleet and orders The fleet plan in Figure 23.4 shows where growth will come from. Mainline and Rouge together go from 256 aircraft at the end of 2025 to 262 in June 2026, 267 at the end of 2026 and 288 at the end of 2027: widebodies from 91 to 99 with the arrival of the 787-10, narrowbodies from 165 to 189. The narrowbody fleet is also

reshuffled: all 737 MAX 8s move to Rouge during 2026, the Airbus A320 family moves to mainline, and the A319s leave.

The order book behind the plan is set out in the MD&A: 30 A321XLRs (15 leased and 15 bought, with rights to 10 more for 2030 to 2032), the first two delivered in the second quarter of 2026 and the last due in 2029; 65 firm A220-300s, of which 46 had been delivered by 30 June 2026, plus 10 options; 14 787-10s, ten by 2028 and four by 2030, plus 12 options; eight A350-1000s from 2030, plus rights to eight more; ten options on the 737 MAX; and a conditional order for 30 Heart Aerospace ES-30 electric-hybrid aircraft not expected before 2031 (Air Canada 2026e). At the published layouts of 137 seats for the A220-300 (Air Canada n.d.a) and 182 for the A321XLR (Air Canada 2026g), the 24 additional A220s and 19 XLRs planned by the end of 2027 add 3,288 and 3,458 seats.

These numbers are plans, not capacity. A planned fleet count or an announced order is not the number of aircraft delivered, the number operationally available, or the number assignable to your desk in a given month (Network Planning Handbook 2026); before building flying on a planned aircraft, ask which constraint binds: tails, engines, maintenance cover, cabin configuration or trained crews.

Network moves The A321XLR is reshaping the network. Air Canada received its first on 24 April 2026 (Air Canada 2026a); it entered service on 9 June on a Montreal–Toronto flight, with its first transatlantic service, Montreal–Toulouse, scheduled for 15 June (Air Canada 2026b). For summer 2027 the airline announced Toronto–Oslo (four weekly, subject to approval) and Toronto–Shannon (three weekly) on the XLR, Toronto–Nice on the A330, and the return of a daytime Toronto–London Heathrow flight on the XLR (Air Canada 2026f), together with Montreal–Basel-Mulhouse (four weekly on the XLR from 16 June) and Montreal–Dubrovnik on the 787, in what it called its largest intercontinental expansion: more than 125 international routes and up to 169,000 weekly seats (PAX News 2026).

On 14 October 2026 it announced Montreal–Seville, Air Canada's fifth Spanish destination and the only nonstop between North America and Seville: three weekly flights from 8 September 2027, four from 1 October, on the 182-seat A321XLR (14 business and 168 economy seats), departing Montreal at 20:30 and arriving at 09:30 the next morning, returning at 12:45 to arrive at 14:40, and operating in two periods, until early January and from mid-February to early June 2028 (Air Canada 2026g). Laurentian's YUL–SVQ case mirrors this public schedule; everything it earns and spends is invented.

Two days later Air Canada and United announced new transborder routes for summer 2027 (Toronto–Louisville, Montreal–Indianapolis and Edmonton–Newark among them) under their transborder joint business agreement, in place since 2022 and credited by the airlines with more than 1,200 connecting itineraries and 20 new routes; they also agreed to renew and expand the agreement, subject to regulatory approval, and applied to Transport Canada (Air Canada 2026h). The expansion would bring in ten city pairs excluded until now, including Montreal–Chicago and Montreal–Washington, and Transport Canada opened a public consultation to weigh public-interest and competition effects (Travelweek 2026a). The excluded markets go back to a 2012 consent agreement with the Commissioner of Competition (Lexology n.d.), and in the United States the carriers need the Department of Transportation to extend their antitrust immunity beyond July 2027 (PYMNTS 2026). The case connects directly to the Competition Bureau's first recommendation, on the minister's power to approve such agreements over a competition review.

Leadership is also changing: Air Canada announced in July 2026 that Michael Rousseau would retire on 31 August, with the executive committee reporting to the board until Anko van der Werff, then chief executive of SAS, takes over by the end of January 2027 (Air Canada 2026c). Figure 23.5 collects the dated milestones of this chapter.

23.7 What changed since 2006, and what did not

The courses this book draws on were taught when distribution ran through GDS terminals, the Boeing 757 flew the thinnest transatlantic routes and nobody priced carbon into a route. Table 23.4 sets the two dates side by side.

2026	2027	2028	2029	2030
Regulation and market access				
Jan EU ETS: allowances for aviation fully auctioned **17 Jul** Commission proposes the ETS for departures to points within 5,000 km, from 2029; not adopted **Oct** Air Canada and United file to renew and widen their transborder JBA	**1 Jan** CORSIA second phase; EU non-CO_2 reporting extends to all flights to and from the EEA **28 Mar** Summer slot series need 7 slots (WASG) **Jul** US antitrust immunity for the JBA ends unless extended	**2028** British Columbia: 1% renewable jet fuel	**2029** Proposed ETS extension would start, if adopted	**2030** ReFuelEU: 6% SAF, at least 1.2% synthetic (two-year average)
Air Canada fleet and network				
24 Apr First A321XLR delivered; in service from 9 Jun **Sep** Summer 2027 expansion announced **Dec (plan)** 9 A321XLR and first 2 787-10s in the fleet	**By end Jan** New chief executive in post **Summer** What the airline calls its largest intercontinental expansion **8 Sep** Montreal–Seville launches on the A321XLR **Dec (plan)** 288 mainline and Rouge aircraft, 19 A321XLR, 65 A220	**2028** 10 of 14 787-10s due by 2028	**2029** Last of 30 A321XLRs	**2030** First A350-1000s; last four 787-10s

Figure 23.5 Milestones of 2026 to 2030 for regulation and market access (top) and for Air Canada's fleet and network (bottom). Year columns are not to scale. Items after October 2026 are plans, proposals or scheduled dates. Sources as cited in this chapter.

Table 23.4 The industry of the teaching material and of 2026. Sources as cited in this chapter and in Chapter 24.

Topic	2006 teaching material	2026
Distribution	GDS; fares filed in advance; at most 26 booking classes	NDC at about 22% of US agency transactions; orders being introduced
Price setting	fixed fare ladders controlled by class	continuous and per-request pricing on some markets
Forecasting	time series and pickup by booking class	machine-learning tools, evidence mostly from vendors
Planning tools	O&D profitability with share and spill models	the same chain, faster, in cloud products
Long-haul fleet	widebodies; 757 on thin routes	A321XLR opens thin transatlantic markets
Aircraft supply	taken for granted	engine groundings and a backlog of 18,100 aircraft
Carbon	absent from route economics	CORSIA (nearly all international flights from 2027), SAF mandates, EU ETS in Europe
US O&D data	DB1B, 10% quarterly sample	DB1C, 40% monthly sample from July 2025
Canadian market	Air Canada and WestJet; Porter enters	Porter jets, Flair, WestJet partly foreign-owned, passenger protection rules

The list of what did not change is shorter and more important. Seats still perish at departure. Demand still arrives by O&D market while capacity is supplied by flight leg, and the spill arithmetic of Chapter 9 still governs what an extra seat is worth. A route's value is still the change it makes to the whole network's contribution, not its allocated profit, and a bid price is still the opportunity cost of a seat. Fares are still endogenous, logit shares still impose proportional substitution, and every forecast is still a model of censored data. A continuous price solves the marginal condition that a fare ladder approximates, a long-range narrowbody is judged by spill-adjusted contribution like any other gauge, and a carbon charge is one more avoidable cost with a scope.

Read every 2026 development as a change to an input, a constraint or a tool, and then apply the unchanged economics: incremental network contribution, opportunity cost and the value of the marginal seat or aircraft.

23.8 Summary

- Pricing is moving from at most 26 filed price points toward continuous, per-request prices sold through NDC; adoption is real but gradual, at about 22 percent of US agency transactions in 2026.
- The contribution lost by rounding a price to a ladder is second order in the step, about $(s\,\eta^*)^2/24$; fine ladders lose tenths of a percent, so larger claimed gains must come from estimation, segmentation or competition.
- Vendor uplift figures are self-reported; the peer-reviewed evidence on continuous pricing is simulation, with gains of up to 2 percent for a first mover.
- Planning systems in 2026 still chain demand, share, spill and recapture, and cost models; what changed is speed and data.
- Engine groundings and delivery backlogs make aircraft time scarce; a missing aircraft costs the marginal line of flying, and scarcity raises the hurdle for new routes.
- The A321XLR opens markets too thin for a widebody: in Laurentian's case only the narrowbody is profitable between about 129 and 176 passengers per departure.

- Carbon costs are route- and direction-specific: CORSIA on international routes (nearly all of them from 2027), SAF mandates on fuel bought in Europe, the EU ETS within Europe.
- Canada's market remains concentrated; no government response to the Competition Bureau's study was found, passenger protection reforms are pending, and the Air Canada and United transborder agreement is under review.
- In 2026 Air Canada pairs strong unit revenue with a fuel and labour cost shock; its plan grows the mainline and Rouge fleet to 288 aircraft by the end of 2027, with new XLR routes such as Montreal–Seville.

23.9 Further reading

The second-quarter MD&A (Air Canada 2026e) and the Competition Bureau's study (Competition Bureau Canada 2025) are the two library documents to read in full; the handbook (Network Planning Handbook 2026) places both in a planner's workflow. On pricing technology, start with Wittman and Belobaba (2019) for definitions, then the PODS studies of continuous pricing (Szymański, Belobaba, and Papen 2021) and of its segmented form (Long and Belobaba 2023); Fiig and coauthors (2010) give the fare adjustment theory behind class-based pricing. Belobaba, Odoni and Barnhart (2015) cover distribution, fleet planning and regulation at textbook depth. For carbon policy, read the IATA CORSIA fact sheet (IATA 2026b) and the ReFuelEU regulation itself (European Parliament and Council of the European Union 2023).

CHAPTER

24

The Analyst's Toolkit

IN THIS CHAPTER

- The four objects of airline data, and the unit discipline that keeps them apart.
- The data sources an analyst meets, what each measures and where each misleads, including the 2025 break in the US origin-and-destination survey.
- A data contract, a reproducible workflow in SQL, Python and spreadsheets, and the audit checks that catch the common errors.
- Clocks: UTC and local time, daylight saving, and the weeks when Montreal and Seville are five hours apart.
- The schedule card, the sensitivity matrix, decision thresholds and the one-page memo.
- A reading plan, and the questions that establish definitions in the first weeks.

DRAWS ON Handbook (data catalogue D1 to D23, analytical workflow, templates, reading room and first ninety days); IATA, ACI and WWACG, Worldwide Airport Slot Guidelines, edition 4; Air Canada, Second quarter 2026 MD&A; Bureau of Transportation Statistics; public sources dated up to 27 October 2026

Most errors in planning analysis are not modelling errors. They are passengers counted twice because a join multiplied rows, seats counted three times because codeshare flights were not removed, a bank that vanishes for a week because a time zone offset was typed in by hand, a market forecast with false precision because its sample was small, and a recommendation that cannot be acted on because it does not say what would change it. The models of the earlier chapters are only as good as the tables that feed them and the memo that carries their answer.

This chapter is the working toolkit. It follows the analytical workflow of the handbook (Network Planning Handbook 2026) step by step and expands each step with a derivation or a worked example on Laurentian. It recommends a method, not a software stack: which

systems a team uses, and which one is authoritative for each number, has to be learned inside the team (Chapter 23). The reader is assumed to know basic SQL and Python; the code is deliberately plain.

24.1 Four objects and unit discipline

Airline data describe four different objects, and the first discipline of the job is to keep them apart (Network Planning Handbook 2026). An **aircraft movement** is one takeoff or one landing at an airport. A **flight leg** is one flight from takeoff to landing, such as YUL–CDG on a given date. A **passenger itinerary** is a journey from an origin to a destination, possibly over several legs. A **ticket** is a commercial document that can cover several journeys, typically an outbound and a return. Someone flying from Boston to Paris through Montreal uses a seat on each of two legs, yet makes only one Boston–Paris trip in one direction, and the same ticket may carry the trip home. Chapter 2 builds the measures on these objects (seat miles, passenger miles, load factor, yield and unit revenue); this section is about tables.

> **DEFINITION 24.1 Grain**
>
> The **grain** of a table is the object one row represents, together with the keys that identify it: one row per flight leg and date, per coupon (one passenger on one leg of a ticket), per directional journey, per ticket, or per market and month. Every aggregate must state the grain it sums over, and every join must state the grain of its result.

WORKED EXAMPLE 24.1 One travel day, counted four ways

Laurentian's coupon table for 8 September 2027 holds five tickets, all illustrative. Ticket T1 is Boston to Paris via Montreal with a return on 20 September; T2 is a one-way Montreal to Paris; T3 and T4 are two passengers travelling together from Quebec City to Seville via Montreal; T5 is Montreal to Seville with a return on 15 September. One row is one coupon: 11 rows in all, 8 of them flown on 8 September. Each row also carries a `journey_id` that identifies one directional journey on a ticket, the

outbound or the return. Grouping the same rows by different keys gives different answers:

```
SELECT leg_orig, leg_dest, COUNT(*) AS leg_pax
FROM coupons
WHERE flight_date = '2027-09-08'
GROUP BY leg_orig, leg_dest;

SELECT od_orig, od_dest,
       COUNT(DISTINCT journey_id) AS od_pax
FROM coupons
WHERE flight_date = '2027-09-08'
GROUP BY od_orig, od_dest;
```

The first query returns leg passengers: 3 on YUL–SVQ, 2 on YUL–CDG, 2 on YQB–YUL and 1 on BOS–YUL, eight in total. The second returns directional journeys: 2 from Quebec City to Seville and 1 each from Boston to Paris, Montreal to Paris and Montreal to Seville, five in total. The day also has 5 tickets and 4 flights, which make 8 aircraft movements (four takeoffs and four landings); the Seville flight lands on 9 September, local time, so even a count of movements by date needs a stated rule.

Revenue shows why grain matters. The five tickets are worth CAD 4,343 together. Joining a table of ticket values to the coupon table and summing gives CAD 10,533, because each ticket's value is repeated on every coupon row: four times for T1, twice for T3, T4 and T5.

Each question has its grain: leg passengers for capacity and load factor, journeys for market size and share, tickets for revenue, movements for airport charges. Choose the grain first, then write the query.

Units follow the same discipline. Passengers per day each way (PDEW) divide a two-way annual total by both the number of days and the two directions: 73,000 passengers in a year across both directions of a symmetric market are 100 PDEW (Network Planning Handbook 2026). Air Canada reports capacity in seat miles (Air Canada 2026e); converting to seat kilometres multiplies by 1.609344, and a ratio must never mix the two.

24.2 The data an analyst will meet

No single source covers an airline market. The supply side is measured well, because schedules are published; the demand side is es-

timated, because no one observes every journey. Table 24.1 summarizes the sources described in the handbook's data catalogue, with the trap that most often catches a new analyst in each.

Table 24.1 Data sources a network analyst meets, adapted from the handbook's catalogue (Network Planning Handbook 2026), with Cirium (2026a) and Bureau of Transportation Statistics (2026a).

Source	What it measures	Grain	Main traps
Schedules (OAG, Cirium)	offered flights, seats, times, aircraft, operating and marketing carriers	flight leg and date, per snapshot	future schedules change; codeshare duplicates; aircraft type does not fix the seat layout
Booking-based market data (OAG, Cirium)	estimated O&D and segment passengers and fares	market, carrier, month	estimates, not a census; uneven coverage of direct, NDC and low-cost sales; later restatements
BTS T-100 Segment and Market	operated flights, seats and passengers on flights touching the United States	segment, or on-flight market	a market record is the trip on one flight number, not the full journey
BTS DB1B, then DB1C	sampled tickets: itineraries and fares	ticket and coupon	10% quarterly sample replaced by 40% monthly in July 2025; international microdata restricted
Statistics Canada	airport activity, movements, aviation indicators	airport or carrier, month	the domestic O&D survey (2702) is inactive
Airport reports	passengers and movements	airport, month	enplaned, screened, connecting and transit passengers differ
Internal systems	bookings, forecasts, fares, cost definitions	varies	learn which system owns each number

Schedules Published schedules measure supply: departures, seats, frequency, timing and the itineraries that can be built from them (Network Planning Handbook 2026). Schedules for future dates are plans that move, so each extraction must keep the date it was taken: the change between two extractions for the same travel week is not the same quantity as the change from last year to this year. A flight sold under several airline codes appears once per code, so operated flights must be counted once, by operating carrier, and knowing the aircraft type does not tell you the seat layout.

Booking-based market estimates Market data products estimate demand by O&D. According to OAG, its traffic database is built mainly from GDS booking records (the Marketing Information Data Tapes, MIDT) together with other inputs, in adjusted and unadjusted versions (Network Planning Handbook 2026; OAG n.d.). Cirium's traffic and fares product covers more than 600 airlines from more than 24 sources, including multi-GDS booking data that it says include NDC bookings, government statistics and the US ticket surveys; it fills gaps by regression and scaling to external control totals, publishes about 55 days after the end of a month and restates at 115 days (Cirium 2026a). Adjusted totals are estimates, not a census, and their precision should be presented accordingly. For each product, find out what happens to bookings made directly with airlines, to low-cost carriers, to cancelled and group bookings, to NDC sales and to countries where sales are poorly covered (Network Planning Handbook 2026).

IN PRACTICE **Market data in 2026**

A growing share of agency sales uses NDC rather than traditional GDS connections: 21.6 percent of US agency transactions settled through ARC in May 2026, against 21.4 percent a year earlier (Airlines Reporting Corporation 2026b); and airlines sell a further, unpublished share directly. Data built on traditional GDS bookings therefore risk covering less of the market than they once did, which is why vendors blend booking data with settlement data, government statistics and statistical models of the gaps (Cirium 2026a). The practical consequence for an analyst is to treat market size and market share as estimates with a coverage problem that differs by carrier and market, and to ask the provider, market by market, what share of the total is observed and what share is modelled.

US government data The US Bureau of Transportation Statistics (BTS) publishes two families of airline data (Network Planning Handbook 2026). The **T-100** tables report operated traffic on flights touching the United States: the Segment table gives nonstop activity and capacity, and the Market table gives passengers between the points where they boarded and left a flight number, which is not necessarily their full journey. The All Carriers version includes foreign airlines but leaves out services with both ends outside the United States, so it is useful as a benchmark for flights between Canada and the United States and useless for sizing a Montreal–Europe market.

The **origin and destination survey** samples tickets. Until June 2025 it was the quarterly 10 percent sample known as DB1B. From travel on 1 July 2025, US carriers report a 40 percent sample monthly, and BTS released the first data from the new program, DB1C, for July to September 2025 on 17 February 2026 (Bureau of Transportation Statistics 2026a, 2026b). The new series differs from the old in how often it is reported, how tickets are sampled, what it covers and how records are laid out. The publicly released tables cover domestic travel and internationally collected records are restricted, so the public DB1C cannot stand in for data on Canada's international markets (Network Planning Handbook 2026).

Canadian and airport data Statistics Canada's aviation data hub brings together airport activity, aircraft movements and monthly aviation indicators, which are useful for aggregate traffic and seasonality; its domestic passenger origin and destination survey (number 2702) is listed as inactive, with releases suspended for confidentiality reasons (Network Planning Handbook 2026). Airport operators

publish their own traffic, sometimes split into domestic, transborder and international; definitions of arriving, departing, screened, connecting and transit passengers differ, so airport totals are a cross-check, not a route forecast (Network Planning Handbook 2026).

What a sample can say A sample of tickets measures a market with error, and the error matters most for the thin markets where new routes are proposed.

RESULT 24.1 Precision of an expanded sample count

Suppose a market has N passengers in a period and each passenger is included in the sample independently with probability p. If X passengers are sampled, the expanded count $\hat{N} = X/p$ is unbiased for N, with variance and coefficient of variation

$$\operatorname{Var}(\hat{N}) = \frac{N(1-p)}{p}, \qquad \frac{\sqrt{\operatorname{Var}(\hat{N})}}{N} = \sqrt{\frac{1-p}{pN}}.$$

Proof. Under the assumption, $X \sim \operatorname{Bin}(N, p)$, so $\mathbb{E}[X] = Np$ and $\operatorname{Var}(X) = Np(1-p)$. Dividing by p gives $\mathbb{E}[\hat{N}] = N$ and $\operatorname{Var}(\hat{N}) = Np(1-p)/p^2 = N(1-p)/p$. Dividing the standard deviation by N gives the coefficient of variation. ∎

Real surveys sample tickets, not passengers, and a ticket can carry several passengers and both directions of a trip; that clustering makes the effective sample smaller and the variance larger than the formula says (Lohr 2021). The formula is a lower bound on the error, which is the useful direction for a sanity check.

WORKED EXAMPLE 24.2 Thin markets and a broken series

Take a US domestic market with 12,000 passengers a month in both directions together. Under the old 10 percent sample about 1,200 would be sampled, and by Result 24.1 the expanded total has a standard error of $\sqrt{12{,}000 \times 0.9/0.1} = 329$ passengers, a coefficient of variation of 2.7 percent. Under the 40 percent sample, 4,800 are sampled and the coefficient of variation is 1.1 percent. For a thin market of 600 passengers a month the figures are 12.2 and 5.0 percent. Quadru-

pling the sampling rate cuts the standard error by a factor of $\sqrt{(0.9/0.1)/(0.6/0.4)} = \sqrt{6} = 2.45$, not four.

Now the splice. Suppose the new monthly sample records 4,800 passengers. Expanded correctly by $1/0.4 = 2.5$, that is 12,000 passengers, or $12{,}000/(2 \times 30) = 200$ PDEW in a 30-day month. An analyst who keeps multiplying by ten, as for the old 10 percent sample, reports 48,000 passengers and an apparent fourfold jump in demand in July 2025. A quarter of the old series with 3,510 sampled passengers expands to 35,100 passengers, or 193 PDEW over 91 days, in line with the new series once each is expanded by its own factor.

The planner's lesson is to record the expansion factor, and the record definitions it comes from, in the data contract, to report intervals rather than points for thin markets, and to remember that a big market estimated from a small sample does not become precise by passing through a sophisticated model (Network Planning Handbook 2026).

PITFALL

Splicing a broken series. Do not join DB1B and DB1C, or any series whose sample design changed, by multiplying every new row by the old factor. Check in the version-specific record definitions whether passenger counts are raw sample counts or already expanded, and mark the break in every chart that crosses it (Network Planning Handbook 2026).

24.3 The data contract

Every table that enters an analysis should arrive with a short written description of what it is. The handbook calls this a data contract (Network Planning Handbook 2026).

DEFINITION 24.2 Data contract

A **data contract** records, for one table, its grain; its source; its extraction (snapshot) date; the travel period it covers; its currency; whether fares are gross or net, and of what; whether carriers are marketing or operating carriers; how airports are grouped into city markets; its revision status; and its coverage. It states explicitly whether rows are tickets, passengers, coupons, legs or flights.

Table 24.2 fills in a contract for two extracts an analyst might pull for the Seville case. The entries are illustrative; the fields are not.

Table 24.2 An illustrative data contract for two extracts. The fields follow the handbook (Network Planning Handbook 2026).

Field	Schedule extract	Market estimate extract
Grain	one operated flight leg per date	one directional O&D market per carrier and month
Source	schedule data provider	traffic and fares product
Snapshot date	26 October 2026	release of 15 October 2026
Travel period	northern winter 2026–27	January 2024 to August 2026
Currency	not applicable	CAD, at the provider's monthly rate
Fare treatment	not applicable	average fare; taxes and commissions to confirm
Carriers	operating carrier only, codeshares removed	as defined by the provider
Geography	airports	city markets; is Montreal Metropolitan Airport grouped with YUL?
Revision status	final for past dates, plan for future dates	first release, restated later
Coverage	all carriers in the provider's file	booking data plus modelled gaps

The contract is the first stage of a lineage that runs from sources to the memo (Figure 24.1). Two habits keep that lineage honest. The first is an **assumption register**: one line per assumption, with its source, owner, date, uncertainty and the decision it affects (Network Planning Handbook 2026). The second is vintage discipline: keep the snapshot date of every input, so that a backtest uses only what

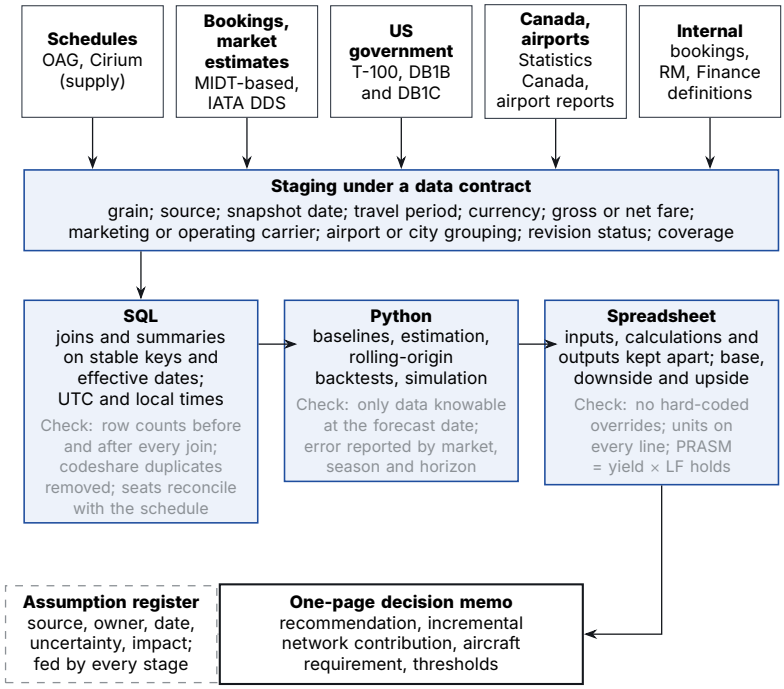


Figure 24.1 The lineage of a route analysis. Each source enters through a data contract; SQL, Python and a spreadsheet each do the work they do best, each with its own audit check; the result reaches a one-page memo, and an assumption register records every assumption along the way.

was knowable on the date of the forecast and not schedules or economic data revised later (Chapter 8).

24.4 A reproducible workflow

The handbook's workflow has six steps: frame the decision and its counterfactual; write a data contract; join and summarize in SQL; estimate and test sensitivity in Python; review scenarios transparently in a spreadsheet; and finish with a short memo (Network Planning Handbook 2026). The first step is the most often skipped. The counterfactual is the schedule the airline would actually fly without the change, with the displaced aircraft put to its next-best use; without it, no number that follows has a meaning (Chapter 12).

SQL for joins SQL is the right tool for joining schedules, bookings and reference tables, because a query is explicit and can be rerun when the data are refreshed. Hold airports, carriers, cabin configu-

rations, dates and currencies in dimension tables, join through stable keys that are valid on the right dates, and keep every local time next to its time zone and its UTC equivalent (Network Planning Handbook 2026). Two queries cover the most common traps. The first counts operated seats only once, whatever the number of codes a flight is sold under; the second attaches the seat configuration that was valid on each date:

```
SELECT SUM(seats) AS seats_operated
FROM schedule
WHERE flight_date = '2027-09-08'
  AND mkt_carrier = op_carrier
  AND mkt_flight = op_flight;

SELECT s.flight_date, s.op_flight,
       c.seats_j, c.seats_y
FROM schedule AS s
JOIN cabin_config AS c
  ON c.aircraft = s.aircraft
 AND s.flight_date BETWEEN c.valid_from
                       AND c.valid_to
WHERE s.mkt_carrier = s.op_carrier;
```

After every join, compare row counts with the grain you expected: a many-to-many join multiplies traffic or revenue without any error message.

WORKED EXAMPLE 24.3 Seats that exist three times

On 8 September 2027 Laurentian's YUL–CDG flight (298 seats) is sold under its own code and under two partner codes, and its YUL–SVQ flight (182 seats) under its own code and one partner code. A schedule extract holds one row per marketed flight, five rows in all. Summing seats over the rows gives 1,258; the first query above returns 480, the seats that will actually fly. The difference, a factor of 2.6, would pass unnoticed into seat miles, load factors and market shares. A check against the published operated schedule catches it in seconds.

Python for estimation and simulation Python (or R) is the right tool for estimation, backtests and simulation. Start from a simple seasonal or analogue baseline, backtest on rolling windows so that each forecast sees only earlier data, and model demand by travel date separately from the pickup of bookings (Network Planning Handbook 2026). The seasonal naive forecast of Chapter 8 and a rolling-origin evaluation take a few lines:

```
import numpy as np

def seasonal_naive(y, h, m=52):
    """Repeat the last observed season."""
    return np.array([y[len(y) - m + i % m]
                     for i in range(h)])

errors = []
for t in range(104, len(y) - 4):
    fc = seasonal_naive(y[:t], h=4)
    errors.append(np.abs(fc - y[t:t + 4]))
mae_by_horizon = np.mean(errors, axis=0)
```

The loop moves the forecast origin t forward one week at a time, so every forecast uses only the weeks before it, and the result is the mean absolute error at each horizon from one to four weeks.

Formulas used repeatedly belong in small, tested functions. The expected spill of Chapter 9 is one:

```
from scipy.stats import norm

def expected_spill(mu, sigma, C):
    z = (C - mu) / sigma
    return sigma * (norm.pdf(z) - z * norm.sf(z))

expected_spill(160, 40, 180)  # 7.91
```

The call reproduces the 7.91 passengers of expected spill in the handbook's Exercise B (Chapter 9).

Spreadsheets for scenarios A spreadsheet is the right tool for the scenario review that colleagues will read and challenge. Put inputs, calculations and outputs on separate sheets, label every line with its unit and forbid quiet hard-coded overrides; show base, downside and upside versions of demand, yield, fuel, exchange rates and reliability; and take the cost definitions from Finance (Network Planning Handbook 2026). Spill can be computed in a cell with `NORM.S.DIST`, as in Chapter 9, and a two-way data table produces the sensitivity matrix of Section 24.6. When a number in the spreadsheet disagrees with the Python result, find out why before showing either.

24.5 Clocks, seasons and the schedule card

Time is the most error-prone dimension of schedule data. The rules are few (Network Planning Handbook 2026): compute in UTC and keep local time for the commercial reading; carry time zone identifiers and date offsets; never use one UTC offset for the whole year;

expect transatlantic flights to cross midnight and Canada and Europe to change their clocks on different dates; and give every weekly pattern an explicit rule for which local date a flight belongs to. The industry exchanges schedules in the format of IATA's Standard Schedules Information Manual (SSIM), whose edition 36 is effective from 1 March 2026 to 28 February 2027 (Network Planning Handbook 2026).

The dates differ by design. Quebec, like most of Canada, puts its clocks forward on the second Sunday in March and back on the first Sunday in November, at 02:00 local time (National Research Council Canada 2025; Québec 2007). The European Union's summer time runs from 01:00 UTC on the last Sunday in March to 01:00 UTC on the last Sunday in October (European Parliament and Council of the European Union 2001). IATA's scheduling seasons begin on the last Sunday in March and the last Sunday in October (IATA, ACI, and WWACG 2025), so they follow Europe, not North America. In software, use the identifiers of the IANA time zone database, which Python's `zoneinfo` module reads: `America/Toronto` covers Montreal and `Europe/Madrid` covers Seville.

RESULT 24.2 Clock changes move block time between directions

Let Seville be Δ hours ahead of Montreal on the dates concerned, and let a schedule keep fixed local clock times: eastbound departure d_E from Montreal and arrival a_E in Seville the next day, westbound departure d_W from Seville and arrival a_W in Montreal the same day, all in hours. The block times the schedule implies are

$$b_E = (a_E + 24 - d_E) - \Delta, \qquad b_W = (a_W - d_W) + \Delta.$$

A one-hour fall in Δ lengthens the eastbound block by an hour and shortens the westbound block by an hour; their sum does not change.

Proof. Let o_Y and o_S be the UTC offsets of Montreal and Seville, so $\Delta = o_S - o_Y$. The eastbound flight leaves at $d_E - o_Y$ UTC and arrives at $a_E + 24 - o_S$ UTC, so $b_E = (a_E + 24 - d_E) - (o_S - o_Y)$. The westbound flight leaves at $d_W - o_S$ UTC and arrives at $a_W - o_Y$ UTC, so $b_W = (a_W - d_W) + (o_S - o_Y)$. Adding the two, Δ cancels. ∎

WORKED EXAMPLE 24.4 Montreal and Seville, five hours apart

Laurentian's YUL–SVQ schedule mirrors the published Air Canada pattern: depart Montreal 20:30, arrive Seville 09:30 the next morning; depart Seville 12:45, arrive Montreal 14:40. When Seville is six hours ahead, Result 24.2 gives $b_E = (33.5 - 20.5) - 6 = 7.0$ hours and $b_W = (14.67 - 12.75) + 6 = 7.92$ hours, 7:00 and 7:55, the blocks of the running case.

In 2026, European summer time ended at 01:00 UTC on Sunday 25 October, but Montreal's daylight time ends only at 02:00 local time on Sunday 1 November. For that week Seville is five hours ahead. A database that keeps the published local times now implies an 8:00 eastbound block and a 6:55 westbound block. The first wastes an hour; the second is impossible, because the aircraft needs about 7:55 to fly west. The schedule must change on one side:

```
from datetime import datetime
from zoneinfo import ZoneInfo

YUL = ZoneInfo("America/Toronto")  # Montreal
SVQ = ZoneInfo("Europe/Madrid")
UTC = ZoneInfo("UTC")

def to_utc(local, tz):
    t = datetime.fromisoformat(local)
    return t.replace(tzinfo=tz).astimezone(UTC)

dep = to_utc("2026-10-28T20:30", YUL)
arr = to_utc("2026-10-29T09:30", SVQ)
print(arr - dep)  # 8:00:00, an hour too long
```

Figure 24.2 shows the two repairs. Keeping Montreal times moves Seville's arrival to 08:30 and its departure to 11:45, local time; keeping Seville times moves Montreal's departure to 21:30 and its arrival to 15:40. After 1 November both cities are on standard time, Seville is again six hours ahead and the published local times work, but every UTC time is an hour later than in summer: the eastbound flight now leaves at 01:30 UTC. Note also that a Wednesday 20:30 departure from Montreal is a Thursday departure in UTC, so a pattern of operating days stored in UTC must shift its weekday flags.

The same problem returns every year at both ends of the season. For the real Seville service, which operates from September 2027 to June 2028, Seville is five hours ahead of Montreal from 31 October to 6 November 2027 and from 12 to 25 March

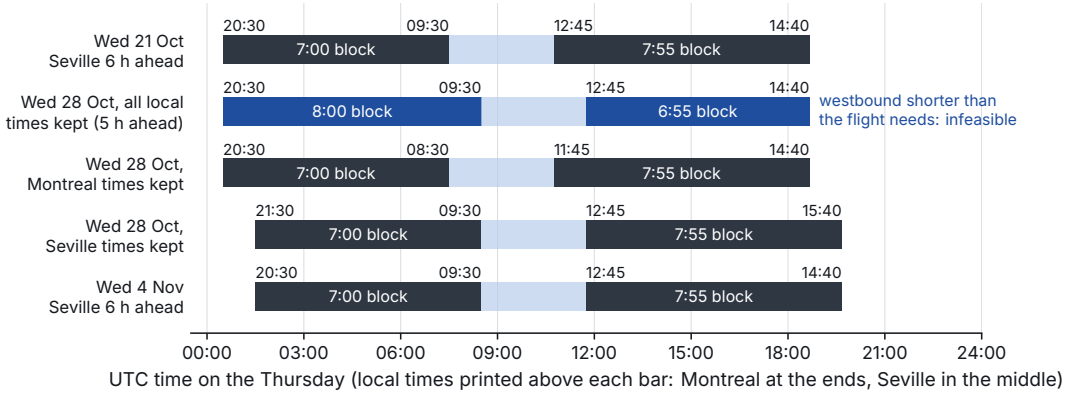


Figure 24.2 Laurentian's YUL–SVQ rotation drawn in UTC on three Wednesdays of autumn 2026, with local times printed above each bar. In the week of 28 October Seville is five hours ahead of Montreal: keeping every published local time implies blocks of 8:00 and 6:55, so one end must move. The light segment is the turn in Seville.

2028; in the spring of 2026 the gap lasted three weeks, from 8 to 28 March. By departure date the affected eastbound flights are those of 30 October to 6 November and 12 to 24 March, because Europe changes its clocks at 21:00 on the Saturday, Montreal time (Table 24.3). At coordinated airports the slot guidelines treat time changes during such periods as ad hoc changes, base historic precedence on the time held for most of the season, and ask coordinators to grant changes for periods of seven days or less as requested where feasible (IATA, ACI, and WWACG 2025).

Which repair is better is a network question: keeping Montreal times protects the connections of the Montreal banks, while keeping Seville times protects the Seville end and may break or create connections at Montreal (Chapter 13). The data question comes first: a pipeline that stores UTC and converts with the time zone database finds the problem automatically; one with a typed-in offset never does.

PITFALL

A typed-in offset. "Seville is UTC+2" is true for seven months of the year. Store time zone identifiers and let the time zone

> database supply the offset for each date; then recompute block times from UTC on dates on both sides of every clock change.

The schedule card Before a schedule is presented, it should fit on a card that a colleague in operations can check. The handbook lists the contents: both directions, operating dates and days, local and UTC times, block times, minimum and planned turns, aircraft and seats by cabin, feasible connections at both ends, the implied rotation, and the owner of each unverified assumption (Network Planning Handbook 2026). Table 24.3 is the card for Laurentian's service; the operating days are invented.

Table 24.3 A schedule card for Laurentian's YUL–SVQ service, with the fields the handbook lists (Network Planning Handbook 2026). Times are computed with the time zone database; operating days are illustrative.

Item	Eastbound LA958	Westbound LA959	Owner
Days	Wed, Fri, Sun from 8 Sep 2027; Mon added from 1 Oct	Thu, Sat, Mon; Tue added	network planning
Local times	YUL 20:30, SVQ 09:30 next day	SVQ 12:45, YUL 14:40	scheduling
UTC, six-hour weeks in summer	00:30 to 07:30	10:45 to 18:40	scheduling
UTC, six-hour weeks in winter	01:30 to 08:30	11:45 to 19:40	scheduling
Five-hour weeks (departure dates)	30 Oct to 6 Nov 2027; 12 to 24 Mar 2028	31 Oct to 6 Nov 2027; 12 to 25 Mar 2028	scheduling
Block	7:00	7:55	flight operations
Turn	3:15 at Seville; minimum to confirm	5:50 at Montreal when a 20:30 departure follows the same day	operations
Aircraft	A321XLR, 182 seats (14 business, 168 economy)	same	fleet
Connections	evening feed from Quebec City and Boston	afternoon onward flights at Montreal	network planning
Rotation	18:10 away from base per round trip		scheduling, maintenance
Unverified	westbound payload by season; Seville handling hours		flight operations, airports

24.6 Sensitivity matrices and decision thresholds

A recommendation is useful when it says what would change it. The handbook asks for base, adverse and favourable cases whose

assumptions hang together, since weak demand tends to come with weak fares and dear fuel with longer routings, rather than for every input moved on its own; the matrix it recommends crosses demand capture with retained fare and then reports the fuel or block-time threshold that reverses the recommendation (Network Planning Handbook 2026).

DEFINITION 24.3 Decision threshold

The **decision threshold** of an input is the value at which the ranking of two feasible alternatives reverses, with other inputs at their base values. For two inputs varied together, the set of reversal points is the **reversal frontier**.

RESULT 24.3 The reversal frontier of a route

Let the candidate's incremental network contribution be

$$C(n,f) = n\,(f - c) - K,$$

where n is passengers per departure, f the average retained fare, c the passenger-variable cost net of ancillary revenue per passenger, and K the remaining net cost, and let the best feasible alternative for the same aircraft contribute V. The candidate is preferred if and only if $n\,(f - c) > K + V$. The reversal frontier is the hyperbola

$$f^{*}(n) = c + \frac{K + V}{n},$$

along which a 1 percent fall in passengers is offset by a 1 percent rise in the fare margin $f - c$. At a base point (n_0, f_0) the single-input thresholds are $f^{*} = c + (K+V)/n_0$ and $n^{*} = (K+V)/(f_0 - c)$.

Proof. The preference condition is $C(n,f) > V$, which rearranges to $n\,(f - c) > K + V$; equality gives the frontier. Along it $\ln(f - c) = \ln(K + V) - \ln n$, so $\mathrm{d}\ln(f - c)/\,\mathrm{d}\ln n = -1$. Setting $n = n_0$ or $f = f_0$ gives the two thresholds. ∎

WORKED EXAMPLE 24.5 **The Seville decision, mapped**

The handbook's Seville case (Network Planning Handbook 2026) has 150 passengers at an average net fare of CAD 620 with CAD 30 of ancillary revenue each, CAD 3,500 of cargo contribution, an avoidable flight cost of CAD 88,000 at that load, CAD 6,000 gained on other flights, CAD 9,000 lost on the airline's own alternatives and CAD 2,000 of expected disruption cost. Laurentian's cost model splits the CAD 88,000 into CAD 80,500 for the flight and CAD 50 per passenger, which lets the load vary. In the notation of Result 24.3, $c = 50 - 30 = 20$ and $K = 80{,}500 - 3{,}500 - 6{,}000 + 9{,}000 + 2{,}000 = 82{,}000$, so $C(150, 620) = 150 \times 600 - 82{,}000 =$ CAD 8,000. A feasible alternative use of the same aircraft contributes $V =$ CAD 11,000, so the candidate is CAD 3,000 worse.

The single-input thresholds follow at once: $f^* = 20 + 93{,}000/150 =$ CAD 640, CAD 20 above the base fare, which is the handbook's threshold, and $n^* = 93{,}000/600 = 155$ passengers. Table 24.4 maps the surface. The candidate wins in the cells above the frontier, for example at 160 passengers and CAD 620, or 150 passengers and CAD 650.

Coherent scenarios read the same surface. An adverse case of 135 passengers at CAD 600 leaves the candidate CAD 14,700 behind; a favourable case of 165 passengers at CAD 640 puts it CAD 9,300 ahead. Other inputs have thresholds too. Each hour of A321XLR block costs CAD 11,000 in Laurentian's model, so the CAD 3,000 gap equals 16.4 minutes of block time on the candidate alone. Fuel is subtler: a fuel price change moves both alternatives, so what matters is the difference in fuel exposure. If fuel were 40 percent of the candidate's block cost of CAD 77,000 (seven hours at CAD 11,000), or CAD 30,800, and the alternative burned CAD 20,000 worth, a fuel price change of x would move the gap by $-10{,}800\,x$, and the candidate would catch up only if fuel fell by 28 percent.

The memo can now say: use the aircraft for the alternative, unless the retained fare reaches CAD 640 at 150 passengers or the expected load reaches 155 at CAD 620. Whether such fares or loads are attainable is a question for the demand model and for revenue management (Chapter 22).

Table 24.4 Candidate contribution minus that of the best alternative, in CAD per departure, by passengers per departure (rows) and average retained fare (columns), for the handbook's Seville case. The base case is in bold; positive cells favour the candidate.

Passengers	CAD 560	CAD 590	CAD 620	CAD 650	CAD 680
130	−22,800	−18,900	−15,000	−11,100	−7,200
140	−17,400	−13,200	−9,000	−4,800	−600
150	−12,000	−7,500	**−3,000**	1,500	6,000
160	−6,600	−1,800	3,000	7,800	12,600
170	−1,200	3,900	9,000	14,100	19,200

PITFALL

Independent toggles. Varying demand down and fares up at the same time, or fuel up without the matching detours, produces scenarios that cannot occur. Build scenarios from a story about the world, keep structural uncertainty, forecast error and execution risk separate, and report thresholds rather than a single point (Network Planning Handbook 2026).

24.7 The one-page decision memo

An analysis ends as a decision memo. The handbook's template asks for what decision is needed and by when; the baseline and its alternatives; dates, frequencies, aircraft and times; the market evidence; revenue kept by cabin and by flow; the definitions of cost and contribution; effects elsewhere in the network; whether aircraft, crews and airports can do it; the downside cases; the recommendation; and the open conditions, each with an owner (Network Planning Handbook 2026). Figure 24.3 arranges these on one page.

The most important line is the recommendation, and the most important part of it is the condition that would reverse it. The handbook's example has the right form: it recommends one schedule unless westbound payload limits pass a stated level or weaker feed pushes its contribution below the alternative's (Network Planning Handbook 2026). A sentence of that form tells the reader which assumption to watch and lets the team act when it changes; the thresholds of Section 24.6 supply the numbers. The handbook also allows the honest conclusion that the evidence is insufficient pending named inputs, and suggests a second page of evidence behind the page of recommendation.

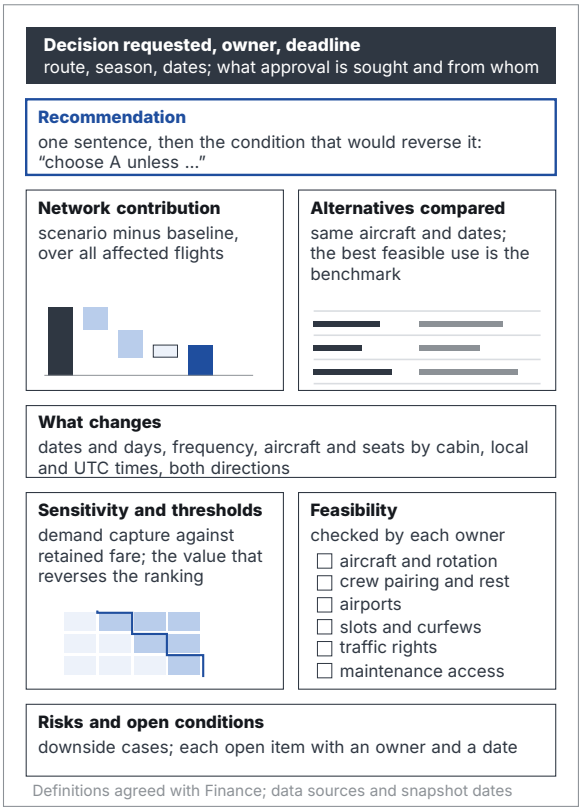


Figure 24.3 The layout of a one-page route decision memo. Each block names what it must contain: the decision and its owner, a one-sentence recommendation with its reversing condition, the incremental contribution bridge, the alternatives on the same aircraft, what changes in the schedule, the sensitivity and thresholds, the feasibility checks with their owners, and the open conditions.

The same discipline applies to two other documents the handbook describes. A **competitor-change note** records the change, the date it was seen and the travel dates it touches, counts only the capacity that will actually fly, and estimates share and fare effects under each plausible response. A **post-launch review** sets the forecast that was approved, not a later revision, against outcomes, explains the gap item by item (volume; the mix of fares, cabins and points of sale; cargo; cost; completion; disruption), and tells a slow ramp-up apart from a forecast that was biased (Network Planning Handbook 2026).

24.8 Audit checks

Checks are cheap and errors are expensive. Table 24.5 expands the handbook's list into the checks worth running on every analysis.

Table 24.5 Audit checks for a route analysis, expanded from the handbook's list (Network Planning Handbook 2026).

Check	What it catches	How to run it
Operated capacity reconciles	codeshare duplicates, missing flights	sum seats over operated flights; compare with the published schedule
Passenger weights reconcile	wrong expansion factors, sample breaks	compare expanded totals with T-100 segment or airport totals
Miles and kilometres	mixed units	one unit per ratio; ASK = 1.609344 × ASM
Directional demand against return capacity	forecasts that fill one direction and empty the other	compare passengers and seats by direction
Codeshares counted once	inflated capacity and share	count by operating carrier and flight
Actual against scheduled dates	cancelled or added flights treated as operated	join to operated records, not to plans
Bridge from market to carried passengers	jumps from market size to load	show market, share, spill, recapture and stimulation as steps
Sample coverage	false precision in thin markets	report coverage and intervals
Load factor weighting	averages of flight load factors	divide total RPK by total ASK
Identities	arithmetic and transcription errors	PRASM = yield × LF; TRASM − CASM = margin per ASM
Join multiplicity	many-to-many joins	row counts before and after each join
Clocks	typed-in offsets, midnight crossings	recompute blocks from UTC around every clock change
Vintage	look-ahead in backtests	keep snapshot dates; use only what was knowable
Currency and fare basis	mixed currencies, gross against net	state currency, rate date and fare treatment in the contract

Two of these deserve a number. Averaging load factors across flights is a classic error: a 137-seat flight from Montreal to Quebec City (233 kilometres) at 60 percent and a 298-seat flight from Mon-

treal to Paris (5,524 kilometres) at 90 percent average 75 percent, but together they fill 89.4 percent of their seat kilometres, because the long flight dominates both traffic and capacity (Chapter 2). And the identity check catches errors even in official filings: the printed passenger miles in Air Canada's second-quarter 2026 MD&A do not reproduce the printed load factor and yield exactly (Table 23.3), a small inconsistency that only recomputation reveals.

Before you trust a number, recompute it a second way: from another source, from an identity, or from the grain below it. Most planning errors are found by the second computation, not by a better model.

24.9 A reading plan and the first weeks

The handbook ends with a reading room and a learning sequence of short sessions, each with one calculation and one written conclusion (Network Planning Handbook 2026). Table 24.6 maps that sequence onto this book. The highest-priority book is The Global Airline Industry, read in the order of its chapters 3, 6, 7, 8, 5 and 10 (Belobaba, Odoni, and Barnhart 2015); Talluri and van Ryzin (2004b) is a reference to consult, not a book to read cover to cover before starting.

Table 24.6 A reading plan, adapted from the handbook's learning sequence and reading room (Network Planning Handbook 2026) and mapped onto this book.

Stage	Read	With	Produce
Sessions 1 to 3	Belobaba's lectures on airline economics, route evaluation and spill (Belobaba 2006d, 2006i, 2006b)	Chapters 2, 9 and 12	O&D against legs explained; the KPI identity and both handbook exercises reproduced
Sessions 4 to 6	schedule development and O&D control lectures (Belobaba 2006f, 2006n); the A321 planning manual (Airbus 2024)	Chapters 11, 13 and 20	a connection matrix, a seven-day rotation, trip cost against seat cost
Sessions 7 to 9	The Global Airline Industry, chapters 3, 6 and 7	Chapters 1, 3 and 4	a route contribution model and a two-page recommendation
Sessions 10 to 12	Air Canada's disclosures; the Competition Bureau study (Air Canada 2026e; Competition Bureau Canada 2025)	Chapter 23	a one-page company and network brief with open questions
First month	The Global Airline Industry, chapters 8, 5 and 10; Train; Hyndman and Athanasopoulos (Train 2009; Hyndman and Athanasopoulos 2021)	Chapters 5, 8 and 16	one analysis replicated, its inputs audited, one forecast backtested
Months 2 and 3	itinerary-based fleet assignment, integrated schedule design, robust routing; the slot guidelines (Barnhart, Kniker, and Lohatepanont 2002; Lohatepanont and Barnhart 2004; Lan, Clarke, and Barnhart 2006; IATA, ACI, and WWACG 2025)	Chapters 14 and 16	one bounded scenario owned, with an alternative recommendation

The first weeks are also a research project with its own method: establish definitions before using numbers. Table 24.7 lists what to establish with each team, adapted from the handbook.

Table 24.7 What to establish in the first weeks, by team, adapted from the handbook (Network Planning Handbook 2026).

Team	What to establish
Manager and network planning	the markets and horizons you own; the planning calendar; what makes a recommendation ready for approval
Forecasting	the agreed baseline; the inputs an analyst may change; how the demand, share, stimulation, spill and recapture models are fitted and tested
Revenue management	where scarce capacity is most valuable; the fare and cabin assumptions planning should adopt; how a schedule change shifts availability and displacement
Finance	the official definitions of contribution; the treatment of connecting, partner, cargo and loyalty revenue; which costs can be avoided at each horizon
Scheduling and fleet	the system of record; publication and freeze dates; how tails, maintenance and reserves are checked
Crew and operations	which changes create step costs or infeasibility; the reliability assumptions and recovery options
Airports and slots	the binding gates, handling windows, restrictions and slots; who validates dates and exceptions
Alliances, legal and regulatory	which partner flows and commercial rights are in scope; how permissions and joint-venture effects enter the case
Commercial, sales and cargo	the customer and cargo commitments that matter; the evidence behind demand claims; what the forecast already includes

Three habits complete the method (Network Planning Handbook 2026). Ask to see a route paper that was approved, one that was rejected and one post-launch review; three real documents teach a team's handling of uncertainty and constraints faster than any process chart. Rebuild a known result from its source inputs, and account for every difference, before trying to improve a model. And hold on to the distinctions that matter more than any aircraft specification: a demand forecast versus passengers carried; allocated profit

versus incremental value; the airline's own contribution versus the joint venture's; a connection that is valid versus one that is reliable; choosing a fleet type versus routing a tail; and a model's output versus an approved operating plan.

24.10 Summary

- Keep four objects apart: aircraft movements, flight legs, passenger journeys and tickets. Every table has a grain, and every aggregate and join must respect it.
- Schedules measure supply well; market data estimate demand from incomplete booking data. Treat market totals as estimates with a coverage problem that differs by carrier and market.
- The US ticket survey changed in July 2025 from a 10 percent quarterly sample to a 40 percent monthly one. Expand each series by its own factor and never splice by rescaling.
- An expanded sample count has a coefficient of variation of at least $\sqrt{(1-p)/(pN)}$; thin markets deserve intervals, not points.
- Write a data contract for every table and keep an assumption register; use SQL for joins, Python for estimation and backtests, and a spreadsheet for transparent scenarios.
- Store UTC and local times with time zone identifiers. When clocks change on different dates, fixed local times move block time from one direction to the other; Montreal and Seville are five hours apart for a week each autumn and two or three weeks each spring.
- Report thresholds, not just points: in the handbook's Seville case the decision reverses at CAD 640 at 150 passengers, or 155 passengers at CAD 620.
- A one-page memo states the recommendation and the condition that would reverse it, with owners for every open item.
- Recompute every important number a second way before trusting it.

24.11 Further reading

The handbook (Network Planning Handbook 2026) is the source of the workflow, the data catalogue, the templates and the reading plan of this chapter, and repays rereading after the first month. For data, read the providers' own descriptions (Cirium 2026a; OAG n.d.) and the BTS pages on the new survey (Bureau of Transportation Statistics 2026b); Lohr (2021) covers the sampling theory be-

hind expanded counts. For forecasting workflow and backtesting, Hyndman and Athanasopoulos (2021) is the practical reference. For schedules and clocks, the Worldwide Airport Slot Guidelines (IATA, ACI, and WWACG 2025) are short enough to read in full. Belobaba, Odoni and Barnhart (2015) remain the desk reference for everything else.

APPENDIX

A

Probability and Statistics Toolkit

This appendix collects the probability and statistics that the chapters use, with short derivations, so that a formula met in a chapter can be checked in one place. It is a toolkit rather than a course: Casella and Berger (2002) and Wooldridge (2010) cover all of it and much more. The sections follow the order in which the tools appear in the book: expectations and conditioning (everywhere), the normal distribution and its loss function (spill, Chapter 9), counting distributions (bookings and no-shows, Chapters 18 and 19), the Gumbel distribution behind logit (Chapter 5), limit theorems, estimation and testing (Chapters 5 and 7), censored data (Chapter 9) and prediction intervals (Chapter 8). Notation follows the rest of the book: φ and Φ are the standard normal density and distribution function, $L(z)$ is the loss function and $k = \sigma/\mu$ the k-factor.

A.1 Random variables, expectation and conditioning

A **random variable** X is described by its distribution function $F(x) = \Pr(X \le x)$, and by a probability mass function $p(x) = \Pr(X = x)$ if it is discrete or a density $f(x) = F'(x)$ if it is continuous. Its expectation is $\mathbb{E}[g(X)] = \sum_x g(x)\,p(x)$ or $\int g(x) f(x)\,\mathrm{d}x$. Expectation is linear, $\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$, whether or not X and Y are independent.

DEFINITION A.1 Moments

For random variables X and Y with means μ_X and μ_Y:

- the **variance** is $\operatorname{Var}(X) = \mathbb{E}[(X-\mu_X)^2] = \mathbb{E}[X^2] - \mu_X^2$, and the standard deviation is $\sigma_X = \sqrt{\operatorname{Var}(X)}$;
- the **coefficient of variation**, called the k-factor when X is demand, is $k = \sigma_X / \mu_X$;
- the **covariance** is $\operatorname{Cov}(X,Y) = \mathbb{E}[(X-\mu_X)(Y-\mu_Y)]$, which equals $\mathbb{E}[XY] - \mu_X\mu_Y$;
- the **correlation** is $\operatorname{Cov}(X,Y)/(\sigma_X\sigma_Y)$, a number between -1 and 1.

Two rules do most of the work: $\operatorname{Var}(aX+b) = a^2 \operatorname{Var}(X)$, and

$$\operatorname{Var}\Big(\sum_i X_i\Big) = \sum_i \operatorname{Var}(X_i) + 2\sum_{i<j} \operatorname{Cov}(X_i, X_j).$$

Independent variables have zero covariance (the converse is false), so their variances add.

RESULT A.1 Expectation from the tail

If $X \geq 0$, then $\mathbb{E}[X] = \int_0^\infty \Pr(X > x)\,\mathrm{d}x$, and for any capacity $c \geq 0$

$$\mathbb{E}[\min(X,c)] = \int_0^c \Pr(X > x)\,\mathrm{d}x.$$

For integer-valued $X \geq 0$, $\mathbb{E}[X] = \sum_{n\geq 0} \Pr(X > n)$.

Proof. Write $X = \int_0^\infty \mathbb{1}\{X > x\}\,\mathrm{d}x$ and $\min(X,c) = \int_0^c \mathbb{1}\{X > x\}\,\mathrm{d}x$, then take expectations inside the integral, which is allowed because the integrand is nonnegative (Fubini's theorem). The integer case replaces the integral by a sum over $n = 0, 1, 2, \ldots$ ∎

This is the formula behind the marginal seat of Chapter 9: seat x is occupied exactly when demand exceeds x. A related fact is **Jensen's inequality**: if g is concave, $\mathbb{E}[g(X)] \leq g(\mathbb{E}[X])$, because a concave function lies below its tangent at the mean, $g(x) \leq g(\mu) + g'(\mu)(x-\mu)$, and the tangent has expectation $g(\mu)$. Load $\min(D, C)$ is concave in demand, so $\mathbb{E}[\min(D,C)] \leq \min(\mu, C)$: replacing demand by its mean always overstates the load.

The **conditional expectation** $\mathbb{E}[Y \mid X]$ is the mean of Y computed from the distribution of Y given X. It is a function of X, and therefore itself a random variable.

RESULT A.2 Iterated expectations and total variance

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y \mid X]],$$
$$\operatorname{Var}(Y) = \mathbb{E}[\operatorname{Var}(Y \mid X)] + \operatorname{Var}(\mathbb{E}[Y \mid X]).$$

Moreover, $\mathbb{E}[Y \mid X]$ is the best predictor of Y given X: it minimizes $\mathbb{E}[(Y - g(X))^2]$ over all functions g.

Proof. For continuous variables,

$$\mathbb{E}[\mathbb{E}[Y \mid X]] = \int \Big(\int y f(y \mid x)\, \mathrm{d}y \Big) f(x)\, \mathrm{d}x = \iint y f(x, y)\, \mathrm{d}y\, \mathrm{d}x = \mathbb{E}[Y];$$

sums replace integrals in the discrete case. For the variance, start from $\operatorname{Var}(Y) = \mathbb{E}[\mathbb{E}[Y^2 \mid X]] - (\mathbb{E}[\mathbb{E}[Y \mid X]])^2$ and add and subtract $\mathbb{E}[(\mathbb{E}[Y \mid X])^2]$: the first two terms form $\mathbb{E}[\operatorname{Var}(Y \mid X)]$ and the last two $\operatorname{Var}(\mathbb{E}[Y \mid X])$. For the predictor, split $Y - g(X)$ into $Y - \mathbb{E}[Y \mid X]$ and $\mathbb{E}[Y \mid X] - g(X)$. The cross term has expectation zero by iterated expectations, so

$$\mathbb{E}[(Y - g)^2] = \mathbb{E}[(Y - \mathbb{E}[Y \mid X])^2] + \mathbb{E}[(\mathbb{E}[Y \mid X] - g)^2],$$

which is smallest when $g(X) = \mathbb{E}[Y \mid X]$. ∎

The total-variance formula explains why demand varies more than any model of individual bookings predicts (Appendix A.4), and the best-predictor property is why forecasts aim at conditional means (Chapter 8).

WORKED EXAMPLE A.1 Pooling two segments

On a flight, business demand has mean 40 and standard deviation 12, and leisure demand has mean 150 and standard deviation 40, the two segments used in Chapter 9. If they are independent, total demand has mean 190, variance $12^2 + 40^2 = 1{,}744$ and standard deviation 41.8, a k-factor of $41.8/190 = 0.22$, below that of either segment (0.30 and 0.27). Pooling reduces relative variability because the segments' shocks partly cancel. If the segments are positively correlated, as when a city-wide event lifts both, the gain shrinks: with correlation 0.5 the variance is $1{,}744 + 2 \times 0.5 \times 12 \times 40 = 2{,}224$ and the k-factor 0.25. A leg that carries many O&D markets (Chapter 10) gains from pooling in the same way, to the extent that the markets' shocks are not common.

A.2 The normal distribution

A standard normal variable Z has density $\varphi(z) = e^{-z^2/2}/\sqrt{2\pi}$ and distribution function $\Phi(z)$. $X \sim \mathcal{N}(\mu, \sigma^2)$ means that $Z = (X-\mu)/\sigma$ is standard normal, so that $\Pr(X \le x) = \Phi((x-\mu)/\sigma)$; this **standardization** reduces every normal calculation to one table. Three facts are used repeatedly: symmetry, $\varphi(-z) = \varphi(z)$ and $\Phi(-z) = 1 - \Phi(z)$; the derivative $\varphi'(z) = -z\,\varphi(z)$; and the quantiles $\Phi^{-1}(0.90) = 1.2816$, $\Phi^{-1}(0.95) = 1.6449$ and $\Phi^{-1}(0.975) = 1.9600$.

Linear combinations of jointly normal variables are normal; in particular, sums of independent normal variables are normal, with means and variances that add. If X and Y are jointly normal with correlation ρ, then X given Y is normal with

$$\mathbb{E}[X \mid Y] = \mu_X + \frac{\operatorname{Cov}(X,Y)}{\operatorname{Var}(Y)}\,(Y - \mu_Y), \qquad \operatorname{Var}(X \mid Y) = \sigma_X^2\,(1-\rho^2),$$

a conditional mean that is linear in Y. This is the property behind the control functions of Chapter 7.

If $\ln X \sim \mathcal{N}(m, s^2)$, X is **lognormal**, with median e^m and mean $e^{m+s^2/2}$. The difference matters for models written in logarithms, such as the log-linear demand functions of Chapter 4: exponentiating a forecast of $\ln X$ gives the median, and the mean is larger by the factor $e^{s^2/2}$, about 4.6 percent when $s = 0.3$.

A.3 The loss function and truncated moments

DEFINITION A.2 Loss function and inverse Mills ratio

The standard normal **loss function** is the expected amount by which a standard normal variable exceeds z,

$$L(z) = \mathbb{E}[[Z - z]^+] = \int_z^\infty (t - z)\,\varphi(t)\,\mathrm{d}t.$$

The **inverse Mills ratio** $\lambda(z) = \varphi(z)/(1 - \Phi(z))$ is the hazard rate of the standard normal.

RESULT A.3 Loss-function identities

For every z:

1. $L(z) = \varphi(z) - z(1 - \Phi(z))$;
2. $L'(z) = -(1-\Phi(z))$ and $L''(z) = \varphi(z) > 0$, so L is decreasing and convex;
3. $L(-z) = L(z) + z$;
4. $L(z) = (1 - \Phi(z))(\lambda(z) - z)$, the probability of exceeding z times the mean excess;
5. $L(0) = \varphi(0) = 0.3989$; $L(z) - [-z]^+ \to 0$ as $|z| \to \infty$; and $0 < L(z) < \varphi(z)/z^2$ for $z > 0$.

If $X \sim \mathcal{N}(\mu, \sigma^2)$ and $z = (c - \mu)/\sigma$, then

$$\begin{aligned} &\mathbb{E}[[X - c]^+] = \sigma L(z), \qquad \mathbb{E}[[c - X]^+] = \sigma(L(z) + z), \\ &\mathbb{E}[\min(X, c)] = \mu - \sigma L(z). \end{aligned} \tag{A.1}$$

Proof. For identity 1, split the integral and use $\int_z^\infty t\,\varphi(t)\,\mathrm{d}t = \varphi(z)$, which follows from $\varphi'(t) = -t\,\varphi(t)$. For 2, differentiate (1) to get $-z\varphi(z) - (1 - \Phi(z)) + z\varphi(z)$, then differentiate again. For 3, note that since $[Z - z]^+ - [z - Z]^+ = Z - z$, taking expectations gives $L(z) - \mathbb{E}[[z - Z]^+] = -z$, and $\mathbb{E}[[z - Z]^+] = \mathbb{E}[[-Z - (-z)]^+] = L(-z)$ because $-Z$ is also standard normal. For 4, substitute $\varphi(z) = \lambda(z)(1 - \Phi(z))$ in (1). For 5, the value at zero is (1) with $z = 0$. As $z \to \infty$, $L(z) \to 0$; as $z \to -\infty$, (3) gives $L(z) + z = L(-z) \to 0$. The bound follows from the classical inequality $1 - \Phi(z) > \varphi(z)(1/z - 1/z^3)$ for $z > 0$, substituted in (1). Finally, for $X = \mu + \sigma Z$ we have $[X - c]^+ = \sigma\,[Z - z]^+$; the other two expectations follow from (3) and from $\min(X, c) = X - [X - c]^+$. ∎

With c a capacity, the three expectations in (A.1) are expected spill, expected empty seats and expected load, and identity 2 is the value of the marginal seat: the derivative of $\mathbb{E}[\min(X, c)]$ with respect to c is $1 - \Phi(z) = \Pr(X > c)$. Table A.1 tabulates the functions and Figure A.1 draws L.

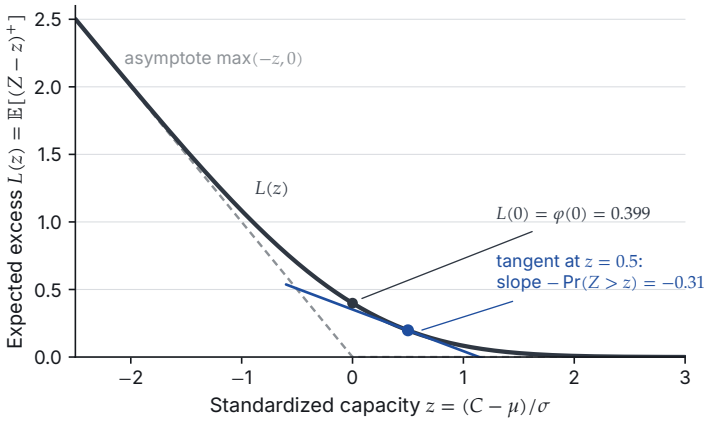


Figure A.1 The standard normal loss function. With capacity z standard deviations above mean demand, $L(z)$ is the expected spill in standard deviations. It approaches the deterministic answer $\max(-z, 0)$ when capacity is far from mean demand and departs from it most at $z = 0$; its slope at any z is minus the probability that demand exceeds capacity.

Table A.1 The standard normal density, tail probability, loss function and inverse Mills ratio. In a spreadsheet, $L(z)$ is `NORM.S.DIST(z,FALSE)-z*(1-NORM.S.DIST(z,TRUE))`.

z	$\varphi(z)$	$1-\Phi(z)$	$L(z)$	$L(-z)$	$\lambda(z)$
−2.00	0.0540	0.9772	2.0085	0.0085	0.0552
−1.00	0.2420	0.8413	1.0833	0.0833	0.2876
−0.50	0.3521	0.6915	0.6978	0.1978	0.5092
0.00	0.3989	0.5000	0.3989	0.3989	0.7979
0.25	0.3867	0.4013	0.2863	0.5363	0.9636
0.50	0.3521	0.3085	0.1978	0.6978	1.1411
0.75	0.3011	0.2266	0.1312	0.8812	1.3288
1.00	0.2420	0.1587	0.0833	1.0833	1.5251
1.25	0.1826	0.1056	0.0506	1.3006	1.7288
1.50	0.1295	0.0668	0.0293	1.5293	1.9387
2.00	0.0540	0.0228	0.0085	2.0085	2.3732
2.50	0.0175	0.0062	0.0020	2.5020	2.8227

The moments of a normal variable beyond a threshold follow from the same integrals.

RESULT A.4 Truncated normal moments

Let $X \sim \mathcal{N}(\mu, \sigma^2)$, let c be a threshold and $z = (c - \mu)/\sigma$. Then

$$\begin{aligned} \mathbb{E}[X \mid X > c] &= \mu + \sigma\lambda(z), \\ \mathrm{Var}(X \mid X > c) &= \sigma^2[1 + z\lambda(z) - \lambda(z)^2], \\ \mathbb{E}[X \mid X < c] &= \mu - \sigma\,\varphi(z)/\Phi(z). \end{aligned} \tag{A.2}$$

In standard units, $\mathbb{E}[Z \mid Z > z] = \lambda(z)$ and $\mathbb{E}[Z^2 \mid Z > z] = 1 + z\lambda(z)$, the two moments used by the EM algorithm of Chapter 9.

Proof. Over the upper tail, $\int_z^\infty t\,\varphi(t)\,\mathrm{d}t = \varphi(z)$, and integration by parts with $u = t$ and $\mathrm{d}v = t\,\varphi(t)\,\mathrm{d}t$ gives $\int_z^\infty t^2\varphi(t)\,\mathrm{d}t = z\varphi(z) + 1 - \Phi(z)$. Dividing by $\Pr(Z > z) = 1 - \Phi(z)$ gives the two standard moments, and the variance is $1 + z\lambda - \lambda^2$; the substitution $X = \mu + \sigma Z$ rescales. For the lower tail, apply the upper-tail result to $-Z$, whose threshold is $-z$, and use $\varphi(-z)/(1 - \Phi(-z)) = \varphi(z)/\Phi(z)$. ■

The mean beyond a threshold lies beyond it, $\lambda(z) > z$, and the excess $\lambda(z) - z$ shrinks as the threshold rises: far in the tail, the departures that exceed capacity exceed it only a little.

WORKED EXAMPLE A.2 One flight, five numbers

Take the flight of Chapter 9, with demand $\mathcal{N}(160, 40^2)$ and 180 seats, so $z = 0.5$. From Table A.1, expected spill is $40 \times 0.1978 = 7.91$, expected empty seats $40 \times 0.6978 = 27.91$ and expected load $160 - 7.91 = 152.09$; load and empty seats add up to the 180 seats, as they must. The flight fills with probability 0.3085, and on those departures demand averages $160 + 40 \times 1.1411 = 205.6$ with a standard deviation of $40\sqrt{1 + 0.5 \times 1.1411 - 1.1411^2} = 20.7$. Identity 4 ties the numbers together: the probability of filling times the mean excess on full departures, 0.3085×25.6, is the expected spill of 7.91.

A.4 Counting distributions

Binomial If each of n independent trials succeeds with probability q, the number of successes $X \sim \mathrm{Bin}(n, q)$ has

$$\Pr(X = x) = \binom{n}{x} q^x (1 - q)^{n-x}, \qquad x = 0, 1, \ldots, n,$$

with $\mathbb{E}[X] = nq$ and $\mathrm{Var}(X) = nq(1-q)$. The moments follow from writing X as a sum of n independent indicators, each with mean q and variance $q(1-q)$. With n bookings and show-up probability q, X counts the passengers who turn up, the starting point of overbooking (Chapter 19). Independence is the weak assumption: families and groups show up or fail to show together, which makes the count more variable than the binomial.

Poisson $N \sim \mathrm{Poisson}(\lambda)$ has $\Pr(N = n) = e^{-\lambda}\lambda^n/n!$ for $n = 0, 1, 2, \ldots$ and $\mathbb{E}[N] = \mathrm{Var}(N) = \lambda$ (this λ is a mean, unrelated to the inverse Mills ratio $\lambda(z)$). It is the limit of $\mathrm{Bin}(n, q)$ as $n \to \infty$ with $nq = \lambda$ fixed: the count of many rare, independent events. Booking requests are usually modelled as a **Poisson process** with rate λ per unit of time, which has four properties that revenue management models rely on (Chapters 18 and 20):

- the numbers of arrivals in disjoint intervals are independent, and the number in an interval of length t is $\mathrm{Poisson}(\lambda t)$ (with a rate that varies over the booking horizon, the mean is the integral of the rate);
- in a short interval of length Δ, one arrival occurs with probability $\lambda\Delta$ and two or more with negligible probability, so a dynamic program can treat each small period as having at most one request;
- if each arrival is independently of type j with probability p_j, the arrivals of each type form independent Poisson processes with rates λp_j (thinning), and independent Poisson processes add up to one with the summed rate (superposition);
- the times between arrivals are independent and exponential, with mean $1/\lambda$.

Thinning takes one line for two types: with $n = n_1 + n_2$,

$$\begin{aligned}\Pr(N_1 = n_1, N_2 = n_2) &= \frac{e^{-\lambda}\lambda^n}{n!}\binom{n}{n_1}p^{n_1}(1-p)^{n_2} \\ &= \frac{e^{-\lambda p}(\lambda p)^{n_1}}{n_1!} \cdot \frac{e^{-\lambda(1-p)}(\lambda(1-p))^{n_2}}{n_2!},\end{aligned}$$

a product of two Poisson probabilities.

Gamma $X \sim \mathrm{Gamma}(a, \theta)$, with shape $a > 0$ and scale $\theta > 0$, has density $f_a(x) = x^{a-1}e^{-x/\theta}/(\Gamma(a)\theta^a)$ for $x > 0$, mean $a\theta$, variance $a\theta^2$ and k-factor $1/\sqrt{a}$. It is skewed to the right for small a and close to normal for large a; $a = 1$ is the exponential distribution, and independent gamma variables with a common scale add up to a gamma

with the shapes added. Its loss function has a closed form. Since $x f_a(x) = a\theta f_{a+1}(x)$,

$$\mathbb{E}[[X-c]^+] = a\theta(1-G_{a+1}(c)) - c(1-G_a(c)),$$

where G_a is the Gamma(a,θ) distribution function. This makes the gamma a convenient skewed alternative to the normal in spill calculations for small flights (Swan 2002).

Overdispersion and the negative binomial A Poisson count has k-factor $1/\sqrt{\lambda}$: 0.08 for a flight with mean demand 150, far below the values of 0.2 to 0.4 found in airline data (Chapter 9). The reason is that the arrival rate itself varies from one departure to the next, with weather, events and the economy.

RESULT A.5 The Poisson-gamma mixture

If $N \mid \Lambda \sim \text{Poisson}(\Lambda)$ and $\Lambda \sim \text{Gamma}(a,\theta)$, then N has the **negative binomial** distribution

$$\Pr(N=n) = \frac{\Gamma(n+a)}{n!\,\Gamma(a)}\, p^a (1-p)^n, \qquad p = \frac{1}{1+\theta},$$

with mean $\mu = a\theta$ and variance $\mu + \mu^2/a$. Its k-factor satisfies $k^2 = 1/\mu + 1/a$, so as demand grows it falls toward the floor $1/\sqrt{a}$ set by the variability of the rate.

Proof. Integrate the Poisson probability against the gamma density:

$$\begin{aligned}\Pr(N=n) &= \int_0^\infty \frac{e^{-\ell}\ell^n}{n!} \cdot \frac{\ell^{a-1}e^{-\ell/\theta}}{\Gamma(a)\,\theta^a}\,\mathrm{d}\ell \\ &= \frac{\Gamma(n+a)}{n!\,\Gamma(a)\,\theta^a}\Big(\frac{\theta}{1+\theta}\Big)^{n+a},\end{aligned}$$

using the gamma integral $\int_0^\infty \ell^{b-1}e^{-\ell/s}\,\mathrm{d}\ell = \Gamma(b)\,s^b$ with $b = n+a$ and $s = \theta/(1+\theta)$; the result rearranges to the stated form. The moments follow from Result A.2: $\mathbb{E}[N] = \mathbb{E}[\Lambda] = a\theta$, and $\text{Var}(N) = \mathbb{E}[\text{Var}(N \mid \Lambda)] + \text{Var}(\mathbb{E}[N \mid \Lambda])$, which is $\mathbb{E}[\Lambda] + \text{Var}(\Lambda) = a\theta + a\theta^2$. ■

WORKED EXAMPLE A.3 How variable must the arrival rate be?

For a flight with mean demand 150, a Poisson model gives $k = 1/\sqrt{150} = 0.08$. To reach a k-factor of 0.30, the negative binomial needs $1/a = 0.30^2 - 1/150 = 0.083$, so $a = 12$: the arrival rate must vary across departures with a coefficient of variation of $1/\sqrt{12} = 0.29$. Almost all of the variability comes from the rate, not from the randomness of individual bookings. The shape of the distribution matters a little for spill. With 180 seats, a normal distribution with mean 150 and standard deviation 45 gives expected spill of 6.80 passengers, while a gamma and a negative binomial with the same mean and variance give 7.62 and 7.59, because their right tails are longer. For a small flight (mean 20, k-factor 0.4, 24 seats) the three give 1.58, 1.73 and 1.71, and the normal model puts 0.6 percent of its probability on negative demand.

A.5 The Gumbel distribution and logit

DEFINITION A.3 Gumbel distribution

The **Gumbel**, or type I extreme value, distribution with location m and unit scale has

$$F(x) = \exp(-e^{-(x-m)}), \qquad f(x) = e^{-(x-m)} \exp(-e^{-(x-m)}),$$

mode m, mean $m + \gamma_{\mathrm{E}}$ with Euler's constant $\gamma_{\mathrm{E}} = 0.5772$, and variance $\pi^2/6$. Logit models assume independent standard Gumbel errors ($m = 0$); fixing the scale at one normalizes the scale of utility (Chapter 5).

RESULT A.6 Max-stability and the logit formula

Let $\varepsilon_1, \dots, \varepsilon_J$ be independent standard Gumbel variables and $V_1, \dots, V_J$ constants.

1. The maximum $\max_j (V_j + \varepsilon_j)$ is Gumbel with location $\ln \sum_j e^{V_j}$, so its mean is the log-sum $\ln \sum_j e^{V_j}$ plus γ_{E}.
2. The difference $\varepsilon_2 - \varepsilon_1$ has the **logistic** distribution, $\Pr(\varepsilon_2 - \varepsilon_1 \le x) = 1/(1+e^{-x})$, with mean 0 and variance $\pi^2/3$.
3. The probability that alternative i has the highest utility $V_i + \varepsilon_i$ is $e^{V_i} / \sum_j e^{V_j}$.

Proof. For part 1, by independence,

$$\Pr(\max_j (V_j + \varepsilon_j) \le x) = \prod_j \exp(-e^{-(x - V_j)}) = \exp\Big(-e^{-x} \sum_j e^{V_j}\Big),$$

which is the Gumbel distribution function with $m = \ln \sum_j e^{V_j}$. For part 2, condition on $\varepsilon_1 = t$ and substitute $w = e^{-t}$:

$$\begin{aligned} \Pr(\varepsilon_2 - \varepsilon_1 \le x) &= \int_{-\infty}^{\infty} \exp(-e^{-(t+x)})\, e^{-t} \exp(-e^{-t})\, \mathrm{d}t \\ &= \int_0^{\infty} e^{-w(1+e^{-x})}\, \mathrm{d}w = \frac{1}{1+e^{-x}}, \end{aligned}$$

and the variance of a difference of independent variables is the sum of their variances. For part 3, by part 1 the best alternative other than i has utility $m_{-i} + \varepsilon^\star$, with $m_{-i} = \ln \sum_{j \ne i} e^{V_j}$ and $\varepsilon^\star$ standard Gumbel and independent of ε_i. By part 2,

$$\Pr(V_i + \varepsilon_i > m_{-i} + \varepsilon^\star) = \frac{1}{1 + e^{m_{-i} - V_i}} = \frac{e^{V_i}}{e^{V_i} + \sum_{j \ne i} e^{V_j}}.$$

■

Part 1 is why the log-sum measures the expected value of a choice set and, divided by the fare coefficient, consumer surplus (Chapter 5). With $V = (1, 0.5, -0.2)$, for example, the log-sum is 1.646, the expected maximum utility 2.223 and the choice probabilities 0.524, 0.318 and 0.158; a simulation with 400,000 sets of Gumbel draws reproduces them to within 0.002.

Figure A.2 also explains why logit and probit coefficients differ when both are estimated on the same data. The logistic has standard deviation $\pi/\sqrt{3} = 1.81$ instead of 1 and density 0.25 instead of 0.399

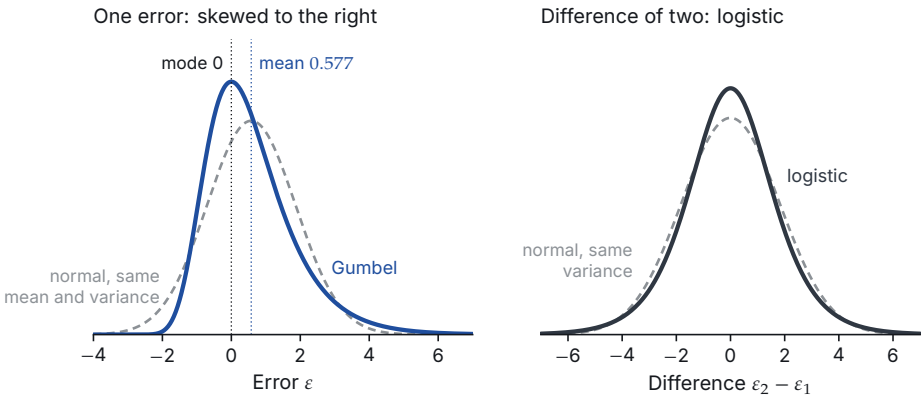


Figure A.2 Left: the standard Gumbel density is skewed to the right, with mode 0 and mean 0.577; the dashed curve is a normal with the same mean and variance. Right: the difference of two independent Gumbel errors is logistic, close to a normal with the same variance but with heavier tails.

at its centre, so logit coefficients are larger than probit coefficients by a factor between about 1.6 (matching slopes at the centre) and 1.8 (matching variances). Only ratios of coefficients, such as a value of time, are comparable across the two models.

A.6 Limit theorems and the normal approximation

Let $X_1, X_2, \ldots$ be independent with mean μ and finite variance σ^2, and let $\bar{X}_n$ be the mean of the first n. The **law of large numbers** says that $\bar{X}_n$ converges to μ. The **central limit theorem** says that $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ converges in distribution to $\mathcal{N}(0,1)$. Versions for independent variables that are not identically distributed (under Lindeberg's condition, which rules out any single term dominating) justify treating a sum of many small independent contributions as normal: the demand for a departure, built from many independent booking decisions (Chapter 9), and the estimators of the next section, which are averages in disguise.

The binomial is such a sum, so for large n it is approximately $\mathcal{N}(nq, nq(1-q))$. Because it is discrete, the approximation is better with a **continuity correction**,

$$\Pr(X \le x) \approx \Phi\Big(\frac{x + 0.5 - nq}{\sqrt{nq(1-q)}}\Big),$$

and it is accurate once $nq(1-q)$ exceeds about 10. The same holds for the Poisson, approximately $\mathcal{N}(\lambda, \lambda)$ for large λ.

WORKED EXAMPLE A.4 Show-ups on an overbooked flight

A 182-seat aircraft has 195 bookings, and each passenger shows up independently with probability 0.92 (illustrative values). Show-ups are $\text{Bin}(195, 0.92)$, with mean 179.4 and standard deviation $\sqrt{195 \times 0.92 \times 0.08} = 3.79$. The exact probability that more than 182 passengers show up is 0.210. The normal approximation with the continuity correction gives $z = (182.5 - 179.4)/3.79 = 0.818$ and $1 - \Phi(0.818) = 0.207$; without the correction it gives 0.246, an error of 17 percent. For expected denied boardings, $\mathbb{E}[[X - 182]^+]$, the exact value is 0.52 passengers per flight, while the normal loss formula $\sigma L(z)$ gives 0.44 with the continuity correction and 0.55 without (the half-seat shift corrects probabilities, not expected excesses). At this scale an exact binomial calculation is worth the extra line of code.

A.7 Estimation and inference

Least squares For $y_i = x_i'\beta + u_i$, the OLS estimator $\hat{\beta} = (X'X)^{-1}X'y$ solves the sample moment conditions $\sum_i x_i \hat{u}_i = 0$. It is consistent when $\mathbb{E}[x_i u_i] = 0$, the exogeneity condition of Definition 7.1, and in large samples it is approximately normal with the heteroskedasticity-robust covariance $(X'X)^{-1}(\sum_i \hat{u}_i^2 x_i x_i')(X'X)^{-1}$. When errors are correlated within groups (departures of the same week, quarters of the same route), the middle term sums $x_i\hat{u}_i$ within each group before squaring: the cluster-robust covariance. Chapter 7 treats what happens when exogeneity fails.

Maximum likelihood For independent observations with density $f(y_i;\theta)$, the log-likelihood is $\ell(\theta) = \sum_i \ln f(y_i;\theta)$ and the maximum likelihood estimator (MLE) $\hat{\theta}$ maximizes it. The gradient $\nabla\ell(\theta)$ is the **score**, and the **Fisher information** per observation is $\mathcal{I}(\theta) = \text{Var}(\nabla \ln f(y_i;\theta))$.

RESULT A.7 Large-sample behaviour of maximum likelihood

If the model is correctly specified, the true θ_0 is an interior point and the likelihood is smooth, then:

1. the score has mean zero at θ_0, and the information equality holds, $\mathcal{I}(\theta_0) = -\mathbb{E}[\nabla^2 \ln f(y_i;\theta_0)]$;
2. $\hat{\theta}$ is consistent and $\sqrt{n}(\hat{\theta}-\theta_0)$ converges to $\mathcal{N}(0, \mathcal{I}(\theta_0)^{-1})$, so standard errors come from the inverse of the negative Hessian, $[-\nabla^2 \ell(\hat{\theta})]^{-1}$;
3. for r restrictions on θ, the likelihood ratio statistic $\mathrm{LR} = 2[\ell(\hat{\theta}) - \ell(\tilde{\theta})]$, where $\tilde{\theta}$ maximizes the likelihood under the restrictions, is approximately χ^2_r when the restrictions are true.

Proof. For part 1, differentiating $\int f(y;\theta)\,\mathrm{d}y = 1$ under the integral gives $\int \nabla f \,\mathrm{d}y = \mathbb{E}[\nabla \ln f] = 0$; differentiating once more gives $\mathbb{E}[\nabla^2 \ln f] + \mathbb{E}[\nabla \ln f \, \nabla \ln f'] = 0$, the information equality. For part 2, expand the first-order condition around θ_0: $0 = \nabla\ell(\hat{\theta}) \approx \nabla\ell(\theta_0) + \nabla^2\ell(\theta_0)(\hat{\theta} - \theta_0)$, so $\sqrt{n}(\hat{\theta} - \theta_0) \approx [-\nabla^2\ell(\theta_0)/n]^{-1} \nabla\ell(\theta_0)/\sqrt{n}$. By the law of large numbers the bracket tends to $\mathcal{I}(\theta_0)$; by the central limit theorem the scaled sum of independent zero-mean scores tends to $\mathcal{N}(0, \mathcal{I}(\theta_0))$; the product therefore tends to $\mathcal{N}(0, \mathcal{I}^{-1}\mathcal{I}\,\mathcal{I}^{-1})$. Part 3, in outline: a second-order expansion of ℓ around $\hat{\theta}$ gives $\mathrm{LR} \approx n(\hat{\theta}-\tilde{\theta})'\mathcal{I}(\hat{\theta}-\tilde{\theta})$, a quadratic form in r asymptotically normal directions, hence χ^2_r. ∎

Two related tests use only one of the two fits: the **Wald test** uses the unrestricted estimate and its standard errors (for one coefficient, the familiar t-statistic), the **score test** only the restricted fit. All three agree in large samples. They need care when the restriction puts a parameter on the boundary of its range, as when testing that a random coefficient's standard deviation is zero (Chapter 6): the LR statistic then has a mixture distribution, and the χ^2 critical values are conservative.

WORKED EXAMPLE A.5 Show-up rates by segment

Over a month, 600 business bookings produced 528 show-ups and 1,400 leisure bookings produced 1,308 (illustrative). For binomial data the MLE of the show-up probability is the sample proportion, with standard error $\sqrt{\hat{q}(1-\hat{q})/n}$. Pooled, $\hat{q} = 1{,}836/2{,}000 = 0.918$ with standard error 0.006. Separately, $\hat{q} = 0.880$ for business and 0.934 for leisure. With $\ell_B(q)$ and $\ell_L(q)$ the log-likelihoods $x \ln q + (n-x)\ln(1-q)$ of the business and leisure samples, the likelihood ratio statistic for equal rates is

$$\mathrm{LR} = 2\left[\ell_B(0.880) + \ell_L(0.934) - \ell_B(0.918) - \ell_L(0.918)\right] = 15.4,$$

far above the 5 percent critical value of 3.84 for χ^2_1. The segments differ, and an overbooking model should treat them separately (Chapter 19).

The delta method Planning quantities are often functions of estimated parameters: a value of time is a ratio of logit coefficients, an elasticity a product of a coefficient and a fare, expected spill a nonlinear function of $\hat{\mu}$ and $\hat{\sigma}$.

RESULT A.8 Delta method

If $\sqrt{n}(\hat{\theta} - \theta) \to \mathcal{N}(0, V)$ and g is differentiable with a nonzero gradient at θ, then

$$\sqrt{n}(g(\hat{\theta}) - g(\theta)) \to \mathcal{N}(0,\ \nabla g(\theta)'\, V\, \nabla g(\theta)).$$

In practice, the standard error of $g(\hat{\theta})$ is $\sqrt{\nabla g(\hat{\theta})'\, \widehat{V}\, \nabla g(\hat{\theta})}$, where $\widehat{V}$ is the estimated covariance matrix of $\hat{\theta}$.

The proof is the first-order expansion $g(\hat{\theta}) \approx g(\theta) + \nabla g(\theta)'(\hat{\theta} - \theta)$, a linear function of an approximately normal vector. For example, a logit model with a time coefficient of -0.012 per minute (standard error 0.002), a fare coefficient of -0.0040 per dollar (0.0006) and a correlation of 0.3 between the two estimates implies a value of time of $g = 60\,\beta_t/\beta_f$, or CAD 180 per hour. Its gradient is $(60/\beta_f,\ -60\beta_t/\beta_f^2)$, or $(-15{,}000,\ 45{,}000)$, and

$$\mathrm{se}(g) = \sqrt{900 + 729 - 486} = 33.8,$$

where the three terms come from the time variance, the fare variance and their covariance. The 95 percent interval is CAD 114 to 246. Simulating the two coefficients from their joint normal distribution gives a similar spread but an asymmetric interval, CAD 121 to 262, because a ratio is skewed when its denominator is imprecise; for such quantities a simulated or bootstrap interval is safer.

The bootstrap The bootstrap treats the sample as if it were the population (Efron and Tibshirani 1993). Draw B samples of the same size with replacement, recompute the estimator on each, and use the B estimates as its sampling distribution: their standard deviation estimates the standard error, and their 5th and 95th percentiles give a 90 percent percentile interval. It needs no derivatives, so it handles estimators that are awkward to differentiate, such as two-step control functions (Chapter 7) or a spill estimate computed from censored data. With dependent data, resample whole clusters (routes, weeks) or blocks of consecutive periods, never single observations.

WORKED EXAMPLE A.6 How sure is a spill estimate?

Chapter 9 unconstrains a season of 26 departures to Seville, eleven of them full, and estimates $\hat{\mu} = 172.7$, $\hat{\sigma} = 46.3$ and expected spill of 14.2 passengers per departure. Resampling the 26 departures 1,000 times and re-running the EM algorithm on each resample gives a standard error of 10.5 for $\hat{\mu}$, against 9.1 from the formula $\hat{\sigma}/\sqrt{n}$ that ignores censoring, and a 90 percent interval for expected spill from 6.3 to 28.1 passengers. The point estimate is the best single number, but a season of data pins spill down only to within a factor of two in either direction; a route review should say so.

A.8 Censoring and truncation

DEFINITION A.4 Censoring and truncation

An observation is **censored** when its value is known only to lie beyond a bound: the load of a full flight says only that demand reached capacity. A sample is **truncated** when observations beyond a bound are missing altogether: a data set that contains only the bookings an airline accepted, or only the markets it chose to serve.

With censoring from above at known bounds c_i, a normal model with mean $x_i'\beta$ and an indicator $\delta_i = 1$ for censored observations has the **Tobit** log-likelihood

$$\ell(\beta,\sigma) = \sum_{i:\,\delta_i=0} \left[\ln \varphi\!\left(\frac{y_i - x_i'\beta}{\sigma}\right) - \ln \sigma \right] + \sum_{i:\,\delta_i=1} \ln\left[1 - \Phi\!\left(\frac{c_i - x_i'\beta}{\sigma}\right) \right]. \tag{A.3}$$

Uncensored observations contribute their density and censored ones the probability of exceeding the bound. Without regressors this is the likelihood that the EM algorithm of Chapter 9 maximizes; with $x_i'\beta$ it lets demand depend on season, day of the week or fare. The classic Tobit model of econometrics is the same construction with censoring from below at zero (Tobin 1958).

A truncated sample needs a different correction: each observation's density is divided by the probability of being observed. If only values below c are recorded,

$$\ell(\beta,\sigma)=\sum_i\left[\ln\varphi\left(\frac{y_i-x_i'\beta}{\sigma}\right)-\ln\sigma-\ln\Phi\left(\frac{c-x_i'\beta}{\sigma}\right)\right].$$

Ignoring either problem biases the estimates. The mean of censored data is $\mu-\sigma L(z)$ rather than μ (Result A.3), and its standard deviation is too small as well. In a truncated sample, $\mathbb{E}[y\mid x, y<c]=x'\beta-\sigma\varphi(z)/\Phi(z)$ with $z=(c-x'\beta)/\sigma$ (Result A.4); the correction grows as $x'\beta$ approaches the bound, so least squares on the observed data estimates slopes that are too flat. In airline data both problems are everywhere: revenue management censors demand class by class long before a flight fills, and the markets an airline serves are the ones it expected to be good.

A.9 Prediction intervals

A **confidence interval** describes uncertainty about a parameter, such as mean Friday demand. A **prediction interval** describes uncertainty about a future observation, such as next Friday's demand, and must include both the parameter uncertainty and the observation's own variability.

For n independent normal observations with sample mean $\bar X$ and sample standard deviation s, the new observation satisfies $X_{n+1}-\bar X\sim\mathcal{N}(0,\sigma^2(1+1/n))$, independently of s, so $(X_{n+1}-\bar X)/(s\sqrt{1+1/n})$ has a Student t distribution with $n-1$ degrees of freedom. The $1-\alpha$ prediction interval is

$$\bar X\pm t_{n-1,\,1-\alpha/2}\,s\sqrt{1+1/n}. \tag{A.4}$$

In a regression the same reasoning gives $\hat y_0\pm t\,s\sqrt{1+x_0'(X'X)^{-1}x_0}$ for a new observation with regressors x_0. Three extensions matter in forecasting (Chapter 8). Forecast uncertainty grows with the horizon: for a random walk the h-step forecast error has standard deviation $\sigma\sqrt{h}$. When the error distribution is not normal, intervals can be built from the quantiles of past out-of-sample forecast errors, and their coverage checked on later data: an 80 percent interval should contain about 80 percent of outcomes (Hyndman and Athanasopoulos 2021). And for models in logarithms, exponentiating the endpoints of an interval for $\ln X$ gives a valid interval for X, because quantiles survive monotone transformations; means do

not (Appendix A.2). When downside exposure drives a decision, the handbook points to quantile or distributional forecasts (Network Planning Handbook 2026).

WORKED EXAMPLE A.7 Next Friday's demand

Twenty Fridays of a departure average 150 passengers with a standard deviation of 40. With $t_{19,\,0.95} = 1.729$, the 90 percent prediction interval for next Friday is $150 \pm 1.729 \times 40 \times \sqrt{1.05}$, that is 150 ± 70.9, from 79 to 221 passengers. The 90 percent confidence interval for mean Friday demand is only $150 \pm 1.729 \times 40/\sqrt{20} = 150 \pm 15.5$. Parameter uncertainty is a small part of the uncertainty about a single departure; the variability of demand itself, the same variability that drives spill, is most of it.

A.10 Summary of distributions

Table A.2 The distributions of this appendix, with the parametrizations used in the text.

Distribution	Mean	Variance	Where it appears
Binomial $\mathrm{Bin}(n, q)$	nq	$nq(1-q)$	Show-ups (Chapter 19)
Poisson $\mathrm{Poisson}(\lambda)$	λ	λ	Booking requests (Chapter 18)
Negative binomial	$a\theta$	$a\theta + a\theta^2$	Overdispersed demand
Normal $\mathcal{N}(\mu, \sigma^2)$	μ	σ^2	Leg demand (Chapter 9)
Gamma $\mathrm{Gamma}(a, \theta)$	$a\theta$	$a\theta^2$	Skewed demand
Gumbel, location m	$m + \gamma_{\mathrm{E}}$	$\pi^2/6$	Logit errors (Chapter 5)
Logistic	0	$\pi^2/3$	Utility differences

APPENDIX

B

Optimization Toolkit

This appendix collects the optimization results that the chapters rely on, with short derivations, so that a statement met in a chapter can be checked in one place. The order follows the book's needs: linear programs and their duals, which give the bid prices of Chapter 20 and the value of an aircraft in Chapter 14; sensitivity analysis, which says how far such prices can be trusted; integer programs and the methods that solve the fleet assignment and crew models of Chapters 14 and 15; network flows, the structure behind time-space networks and passenger flows; dynamic programming, the engine of seat inventory control in Chapter 18; and the two decomposition ideas that make airline-sized models tractable, Lagrangian relaxation and column generation. A last section turns to solvers and modelling languages. The standard references go much further: Bertsimas and Tsitsiklis (1997) for linear programming, Wolsey (1998) and Nemhauser and Wolsey (1988) for integer programming, Ahuja, Magnanti, and Orlin (1993) for network flows, Puterman (1994) for dynamic programming and Desrosiers and Lübbecke (2005) for column generation. Vectors are columns, $x^\top$ is the transpose of x, and inequalities between vectors hold component by component.

One small planning question runs through the appendix. Laurentian has spare A220-300 time on a summer day: 42 aircraft-hours beyond its existing flying. A Montreal–Toronto round trip takes 3 aircraft-hours, turns included, and contributes CAD 6,000; a Montreal–Halifax round trip takes 4 aircraft-hours and contributes CAD 10,000. Montreal can offer 13 more departure slots, and Halifax demand supports at most 6 more round trips. The numbers are illustrative. With x_1 Toronto and x_2 Halifax round trips, and contributions in CAD thousands, the question is the linear program

$$\begin{aligned} \max_{x_1, x_2} \; & 6x_1 + 10x_2 \\ \text{s.t. } & 3x_1 + 4x_2 \le 42 \quad \text{(aircraft-hours)}, \\ & x_1 + x_2 \le 13 \quad \text{(slots)}, \\ & x_2 \le 6 \quad \text{(Halifax demand)}, \\ & x_1,\ x_2 \ge 0. \end{aligned} \tag{B.1}$$

B.1 Linear programming

A **linear program** (LP) maximizes or minimizes a linear function of continuous variables subject to linear equations and inequalities. Every model in the book that allocates a divisible resource is one, and the integer models of Chapters 14 and 15 are solved through linear programs.

> **DEFINITION B.1 Standard form and basic solutions**
>
> A linear program is in **standard form** when it reads $\min\{c^\top x : Ax = b,\ x \ge 0\}$, with A an $m \times n$ matrix of rank m. A **basis** is a set of m linearly independent columns of A, collected in a matrix B. Setting the other, nonbasic, variables to zero and solving $Bx_B = b$ gives a **basic solution**; it is a **basic feasible solution** if $x_B \ge 0$.

Any linear program can be put in standard form. A maximization becomes the minimization of $-c^\top x$; an inequality $a^\top x \le \beta$ becomes the equation $a^\top x + s = \beta$ with a nonnegative **slack variable** s (a surplus is subtracted for $\ge$); and a variable free in sign is written as the difference of two nonnegative ones. The example acquires three slacks: unused aircraft-hours, unused slots and unserved Halifax demand.

The feasible set of a linear program is a **polyhedron**, an intersection of finitely many half-spaces. It is convex, and its corners are its **vertices**; algebraically, the vertices are exactly the basic feasible solutions (Bertsimas and Tsitsiklis 1997). Figure B.1 draws the example's polyhedron, a pentagon with vertices $(0,0)$, $(13,0)$, $(10,3)$, $(6,6)$ and $(0,6)$, together with lines along which the objective is constant. Pushing such a line outward until it last touches the region finds the optimum at a corner, and this is general.

RESULT B.1 Optima occur at vertices

If a linear program in standard form has a feasible solution and its objective is bounded below on the feasible set, then it has an optimal solution, and some basic feasible solution is optimal.

Proof. We show that for every feasible x there is a basic feasible solution x' with $c^\top x' \le c^\top x$; since there are finitely many bases, the best basic feasible solution is then optimal. Let $S = \{j : x_j > 0\}$. If the columns A_j, $j \in S$, are linearly independent, they can be completed to a basis and x is basic. Otherwise some $d \neq 0$ has $Ad = 0$ and $d_j = 0$ for $j \notin S$. Choose the sign of d so that $c^\top d \le 0$ and, if $c^\top d = 0$, so that d has a negative component. If $d \ge 0$, then $c^\top d < 0$ and $x + \theta d$ is feasible for every $\theta \ge 0$, so the objective is unbounded below, a contradiction. So d has a negative component; take the largest θ that keeps $x + \theta d \ge 0$. The new point is feasible, no worse, and has at least one more zero component. Repeating this at most n times yields a basic feasible solution. ∎

The **simplex method** (Dantzig 1963) exploits Result B.1 by walking from vertex to adjacent vertex, two vertices being adjacent when their bases differ in one column. At a basis B, each nonbasic variable x_j has a **reduced cost**

$$\bar{c}_j = c_j - c_B^\top B^{-1} A_j,$$

the rate at which the objective changes when x_j is raised from zero and the basic variables adjust to keep $Ax = b$. If no reduced cost is negative, no edge leaving the vertex improves the objective and the basis is optimal; Result B.2 shows that this local test is also global. Otherwise the method raises a variable with a negative reduced cost until a basic variable falls to zero (the **ratio test**) and exchanges the two. At a degenerate vertex, where a basic variable is already zero, the method can stall, and anti-cycling rules guarantee that it terminates. **Interior-point methods** cross the inside of the polyhedron instead and are often faster on very large models; solvers provide both.

In the example, start at the origin. A Halifax trip contributes more, so raise x_2 until the demand cap binds at $(0, 6)$, where $z = 60$. Then raise x_1 along the cap until the hours run out at $(6, 6)$, where $z = 96$. From $(6, 6)$ one edge leads back to $(0, 6)$ and the other runs along the hours line to $(10, 3)$, trading three Halifax trips for four Toronto trips at a loss of $3 \times 10 - 4 \times 6 = 6$. Neither improves, so Laurentian

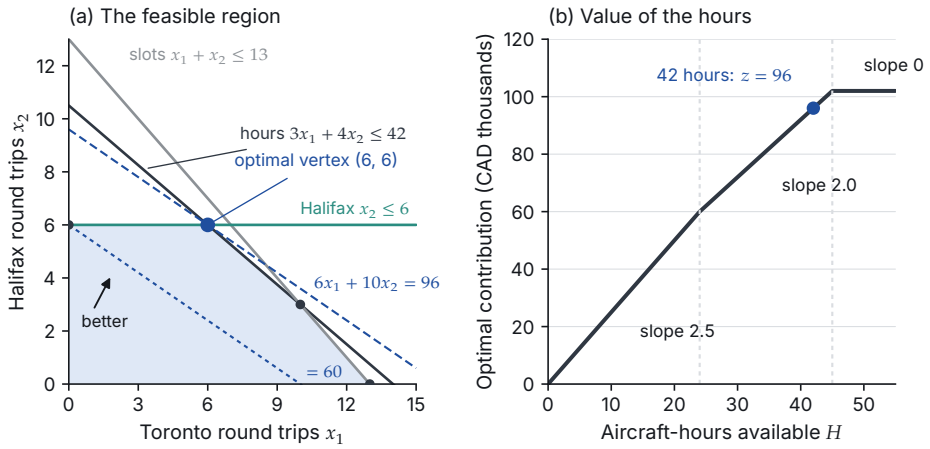


Figure B.1 (a) The feasible region of the example (B.1). Aircraft-hours, slots and Halifax demand bound it, and the dashed lines are objective contours; the optimum is the vertex $(6,6)$, worth CAD 96,000 a day. (b) The optimal contribution as the aircraft-hours vary. Its slopes, 2.5, 2.0 and 0 thousand CAD per hour, are the shadow price of an aircraft-hour over successive ranges.

should fly six Toronto and six Halifax round trips, for CAD 96,000 a day, using all 42 aircraft-hours and 12 of the 13 slots.

B.2 Duality

Every linear program has a twin built from the same data, its **dual**. The dual's variables are prices on the primal's constraints, and they are the shadow prices and bid prices used throughout the book.

The dual arises from the search for upper bounds. Take the primal in the form $\max\{c^\top x : Ax \le b,\ x \ge 0\}$, multiply constraint i by $y_i \ge 0$ and add the constraints: $y^\top Ax \le y^\top b$ for every feasible x. If the multipliers satisfy $A^\top y \ge c$, then, because $x \ge 0$,

$$c^\top x \le (A^\top y)^\top x = y^\top Ax \le b^\top y,$$

so every such y certifies the upper bound $b^\top y$. The dual looks for the best certificate:

$$\begin{aligned} &\text{(P)} \quad \max_x\ c^\top x \quad \text{s.t. } Ax \le b,\ \ x \ge 0, \\ &\text{(D)} \quad \min_y\ b^\top y \quad \text{s.t. } A^\top y \ge c,\ \ y \ge 0. \end{aligned} \tag{B.2}$$

Constraints of other kinds change only signs, as Table B.1 lists: an equation has a dual variable free in sign, and a variable free in sign has an equation as its dual constraint. The dual of the dual is the primal.

Table B.1 Rules for writing the dual of a maximization problem. Read from right to left, the table gives the dual of a minimization problem.

Primal (maximize)	Dual (minimize)
Constraint i is $\leq$	$y_i \geq 0$
Constraint i is $=$	y_i free
Constraint i is $\geq$	$y_i \leq 0$
$x_j \geq 0$	Dual constraint j is $\geq$
x_j free	Dual constraint j is $=$
$x_j \leq 0$	Dual constraint j is $\leq$

RESULT B.2 Duality theorems

For the pair (B.2):

1. Weak duality: if x and y are feasible, then $c^\top x \leq b^\top y$. Feasible solutions with equal values are both optimal.
2. Strong duality: if (P) has an optimal solution, so does (D), and the optimal values are equal. If (P) is unbounded, (D) is infeasible.
3. Complementary slackness: feasible x and y are both optimal if and only if $y_i\,(b_i - a_i^\top x) = 0$ for every constraint i and $x_j\,(A_j^\top y - c_j) = 0$ for every variable j, where $a_i^\top$ is row i and A_j column j of A.

Proof. Part 1 is the chain of inequalities above. For part 2, add slacks s to (P) and apply the simplex method, with an anti-cycling rule, to $\max\{c^\top x : Ax + s = b,\ x \geq 0,\ s \geq 0\}$. It ends at a basis B whose reduced costs are all non-positive, since the problem is a maximization (Bertsimas and Tsitsiklis 1997). Let $y^\top = c_B^\top B^{-1}$. For the columns of x, non-positive reduced costs say $c_j - y^\top A_j \leq 0$, that is $A^\top y \geq c$; the slack columns have zero cost and unit-vector columns, and for them the condition says $y \geq 0$. So y is dual feasible, and $b^\top y = c_B^\top B^{-1} b = c^\top x$; by part 1 both are optimal. If (P) is unbounded, a feasible y would bound it, so none exists. For part 3, the two gaps in the chain of inequalities are $\sum_j x_j (A_j^\top y - c_j)$ and $\sum_i y_i (b_i - a_i^\top x)$, sums of nonnegative terms. The values are equal exactly when every term is zero, and by parts 1 and 2 equal values are equivalent to joint optimality. ∎

Complementary slackness is the economics of the dual. Read y_i as a price on a unit of resource i. Dual feasibility says that no activity earns more than the value, at these prices, of the resources it

uses. Complementary slackness says that every activity in use earns exactly that value, making no profit beyond the prices, and that a resource left over has a price of zero. Strong duality says that the total contribution equals the value of all resources at these prices. When the optimal solution is nondegenerate, y_i is also the rate at which the optimal value grows with b_i: the **shadow price** of resource i, the most it is worth paying for one more unit (Appendix B.3).

WORKED EXAMPLE B.1 Pricing aircraft-hours

The dual of (B.1) prices aircraft-hours (y_1), slots (y_2) and Halifax demand (y_3):

$$\begin{aligned} &\min\ 42y_1 + 13y_2 + 6y_3 \\ &\text{s.t. } 3y_1 + y_2 \geq 6, \quad 4y_1 + y_2 + y_3 \geq 10, \quad y \geq 0. \end{aligned}$$

Complementary slackness pins it down. One slot is idle, so $y_2 = 0$. Both routes are flown, so both dual constraints hold with equality: $3y_1 = 6$ and $4y_1 + y_3 = 10$, giving $y_1 = 2$ and $y_3 = 2$. The dual value $42 \times 2 + 6 \times 2 = 96$ equals the primal value, which proves that $(6, 6)$ is optimal without examining any other vertex. An aircraft-hour is worth CAD 2,000 a day, a third of a Toronto trip's contribution. A unit of Halifax demand, enough for one more round trip, is worth CAD 2,000 too: a seventh Halifax trip contributes 10 but takes 4 aircraft-hours from Toronto flying worth $4 \times 2 = 8$. A slot is worth nothing while one is idle. For the planner, the dual prices are the exchange rates between resources implied by the optimal plan: any new use of the A220's spare time must earn more than CAD 2,000 an hour to deserve it.

WORKED EXAMPLE B.2 Bid prices as dual prices

Take the A–B–C network of Result 14.4: 80 local A–B passengers at USD 200, 90 local B–C passengers at USD 250 and 50 A–C connecting passengers at USD 400, historical teaching numbers from Barnhart's lectures (Barnhart 2006), now with 100 seats on each leg. The deterministic LP (20.3) of Chapter 20 carries all 80 A–B locals, 80 of the 90 B–C locals and 20 connections, for USD 44,000. Complementary slackness gives its dual. B–C locals and connections are carried below their demand, so their

rents (the dual prices β of the demand bounds) are zero and their fares equal the prices of the seats they use: $\pi_{BC} = 250$ and $\pi_{AB} + \pi_{BC} = 400$, so $\pi_{AB} = 150$. A–B locals are carried in full, and their rent is $\beta_{AB} = 200 - 150 = 50$. The dual value $100 \times 150 + 100 \times 250 + 80 \times 50 = 44{,}000$ matches the revenue.

Neither seat price is a fare. A seat on B–C is worth 250 because it would carry one more B–C local. A seat on A–B is worth 150 because it would carry one more connection, at 400, only by taking a B–C seat from a local who pays 250. These are the bid prices of Chapter 20: a request is accepted when its fare is at least the sum of the prices of the seats it uses. The same reading gives the value of one more aircraft, the dual price of the count constraint in fleet assignment when that price is unique (Example 14.4), and the marginal crew cost of a flight, the dual price of its cover constraint in crew pairing (Chapter 15).

B.3 Sensitivity analysis

How far can the data move before the optimal plan changes, and how does the optimal value respond? Sensitivity analysis answers from the final basis, without solving again.

Right-hand sides Changing b changes the basic solution $x_B = B^{-1}b$ but not the reduced costs, so the basis remains optimal as long as $B^{-1}b \geq 0$. Over that range the optimal value is linear in b, with the dual prices as slopes. At the end of the range a basic variable reaches zero, the basis changes and so do the slopes. Because the dual feasible set does not involve b, the optimal value of a maximization, $z^*(b) = \min\{b^\top y : A^\top y \geq c,\ y \geq 0\}$, is a minimum of linear functions of b, hence concave and piecewise linear: each additional unit of a resource is worth no more than the one before. Result 20.2 states the same property for seats.

Panel (b) of Figure B.1 traces the example's optimal contribution as the aircraft-hours H vary. Below 24 hours every hour goes to Halifax, at $10/4 = 2.5$ thousand CAD per hour, because Halifax earns more per hour than Toronto ($6/3 = 2.0$). At 24 hours Halifax reaches its cap of 6 trips, and from 24 to 45 hours each additional hour flies Toronto at 2.0, the dual price y_1 found above. At 45 hours, seven Toronto and six Halifax trips fill the 13 slots, and further hours are worth nothing. At a kink the shadow price is not unique: at 24 hours a removed hour is worth 2.5 and an added one 2.0, and a solver reports one of them.

Objective coefficients Changing the contribution c_j of a nonbasic variable changes only its own reduced cost; changing that of a basic variable changes all of them, through c_B. Either way the basis, and with it the plan, stays optimal while every reduced cost keeps its sign. In the example the plan $(6, 6)$ stays optimal while the Toronto contribution c_1 lies between 0 and 7.5 thousand CAD. At 7.5 a Toronto trip earns 2.5 per aircraft-hour, as much as a Halifax trip, and above it the optimum jumps to $(10, 3)$, flying Toronto until the slots bind.

> **PITFALL**
>
> A dual price is a derivative and holds only within its range. Ten more aircraft-hours at CAD 2,000 each suggest CAD 20,000 more contribution a day; the example gains CAD 6,000, because the slots bind after three hours. Removing hours costs more than the price suggests once fewer than 24 remain. For large changes, solve again, as Chapters 20 and 22 insist for bid prices. An integer program has no dual in this sense: the dual values that a solver reports after branch and bound come from a single linear program, usually the last relaxation or the one with the integer variables fixed, and need not price anything correctly.

B.4 Integer programming

An **integer program** (IP) is a linear program whose variables must take integer values; a **mixed-integer program** (MIP) has both integer and continuous variables. Integers model indivisible things, such as aircraft, frequencies and crew pairings, and binary variables model yes-or-no decisions.

DEFINITION B.2 Modelling with binary variables

Let x, y and z be binary decisions and w a continuous one with $0 \le w \le M$.

- Choice: $\sum_k x_k = 1$ picks exactly one option, such as the fleet type of a flight (Chapter 14); $\sum_k x_k \le 1$ picks at most one.
- Implication: $x \le y$ allows x only if y, such as a connecting service only if both of its legs operate.
- Fixed charge: the cost $Fy + c\,w$ with $w \le My$ charges F whenever $w > 0$, such as the cost of opening a station before any flight serves it.
- Either-or: $g(w) \le Mz$ and $h(w) \le M(1 - z)$ impose at least one of $g(w) \le 0$ and $h(w) \le 0$, for a large enough M.
- Partitioning and covering: $\sum_p a_{fp} x_p = 1$ (or ≥ 1) for each flight f, with $a_{fp} = 1$ if column p covers f, chooses crew pairings or aircraft routes (Chapter 15).

Each constant M should be the smallest valid bound: a large M gives a weak relaxation.

Dropping integrality gives the **LP relaxation**. Its feasible set contains every integer solution, so its optimal value bounds the integer optimum, from above for a maximization. The difference is the **integrality gap**, and the bound is what allows an algorithm to prove that an integer solution is optimal, or within a stated percentage of optimal. Rounding the relaxation's solution is tempting and unreliable: the rounded point may be infeasible, and when it is feasible it may be far from optimal.

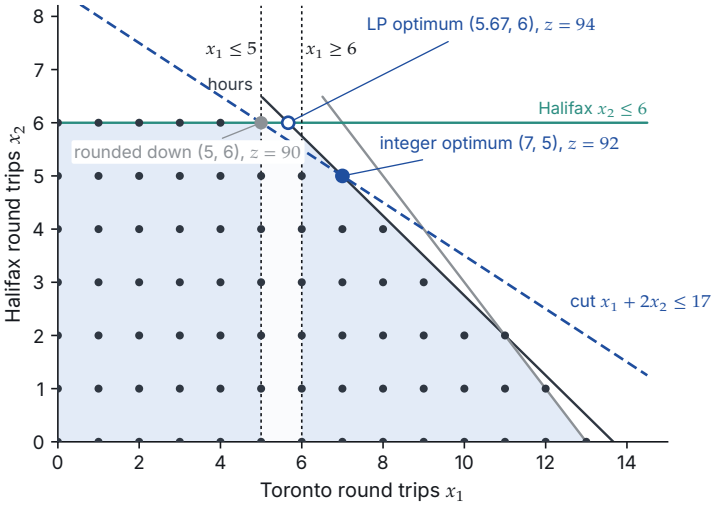


Figure B.2 The integer version of the example, with 41 aircraft-hours. The shaded region is the LP relaxation and the dots are its integer points; the grey line is the slot limit. The relaxation's optimum $(5.67, 6)$ is fractional, rounding it down gives $(5, 6)$, and the integer optimum $(7, 5)$ lies further along the hours line. The dotted lines show the first branching of Figure B.3, which removes the strip $5 < x_1 < 6$; the dashed line is the cut $x_1 + 2x_2 \le 17$, which removes the fractional optimum without removing any integer point.

WORKED EXAMPLE B.3 **Integer trips**

Trips are indivisible, so require x_1 and x_2 to be integers, and cut the aircraft-hours to 41. The relaxation's optimum moves to $(5.67, 6)$ with $z = 94$. Rounding x_1 down gives $(5, 6)$, feasible with $z = 90$. Listing the 72 integer points of the region shows that the integer optimum is $(7, 5)$ with $z = 92$: seven Toronto trips and five Halifax trips, which no rounding of $(5.67, 6)$ produces. The relaxation's value of 94 bounds every integer plan, so it shows that the rounded plan is at most 4 short of the optimum before the optimum is known. A planner who rounds an LP obtains a feasible plan and a bound, not an optimum.

Branch and bound (Land and Doig 1960) organizes the search. Solve the relaxation; if its solution is integral, it is optimal. Otherwise choose a variable x_j with a fractional value v and split the problem in two, adding $x_j \le \lfloor v \rfloor$ to one copy and $x_j \ge \lceil v \rceil$ to the other. The fractional point is in neither copy, and every integer point is in one. Each new problem is solved as a linear program, and the pro-

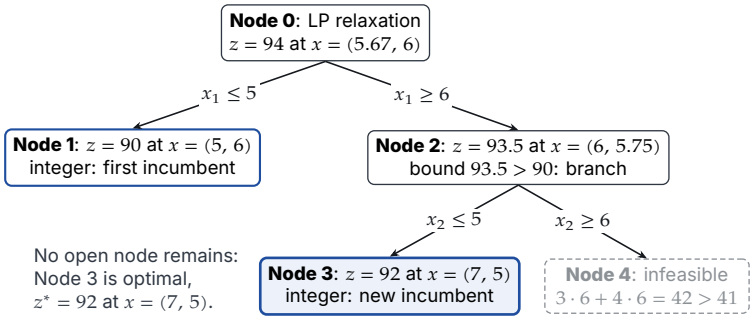


Figure B.3 Branch and bound on the integer example. Each box gives the value and solution of the relaxation at a node; branching on x_1 and then on x_2 closes every node after five linear programs.

cess repeats, building a tree. A node is closed when its relaxation is infeasible; when its solution is integral, in which case it becomes the **incumbent** if it beats the best solution so far; or when its bound is no better than the incumbent (**pruning by bound**). At any moment the incumbent and the best bound among the open nodes bracket the optimum, and their relative difference is the **gap** that solvers report.

In the example (Figure B.3), branching on $x_1 = 5.67$ creates two nodes. With $x_1 \le 5$ the relaxation's optimum is $(5, 6)$, integral with $z = 90$: the first incumbent. With $x_1 \ge 6$ it is $(6, 5.75)$, with a bound of 93.5 that exceeds 90, so the node is branched on x_2. With $x_2 \le 5$ the optimum $(7, 5)$ is integral with $z = 92$, a new incumbent; with $x_2 \ge 6$ the problem is infeasible, since six trips of each kind need 42 hours. No open node remains, so $(7, 5)$ is optimal: five linear programs prove what listing 72 points proved. In large problems the choice of branching variable and of the next node matters a great deal. Exploring deep nodes first finds incumbents quickly, exploring the best bound first keeps the tree small, and branching on groups of variables balances the tree (Section 14.5 discusses fleet assignment).

Cutting planes tighten the relaxation instead of splitting it. A **valid inequality** is satisfied by every integer solution, and a **cut** is a valid inequality that the relaxation's current optimum violates. Chvátal–Gomory rounding builds cuts from the constraints themselves (Gomory 1958; Chvátal 1973). For multipliers $u \ge 0$ on the constraints $Ax \le b$, and $x \ge 0$ integer, the inequality

$$\sum_j \lfloor u^\top A_j \rfloor x_j \le \lfloor u^\top b \rfloor$$

is valid. Its left side is at most $u^\top Ax \le u^\top b$, because rounding down the coefficient of a nonnegative variable cannot increase the sum,

and it is an integer, so it is at most the integer part of $u^{\top} b$. In the example, multipliers $1/3$ on the hours constraint and $2/3$ on the Halifax cap give $x_1 + 2x_2 \le 41/3 + 4 = 17.67$, hence the cut $x_1 + 2x_2 \le 17$. The relaxation's optimum violates it ($5.67 + 12 = 17.67$), and once the cut is added the relaxation's optimum is $(7, 5)$: a single cut closes the gap. Solvers combine the two ideas in **branch and cut**, adding general and problem-specific cuts at the nodes of the tree (Wolsey 1998).

Some integer programs need neither branching nor cuts, because their relaxations have only integral vertices.

DEFINITION B.3 Total unimodularity

A matrix is **totally unimodular** (TU) if every square submatrix has determinant 0, 1 or -1. In particular, every entry is 0, 1 or -1.

RESULT B.3 Integral polyhedra

1. If A is totally unimodular and b is integral, every vertex of $\{x : Ax \leq b,\ x \geq 0\}$ is integral (Hoffman and Kruskal 1956).
2. The node-arc incidence matrix of a directed graph, with one row per node, one column per arc, $+1$ in the row of the arc's tail and -1 in the row of its head, is totally unimodular.

1. Stacking $-I$ below A, for the constraints $-x \leq 0$, preserves total unimodularity: expanding the determinant of a square submatrix along its rows from $-I$, each with at most one nonzero entry, leaves either zero or, up to sign, the determinant of a square submatrix of A. A vertex is the unique solution of n linearly independent constraints that hold with equality, $Mx = d$, where M is a nonsingular square submatrix of the stacked matrix and d is integral. By Cramer's rule $x_j = \det M_j / \det M$, where M_j is M with column j replaced by d. Since $\det M = \pm 1$ and $\det M_j$ is an integer, x_j is an integer.
2. By induction on the size of a square submatrix M; a single entry is 0 or ± 1. If a column of M is zero, $\det M = 0$. If a column has exactly one nonzero entry, expanding along it gives $\det M = \pm \det M'$ for a smaller square submatrix M', which is 0 or ± 1 by induction. Otherwise every column of M contains both its $+1$ and its -1, the rows of M sum to zero, and $\det M = 0$.

The same argument covers equations and upper bounds, since stacking $-A$ or I also preserves total unimodularity. A network flow problem with integer supplies and capacities (Appendix B.5) therefore has integral optimal vertices, and the simplex method, which returns a vertex, finds integer flows without any branching. This is why one fleet type's time-space network solves as a linear program, and why the relaxation of the fleet assignment model is often nearly integral: only the cover constraints that tie the fleet types together, the count constraints that cap them and side constraints such as through flights create fractional vertices (Section 14.5 and Example 14.3). A matrix with consecutive ones in each column, an **interval matrix**, is also totally unimodular; the example's constraints $x_1 + x_2 \leq 13$ and $x_2 \leq 6$ form one, a fact used in Appendix B.7.

B.5 Network flows

A network is a directed graph with node set N and arc set A. Many airline models are flows on networks: aircraft on time-space networks (Chapter 14), crews on networks of duties (Chapter 15) and passengers on flight legs (Chapter 20).

DEFINITION B.4 Minimum-cost flow

Each arc a has a unit cost c_a and a capacity u_a, and each node v a net supply b_v, positive at sources and negative at sinks, with $\sum_v b_v = 0$. Writing $\delta^+(v)$ and $\delta^-(v)$ for the arcs leaving and entering v, the **minimum-cost flow problem** is

$$\begin{aligned} \min_x \ & \textstyle\sum_{a\in A} c_a x_a \\ \text{s.t.} \ & \textstyle\sum_{a\in\delta^+(v)} x_a - \sum_{a\in\delta^-(v)} x_a = b_v, \quad v \in N, \\ & 0 \le x_a \le u_a, \quad a \in A. \end{aligned}$$

The balance constraints' matrix is the node-arc incidence matrix, so by Result B.3 integer supplies and capacities give integer optimal flows. Familiar problems are special cases. The **shortest path** problem sends one unit from an origin to a destination without capacities; the **transportation** and **assignment** problems have only sources and sinks; and a **circulation** has every $b_v = 0$. A fleet's time-space network in Chapter 14 carries a circulation: flight arcs carry at most one aircraft, ground arcs any number, and the flow across the count line is the number of aircraft in use.

Shortest paths Let $d(v)$ be the length of a shortest path from the origin s to node v. The lengths satisfy **Bellman's equations**

$$d(s) = 0, \qquad d(v) = \min_{(w,v)\in A} \{d(w) + c_{wv}\},$$

which say that a shortest path to v is a shortest path to some predecessor w followed by the arc from w to v. In an **acyclic** network, such as a time-space or connection network in which every arc points forward in time, processing the nodes in time order solves the equations in one pass over the arcs, whatever the signs of the costs. With cycles and nonnegative costs, Dijkstra's algorithm labels the nodes in order of distance (Dijkstra 1959); with negative costs but no negative cycle, the Bellman–Ford method repeats the update over all arcs at most $|N| - 1$ times (Ahuja, Magnanti, and Orlin 1993). The linear programming dual of the shortest path problem maximizes

$d(t) - d(s)$ subject to $d(j) - d(i) \le c_{ij}$ on every arc, the triangle inequality, and the shortest distances solve it. The pricing problems of column generation (Appendix B.8) are shortest path problems whose arc costs include dual prices and are often negative. When a path must also respect limits on resources, such as duty time or the number of landings in a crew pairing, each node keeps several labels, one for each combination of cost and resource use that no other label beats on every count: the **resource-constrained shortest path** method of Chapter 15 (Irnich and Desaulniers 2005).

Multicommodity flows In a **multicommodity flow** problem several commodities share the arcs' capacities. Commodity $k \in K$ is a quantity q_k to be moved from an origin to a destination. The **node-arc formulation** has a flow variable x_a^k for every commodity and arc, the balance constraints of Definition B.4 for each commodity separately, and a **bundle constraint** $\sum_k x_a^k \le u_a$ on each arc. The **path formulation** lists for each commodity its set P_k of origin-destination paths and lets f_p be the fraction of commodity k sent along path $p \in P_k$:

$$\begin{aligned} \min_{f} \ & \textstyle\sum_k \sum_{p \in P_k} c_p f_p \\ \text{s.t.} \ & \textstyle\sum_k \sum_{p \in P_k} q_k \, \delta_{ap} f_p \le u_a, \quad a \in A, \\ & \textstyle\sum_{p \in P_k} f_p = 1, \quad k \in K, \qquad f \ge 0, \end{aligned} \tag{B.3}$$

where $\delta_{ap} = 1$ if path p uses arc a, and $c_p = q_k \sum_{a \in p} c_a$. The formulations are equivalent, since any feasible flow of one commodity splits into flows along paths, but their sizes differ enormously. The node-arc form has $|N||K| + |A|$ rows and $|A||K|$ columns; the path form has only $|A| + |K|$ rows but one column per path, far too many to list. A historical example in Barnhart's lectures had 807 nodes, 1,363 arcs and 17,539 commodities: 14,155,336 rows and 23,905,657 columns in node-arc form against 18,902 rows in path form. On an IBM RS6000 workstation the path model took 44 minutes to solve, and a formulation that grouped the commodities by their 136 origins, with 1,499 rows, took under a minute (Barnhart 2003).

Unlike a single-commodity flow, the multicommodity constraint matrix is not totally unimodular, so the relaxation can be fractional, and requiring each commodity to use a single path makes the problem hard. The passenger mix model of Chapter 14 is a path-based multicommodity flow whose commodities are itineraries and whose arcs are flight legs, and the deterministic LP of Chapter 20 has the same structure.

B.6 Dynamic programming

Dynamic programming solves sequential decision problems by working backward from the end. It rests on Bellman's **principle of optimality**: whatever the first decision, the remaining decisions of an optimal policy are optimal for the state that the first decision leads to (Bellman 1957). Count stages by the number t remaining, and let x be the state, $a \in A_t(x)$ a decision, $r_t(x,a)$ the immediate reward and X' the random next state. The value functions satisfy the **Bellman equation**

$$V_t(x) = \max_{a \in A_t(x)} \left\{ r_t(x,a) + \mathbb{E}\left[V_{t-1}(X') \mid x, a \right] \right\}, \tag{B.4}$$

with V_0 the terminal value. **Backward induction** computes $V_1, V_2, \ldots$ in turn, and the maximizing decisions form an optimal policy (Puterman 1994). The work grows with the number of states. That is why the method suits a single flight leg, whose state is the number of seats left (Chapter 18), and not a whole network, whose state is the vector of seats left on every leg (Chapter 20).

WORKED EXAMPLE B.4 Three seats, four periods

A Laurentian flight has 3 seats left and 4 booking periods to go, each short enough that at most one request arrives: a high-fare request at CAD 400 with probability 0.3, a low-fare request at CAD 150 with probability 0.5, and none with probability 0.2. Unsold seats are worth nothing at departure, so $V_0(x) = 0$, and $V_t(0) = 0$. Accepting a request with fare f is worth $f + V_{t-1}(x-1)$ and rejecting it $V_{t-1}(x)$, so the request is accepted when $f \geq \Delta V_{t-1}(x) = V_{t-1}(x) - V_{t-1}(x-1)$, the marginal value of the x-th seat. Hence

$$V_t(x) = V_{t-1}(x) + 0.3\,\left[400 - \Delta V_{t-1}(x)\right]^+ + 0.5\,\left[150 - \Delta V_{t-1}(x)\right]^+ .$$

Table B.2 Backward induction: expected revenue (CAD) with t periods to go and x seats left, and the resulting rule for the low fare when $t+1$ periods are left.

t	$V_t(1)$	$V_t(2)$	$V_t(3)$	Low fare at $t+1$
0	0	0	0	accept if $x \geq 1$
1	195.00	195.00	195.00	accept if $x \geq 2$
2	256.50	390.00	390.00	accept if $x \geq 2$
3	299.55	478.20	585.00	accept if $x = 3$

The high fare is always accepted, since no seat is ever worth more than CAD 400. The low fare is accepted in the last period whatever the seats left, with two or three periods to go only if at least two seats remain, and with four periods to go only if all three do: $\Delta V_3(3) = 585.00 - 478.20 = 106.80$ is below 150, but $\Delta V_3(2) = 178.65$ is above. The optimal policy earns $V_4(3) =$ CAD 694.56 in expectation, against 680.16 for accepting every request and 476.76 for selling only high fares. The marginal values fall as the seats left increase and rise with the time to go, the structure behind nested protection levels; Chapter 18 analyzes this model, due to Lee and Hersh (1993), in full. The marginal value ΔV is the single-leg bid price, the opportunity cost of a seat.

B.7 Lagrangian relaxation

Many large integer programs would be easy without a few complicating constraints. **Lagrangian relaxation** moves those constraints into the objective with prices. Write the problem as $z = \max\{c^\top x : Ax \leq b,\ x \in X\}$, where $Ax \leq b$ are the complicating constraints and X is a set over which optimization is easy, for example the integer points of a polyhedron with network structure. For multipliers $\lambda \geq 0$ the **Lagrangian function** is

$$L(\lambda) = \max_{x \in X} \{c^\top x + \lambda^\top (b - Ax)\}, \tag{B.5}$$

and the **Lagrangian dual** is $z_{\mathrm{LD}} = \min_{\lambda \geq 0} L(\lambda)$.

RESULT B.4 Lagrangian bounds

1. $L(\lambda) \geq z$ for every $\lambda \geq 0$, and L is convex; it is piecewise linear when X is finite.
2. If X is the set of integer points of a polyhedron Q, and z_{LP} is the optimal value of $\max\{c^{\top}x : Ax \leq b,\ x \in Q\}$, then $z \leq z_{\mathrm{LD}} \leq z_{\mathrm{LP}}$.
3. If Q has only integral vertices (the **integrality property**), then $z_{\mathrm{LD}} = z_{\mathrm{LP}}$ (Geoffrion 1974).

1. If $Ax \leq b$ and $x \in X$, then $c^{\top}x \leq c^{\top}x + \lambda^{\top}(b - Ax) \leq L(\lambda)$, because $\lambda \geq 0$. And L is a maximum of functions linear in λ, one for each $x \in X$. 2. The first inequality follows from part 1. For the second, let $L_Q(\lambda)$ be the same maximum taken over Q; since $X \subseteq Q$, $L(\lambda) \leq L_Q(\lambda)$, and linear programming duality applied to the constraints $Ax \leq b$ alone gives $\min_{\lambda \geq 0} L_Q(\lambda) = z_{\mathrm{LP}}$. 3. A linear function's maximum over Q is attained at a vertex, which is in X, so $L = L_Q$.

When X has the integrality property the Lagrangian dual is no better than the linear relaxation, and its value lies elsewhere: each evaluation of L may be much faster than solving the LP, and the maximizers of L suggest feasible solutions. When X lacks the property, the Lagrangian bound can be strictly better than the LP bound (Fisher 1981).

Relax the hours constraint of the integer example, keeping $X = \{x \text{ integer} : x \geq 0,\ x_1 + x_2 \leq 13,\ x_2 \leq 6\}$. Then

$$L(\lambda) = 41\lambda + \max_{x \in X} \{(6 - 3\lambda)\, x_1 + (10 - 4\lambda)\, x_2\}.$$

For $\lambda \leq 2$ both coefficients are nonnegative, the maximizer is $(7, 6)$ and $L = 102 - 4\lambda$; for $2 \leq \lambda \leq 2.5$ only Halifax pays, the maximizer is $(0, 6)$ and $L = 60 + 17\lambda$; above 2.5 nothing pays and $L = 41\lambda$ (Figure B.4). The minimum, at $\lambda = 2$, is 94, the LP bound, as part 3 predicts, because X is defined by an interval matrix (Appendix B.4). The minimizing price is the relaxation's dual price of an aircraft-hour. The maximizer $(7, 6)$ needs 45 hours and is infeasible, but dropping one Halifax trip repairs it to $(7, 5)$, which the bound guarantees to be within 2 of the optimum; here it is the optimum.

Since L has kinks, it is minimized by the **subgradient method**. At λ_k, find a maximizer x_k. The vector $g_k = b - Ax_k$ is a subgradient of L at λ_k: for every λ, $L(\lambda) \geq c^{\top}x_k + \lambda^{\top}g_k = L(\lambda_k) + (\lambda - \lambda_k)^{\top}g_k$. Set

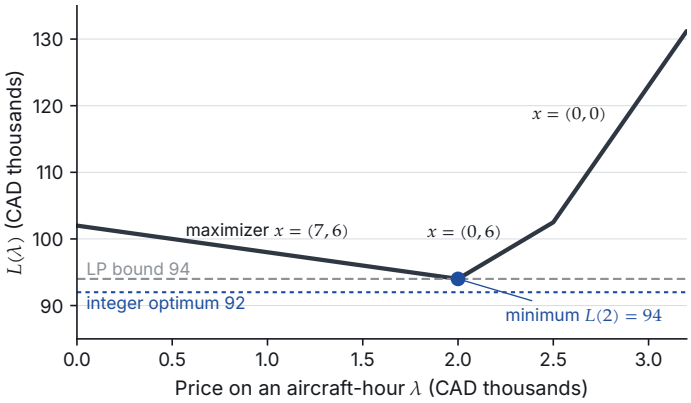


Figure B.4 The Lagrangian function of the integer example with the hours constraint relaxed. Each linear piece corresponds to one maximizer. The minimum, 94 at a price of CAD 2,000 per aircraft-hour, equals the LP bound and lies above the integer optimum of 92: the Lagrangian dual cannot close this gap because the remaining constraints have the integrality property.

$\lambda_{k+1} = \max\{0,\ \lambda_k - s_k g_k\}$ with a step $s_k > 0$. In the example, from $\lambda_0 = 0$ with steps of 0.5, the maximizer $(7, 6)$ gives $g_0 = 41 - 45 = -4$ and $\lambda_1 = 2$, the optimum. But at $\lambda = 2$ the maximizer is not unique; if the subproblem returns $(0, 6)$, then $g_1 = 17$, a step of 0.5 throws λ back to 0, and constant steps zigzag forever. Steps that shrink to zero while their sum diverges guarantee convergence (Fisher 1981).

In airline models the complicating constraints are usually those that tie otherwise separate problems together. Relaxing the cover constraints of the fleet assignment model, which require each flight to be flown by exactly one type, leaves one network flow problem per fleet type, each with its count constraint (Chapter 14). Relaxing the leg capacities of the deterministic LP of Chapter 20 leaves one problem per fare product, solved by comparing the fare with the sum of the prices on its legs. Column generation, next, reaches the same bound by another route.

B.8 Dantzig–Wolfe decomposition and column generation

Consider a linear program $\min\{c^\top x : Ax \geq b,\ x \in Q\}$ in which the **linking constraints** $Ax \geq b$ tie together blocks that Q keeps separate: $x = (x_1, \ldots, x_K)$ and $Q = Q_1 \times \cdots \times Q_K$, each Q_k a bounded polyhedron. Let c_k and A_k be the parts of c and A that belong to block k. Every point of Q_k is a convex combination of its vertices x^p, $p \in P_k$, and

substituting $x_k = \sum_{p \in P_k} \lambda_p x^p$ gives the **master problem** (Dantzig and Wolfe 1960)

$$\begin{aligned} \min_{\lambda \geq 0} \ & \textstyle\sum_k \sum_{p \in P_k} (c_k^\top x^p)\, \lambda_p \\ \text{s.t.} \ & \textstyle\sum_k \sum_{p \in P_k} (A_k x^p)\, \lambda_p \geq b, \\ & \textstyle\sum_{p \in P_k} \lambda_p = 1, \quad k = 1, \dots, K. \end{aligned} \tag{B.6}$$

It has few rows, the linking constraints and one **convexity constraint** per block, and an enormous number of columns, one per vertex. The path formulation (B.3) is a master problem of this kind, with the commodities as blocks and their paths as columns, and so is crew pairing (Chapter 15), whose columns are pairings.

Column generation solves the master problem without listing its columns (Gilmore and Gomory 1961; Desrosiers and Lübbecke 2005). Solve a **restricted master problem** (RMP) over a subset of the columns, and read its dual prices: $\pi \geq 0$ on the linking constraints and σ_k on the convexity constraints. A column p of block k can improve the RMP only if its **reduced cost**

$$\bar{c}_p = c_k^\top x^p - \pi^\top A_k x^p - \sigma_k \tag{B.7}$$

is negative. Finding the most negative one is the **pricing problem**, $\min_{x \in Q_k} (c_k - A_k^\top \pi)^\top x - \sigma_k$ for each block: an optimization over the easy set with costs adjusted by the dual prices. For the path formulation it is a shortest path problem per commodity; for crew pairing, a resource-constrained shortest path. Columns with negative reduced costs join the RMP, and the loop repeats until none remains.

RESULT B.5 Optimality test and bound

Let z_{RMP} be the optimal value of the restricted master problem, $\bar{c}_p$ the reduced costs at its dual prices, and z_{M} the optimal value of the full master problem. Then

$$z_{\text{RMP}} + \sum_k \min_{p \in P_k} \bar{c}_p \ \leq \ z_{\text{M}} \ \leq \ z_{\text{RMP}}.$$

In particular, if no column has a negative reduced cost, the RMP's solution is optimal for the master problem.

Proof. The upper bound holds because the RMP is the master problem with columns removed. For the lower bound, take any feasible λ

of the master problem and write each cost as $c_k^\top x^p = \bar{c}_p + \pi^\top A_k x^p + \sigma_k$. Summing over all columns,

$$\sum_{k,p} (c_k^\top x^p)\lambda_p = \sum_{k,p} \bar{c}_p \lambda_p + \pi^\top \sum_{k,p} (A_k x^p)\lambda_p + \sum_k \sigma_k .$$

The first sum is at least $\sum_k \min_p \bar{c}_p$, because the weights of each block are nonnegative and sum to one, and the second term is at least $\pi^\top b$, because $\pi \geq 0$ and the linking constraints hold. By strong duality for the RMP, $\pi^\top b + \sum_k \sigma_k = z_{\text{RMP}}$. If no reduced cost is negative, the lower bound is at least z_{RMP} and the two bounds meet. ∎

The lower bound is the Lagrangian function of the linking constraints evaluated at the prices π, so column generation and Lagrangian relaxation compute the same bound by different routes (Lübbecke and Desrosiers 2005). Result 15.2 in Chapter 15 gives the version used for crew pairing.

WORKED EXAMPLE B.5 Column generation on a small network

Barnhart's lectures introduce multicommodity flows with a five-arc network (Barnhart 2003). Arcs a from node 1 to 2, b from 1 to 3, c from 2 to 3, d from 2 to 4 and e from 3 to 4 have unit costs 1, 2, 3, 4 and 5 and capacities 20, 10, 20, 10 and 40. Four commodities must move: 5 units from node 1 to 3, 15 from 1 to 4, 5 from 2 to 4 and 10 from 3 to 4. With prices $\pi_a \geq 0$ on the arc capacities, the reduced cost of a path p of commodity k in (B.3) is $\bar{c}_p = q_k \sum_{a \in p} (c_a + \pi_a) - \sigma_k$, so pricing is a shortest path problem with arc lengths $c_a + \pi_a$.

Start the RMP with one path per commodity: b, a-c-e, d and e. Path a-c-e is the only one with room for all 15 units of the second commodity. The RMP costs $10 + 135 + 20 + 50 = 215$, no arc is full, so every $\pi_a = 0$ and each σ_k equals the cost of its commodity's path. Pricing finds two better paths for the second commodity: a-d, of length 5 and reduced cost $15 \times 5 - 135 = -60$, and b-e, of length 7 and reduced cost -30. Result B.5 brackets the optimum between $215 - 60 = 155$ and 215.

With the two new columns the RMP splits the second commodity equally over a-c-e, a-d and b-e, at a cost of 185. Arcs b and d are now full, with prices $\pi_{\text{b}} = 2$ and $\pi_{\text{d}} = 4$, and the convexity prices become 20, 135, 40 and 50. With lengths $c_a + \pi_a$, every path of every commodity has a reduced cost of exactly zero, so the bound meets the RMP's value and 185 is optimal.

If each commodity must travel on a single path, the second commodity fits only on a-c-e, because arcs b and d hold 10 units, and the best such solution costs 215, 16 percent above the linear optimum. Closing gaps of this kind calls for **branch and price**, branch and bound with column generation at every node, using branching rules that keep the pricing problem a shortest path problem (Barnhart, Johnson, et al. 1998; Ryan and Foster 1981). The planner's lesson is in the prices: a full arc becomes longer by its price, and the pricing step routes flow around it, much as bid prices steer connecting passengers away from full legs in Chapter 20.

B.9 Solvers and modelling languages

The models of this book are small enough for open-source software. SciPy (Virtanen et al. 2020) provides `linprog` for linear programs and `milp` for mixed-integer programs; both call the HiGHS solver (Huangfu and Hall 2018), which implements the simplex and interior-point methods and branch and cut. The examples of this appendix take a few lines:

```
import numpy as np
from scipy.optimize import linprog, milp
from scipy.optimize import LinearConstraint

c = np.array([6, 10])  # CAD thousands
# rows: hours, slots, Halifax demand
A = np.array([[3, 4], [1, 1], [0, 1]])
b = np.array([42, 13, 6])
lp = linprog(-c, A_ub=A, b_ub=b,
             method="highs")
print(lp.x, -lp.fun)          # [6. 6.] 96.0
print(-lp.ineqlin.marginals)  # [2. 0. 2.]

b[0] = 41                     # integer case
con = LinearConstraint(A, ub=b)
ip = milp(-c, constraints=con,
          integrality=[1, 1])
print(ip.x, -ip.fun)          # [7. 5.] 92.0
```

Two details trip up newcomers. SciPy minimizes, so a maximization is solved by negating c, and the reported marginals, the derivatives of the minimized objective with respect to the right-hand sides, must be negated as well to give the shadow prices of Example B.1. And `milp` reports no dual prices, in keeping with the pitfall of Appendix B.3.

Larger models are written in an algebraic **modelling language**, which keeps the model separate from the data and can pass it to different solvers. AMPL (Fourer, Gay, and Kernighan 2003), and the Python and Julia libraries Pyomo (Hart, Watson, and Woodruff 2011) and JuMP (Dunning, Huchette, and Lubin 2017), are examples; the problem sets of the MIT course on which Chapters 14 and 15 draw used ILOG's OPL Studio, a historical choice (Barnhart 2003). Commercial solvers such as CPLEX and Gurobi and open-source ones such as HiGHS combine the ideas of Appendix B.4, branch and bound and cutting planes, with preprocessing and heuristics (Wolsey 1998). Commercial airline planning systems embed such methods in their fleet assignment and crew modules (Chapter 14).

A few habits prevent most errors. Check the solver's status before reading a solution. Scale the data so that coefficients span few orders of magnitude, as the CAD-thousand units of the example do. Treat a value such as 0.9999998 as 1 only within the solver's integrality tolerance. Give large integer programs a time limit and a target gap, and report the gap with the solution. And to obtain prices from an integer solution, fix the integer variables at their optimal values and solve the remaining linear program.

B.10 Where the tools are used

Table B.3 Where the tools of this appendix are used in the book.

Tool	Where it appears
Linear programming and duality	Bid prices from the deterministic LP (Chapter 20); the passenger mix model (Chapter 14); displacement in route evaluation (Chapter 12)
Sensitivity analysis	The value of seats, aircraft and hours (Chapters 14 and 20)
Integer programming	Schedule design (Chapter 13); fleet assignment (Chapter 14); aircraft routing and crew pairing (Chapter 15)
Network flows	Time-space networks (Chapter 14); shortest path pricing (Chapter 15)
Multicommodity flows	Passenger mix and itinerary-based fleet assignment (Chapter 14)
Dynamic programming	Seat inventory control on a single flight (Chapter 18)
Lagrangian relaxation and column generation	Crew pairing (Chapter 15); bounds for large fleet assignment models (Chapter 14)

APPENDIX

C

Glossary

This glossary defines the terms and acronyms used in the book and in daily conversation in a network planning department. Definitions are deliberately short; the chapter given in brackets develops each idea. Where a term has a common trap, the entry names it.

A

A14 An on-time measure: the share of flights arriving no more than 14 minutes after their scheduled arrival time. Public statistics usually use the same 15-minute tolerance under other names. Flight punctuality is not passenger punctuality: a late feeder can strand connecting passengers on a flight that itself arrives on time (Chapter 16).

ACMI Aircraft, crew, maintenance and insurance supplied by one airline to another for a fee per block hour; also called a wet lease. The lessee sells the seats and pays fuel, airport and navigation charges. Contrast with a dry lease, which supplies the aircraft only (Chapter 11).

Advance purchase A fare rule requiring the ticket to be bought a minimum number of days before departure. It is the classic fence separating leisure travellers, who plan early, from business travellers, who often cannot (Chapter 17).

Aircraft movement One take-off or one landing at an airport. Airport statistics count movements; a single flight leg produces two of them, one at each end.

Aircraft routing The sequence of flight legs assigned to an individual aircraft (a tail), chosen so that every aircraft reaches a maintenance station at the required intervals. Also called maintenance routing (Chapter 15).

Alliance A multilateral cooperation among airlines covering codeshares, frequent flyer reciprocity, lounge access and coordinated connections. The three global alliances are Star Alliance, oneworld and SkyTeam. An alliance does not by itself permit

joint pricing; that requires a joint venture with antitrust approval (Chapter 10).

Ancillary revenue Revenue from products sold alongside the fare: bags, seat selection, upgrades, onboard sales, change fees. Check whether a revenue figure already includes it before adding it to a route case.

ASK, ASM Available seat kilometre and available seat mile: one seat, occupied or not, flown one kilometre or one statute mile. The basic unit of capacity. Never mix kilometres and miles in one ratio; Air Canada reports in miles (Chapter 2).

Authorization level (AU) The number of reservations an airline is willing to accept on a flight or cabin, set above physical capacity to allow for no-shows (Chapter 19).

Avoidable cost A cost that disappears if the decision under study is not taken, over the horizon of that decision. The relevant cost for incremental route and schedule decisions. Contrast with fully allocated cost (Chapter 3).

B

Bank A cluster of arrivals at a hub followed by a cluster of departures, timed so that arriving passengers can connect to many departing flights. A banked hub creates connections at the price of peaks in gate, runway and staff use; a rolling hub spreads flights evenly and creates fewer good connections (Chapter 13).

Bid price The opportunity value of one seat on a flight leg, used as a threshold: a request is accepted if its fare is at least the sum of the bid prices of the legs it uses. Bid prices come from the dual of a network optimization (Chapters 20 and 22).

Block time, block hour The time from the aircraft leaving its departure gate (off blocks) to arriving at its destination gate (on blocks). The main driver of crew, maintenance and ownership costs, and the unit in which many costs are budgeted (Chapter 3).

Booking class An inventory bucket identified by a letter (the reservations booking designator) to which fare products are mapped. Revenue management opens and closes booking classes, not individual fares (Chapter 18).

Booking curve Cumulative bookings for a future flight plotted against days before departure. Its shape differs by market and fare class and is the basis of pickup forecasting (Chapters 8 and 18).

Booking limit The maximum number of seats that may be sold in a booking class (and, under nesting, in all lower classes). The complement of a protection level (Chapter 18).

Branded fare A fare family sold under a name (for example a basic, a standard and a flexible economy product), each brand bundling a set of attributes such as bags, seat choice and change rules. Branded fares replace many of the restrictions of older fare structures with explicit product differences (Chapter 17).

Break-even load factor The load factor at which revenue equals cost, approximately unit cost divided by yield. It depends on the cost concept used and ignores the fact that yield falls as load factor rises (Chapter 2).

Buffer Extra time added to a scheduled block time or turn time beyond what is needed on average, to absorb delays. Buffers improve punctuality and reduce aircraft utilization (Chapter 16).

Buy-down A passenger buying a cheaper fare product than the one he or she would have bought if the cheaper one had been unavailable. Revenue management that ignores buy-down can trigger spiral down (Chapter 21).

C

CASK, CASM Operating cost per available seat kilometre or mile. Falls with stage length and aircraft size, so comparisons across routes or airlines need a stage-length adjustment. Many airlines also report adjusted unit costs that exclude fuel and items they consider unusual; check each company's definition (Chapter 3).

Codeshare An arrangement in which an airline (the marketing carrier) sells seats under its own flight number on a flight operated by another airline (the operating carrier). When counting flights from schedule data, remove the duplicate marketing records.

Column generation A method for linear programs with too many variables to list, such as crew pairing models: solve the problem with a subset of columns, then use the dual prices to search for a new column with negative reduced cost, and repeat (Chapter 15 and Appendix B).

Connection window The range of acceptable connecting times at a hub, from the minimum connecting time to a maximum beyond which the connection is unattractive or not offered (Chapter 13).

CORSIA The Carbon Offsetting and Reduction Scheme for International Aviation, adopted by ICAO, under which airlines offset growth in international aviation emissions above a baseline (Chapter 23).

Crew pairing A sequence of duty periods, starting and ending at the same crew base, that covers a set of flights and satisfies the legal and contractual rules. Pairings are later assembled into individual monthly schedules by rostering (Chapter 15).

Curfew A period, usually at night, during which an airport restricts or prohibits operations. Curfews shape which departure and arrival times a schedule can use (Chapter 13).

D

DAVN Displacement-adjusted virtual nesting: a network revenue management method that maps each origin-destination fare product into the value buckets of each leg it uses, after subtracting the displacement cost it imposes on the other legs of its itinerary (Chapter 20).

DB1B, DB1C The United States Department of Transportation's airline origin and destination survey. DB1B was a quarterly ten percent sample of tickets until mid-2025; its successor, DB1C, is a monthly sample of forty percent. The change is a methodological break: the two series cannot be spliced by rescaling (Network Planning Handbook 2026).

Deadhead A crew member travelling as a passenger to position for a later duty. Deadheads cost money and seats, and crew pairing models minimize them (Chapter 15).

Decision window The range of departure and arrival times a traveller considers acceptable, used in the Boeing path preference model and in the PODS simulator to decide which itineraries enter a traveller's choice set (Chapter 6).

Demand factor Unconstrained demand divided by capacity. Unlike the load factor, it can exceed one (Chapter 9).

Denied boarding Refusing to carry a passenger holding a confirmed reservation because more passengers showed up than there are seats. Voluntary denied boardings are passengers who accept compensation to give up their seat; involuntary ones are not given the choice (Chapter 19).

Displacement cost The revenue lost on other passengers when a seat is used for a given booking. On a full leg, a connecting passenger displaces a local passenger, so the value of the connecting booking to the network is its fare minus the displacement cost of the legs it uses (Chapters 12 and 20).

DLP The deterministic linear program of network revenue management: maximize fare revenue from seats allocated to origin-destination fare products, subject to leg capacities and to the

allocation not exceeding mean demand. Its dual prices are bid prices (Chapter 20).

Dry lease A lease of an aircraft without crew, maintenance or insurance (Chapter 11).

Duty period A working day of a crew member, from report time to release, containing one or more flights and limited by fatigue rules (Chapter 15).

E

EDTO, ETOPS Extended diversion time operations, the current ICAO term, and its predecessor, extended-range twin-engine operations: rules for flying routes that go beyond a threshold time from an adequate diversion airport. They require approval of the aircraft type, of the operator and of the specific operation (Chapter 11).

EMSR Expected marginal seat revenue: heuristics for setting nested protection levels for several fare classes on one flight. EMSR-a combines pairwise Littlewood protection levels; EMSR-b aggregates the higher classes into one artificial class (Chapter 18).

Endogeneity Correlation between an explanatory variable and the error term of a model. In airline data, fares are endogenous because revenue management raises them when demand is high, which biases naive estimates of price elasticity toward zero (Chapter 7).

EU ETS The European Union Emissions Trading System, which requires airlines to surrender emission allowances for covered flights. Its scope for flights to and from outside Europe has been restricted and must be checked for each planning year (Chapter 23).

F

Fare basis The code identifying a fare and its rules on a ticket. Several fare bases can share one booking class.

Fare product A fare level together with its rules (advance purchase, minimum stay, refundability, changes, bags) for a market. In network revenue management the unit of control is the origin-destination fare product (Chapter 17).

Fence A restriction that stops high willingness-to-pay customers from buying a low fare, such as advance purchase, a Saturday night stay or non-refundability. Fences make price discrimination possible (Chapter 17).

Fleet assignment The assignment of an aircraft type to each flight of a schedule so as to maximize contribution (revenue lost to spill counted as a cost) subject to aircraft availability and flow balance; solved with the fleet assignment model (FAM) on a time-space network (Chapter 14).

Freedoms of the air The rights a country may grant to foreign airlines. The first and second allow overflight and technical stops; the third and fourth allow carriage to and from the airline's home country; the fifth allows carriage between two foreign countries on a service that starts or ends at home. The sixth, carriage between two foreign countries through the home hub, is the basis of hub connecting traffic (Chapter 10).

Frequency share An airline's share of the departures in a market. Market share rises more than proportionally with frequency share, the S-curve effect (Chapter 4).

Fully allocated cost A cost per flight that includes a share of every company cost, including overheads that would not change if the flight were cancelled. Useful for long-run pricing and reporting, misleading for incremental decisions (Chapters 3 and 12).

G

Gauge The size of the aircraft on a route, measured in seats. Upgauging replaces an aircraft with a larger one; the frequency-gauge trade-off is the central schedule design choice (Chapter 13).

GDS A global distribution system (Amadeus, Sabre, Travelport) through which travel agents search for and book flights. GDS bookings are the main input of MIDT data and are not the whole market.

Go-show A passenger who arrives for a flight without a reservation, or with one for a different flight, and is accommodated (Chapter 19).

H

Hub premium The higher average fare an airline earns in markets touching its hub, explained by frequent flyer loyalty, schedule dominance, airport constraints and a passenger mix rich in business travel (Chapter 10).

Hub-and-spoke network A network in which most routes connect outlying airports (spokes) to one or a few central airports (hubs), so that passengers between two spokes connect at the hub.

With n spokes, one hub connects of the order of $n^2/2$ markets with n routes (Chapter 10).

I

IATA season The two scheduling seasons used worldwide: the northern summer season, which starts on the last Sunday of March, and the northern winter season, which starts on the last Sunday of October (Chapter 13).

IFAM The itinerary-based fleet assignment model, which evaluates passenger revenue by itinerary rather than by leg, so that spilling a connecting passenger counts on all the legs of the journey (Chapter 14).

IIA Independence from irrelevant alternatives: in a logit model, the ratio of the probabilities of two alternatives does not depend on the other alternatives. It implies that a new itinerary draws share from every existing one in proportion, which is unrealistic for near-identical itineraries (Chapter 5).

Interline An agreement that lets an airline sell a single ticket that includes another airline's flights, with baggage checked through and revenue prorated between the two (Chapter 10).

Itinerary The sequence of flight legs a passenger uses to travel in one direction from origin to destination. A nonstop itinerary uses one leg; a connection uses two or more.

J

Joint venture A deep partnership in which airlines share revenue or profit on a set of markets regardless of which partner operates the aircraft (metal neutrality), usually with antitrust immunity or regulatory approval to coordinate prices and capacity (Chapter 10).

K

k-factor The coefficient of variation of demand, $k = \sigma/\mu$. Empirical studies typically find values between 0.20 and 0.40 for flight demand (Chapter 9).

L

Leg A single nonstop operation of an aircraft between two airports, also called a flight leg or segment. Capacity is sold by leg; demand is generated by origin and destination.

Load factor Revenue passenger kilometres divided by available seat kilometres. For a single flight it is passengers divided by seats; for a system it must be computed as a ratio of sums, not an average of ratios (Chapter 2).

Logit The discrete choice model derived from independent, identically distributed extreme value errors, in which the probability of choosing alternative i is $e^{V_i} / \sum_j e^{V_j}$ (Chapter 5).

Logsum The expected maximum utility of a choice set, $\ln \sum_j e^{V_j}$ up to a constant. Divided by the marginal utility of money, it measures consumer surplus, which makes it the right tool to value a schedule improvement (Chapter 5).

M

Marketing carrier, operating carrier The airline that sells a seat under its flight number, and the airline that flies the aircraft. They differ on codeshare flights.

MCT Minimum connecting time: the shortest interval between an arrival and a departure at an airport for which a connection may be sold, set by airport, terminal and type of connection (Chapter 13).

MIDT Marketing Information Data Tapes: booking records collected from global distribution systems and sold as market intelligence. Coverage of direct and some new distribution channels is incomplete, so market totals built from MIDT are estimates (Network Planning Handbook 2026).

Mixed logit A logit model whose coefficients vary randomly across travellers, which allows taste heterogeneity and flexible substitution patterns. Choice probabilities are computed by simulation (Chapter 6).

MTOW, MZFW, MLW, OEW Maximum take-off weight, maximum zero-fuel weight, maximum landing weight and operating empty weight: the structural weights that, together with fuel capacity, define an aircraft's payload-range diagram (Chapter 11).

N

NDC New Distribution Capability: an IATA data standard that lets airlines send offers, including ancillaries and dynamic prices, directly to travel agents and other sellers through programming interfaces rather than through filed fares in a GDS (Chapter 23).

Nested logit A generalized extreme value model that groups similar alternatives into nests, so that alternatives in the same nest are

closer substitutes for one another than for alternatives in other nests (Chapter 6).

Nesting The inventory rule that a seat available to a lower booking class is also available to every higher class. Under serial nesting, protection levels and booking limits apply to cumulative groups of classes (Chapter 18).

Network contribution The change in the whole network's revenue minus avoidable cost caused by a decision, including effects on other flights through feed, displacement and recapture (Chapter 12).

No-show A passenger holding a reservation who does not turn up for the flight (Chapter 19).

O

O&D Origin and destination: the true start and end points of a passenger's journey, as opposed to the legs it uses. Demand is generated by O&D markets; capacity is supplied by legs.

ODF An origin-destination fare product: an itinerary combined with a fare class, the unit of control in network revenue management (Chapter 20).

Overbooking Accepting more reservations than seats to compensate for no-shows, balancing the cost of empty seats (spoilage) against the cost of denied boardings (Chapter 19).

P

Pairing See crew pairing.

Payload-range diagram A chart of the maximum payload an aircraft can carry as a function of range, bounded by the maximum zero-fuel weight, the maximum take-off weight and the fuel capacity (Chapter 11).

PDEW Passengers per day each way: a market's annual passengers divided by 365 and by two directions. Check whether a source has already split directions before dividing by two (Network Planning Handbook 2026).

PMM The passenger mix model: a linear program that chooses which passengers to carry on a network of given capacities to maximize revenue, accounting for recapture. A building block of IFAM (Chapter 14).

PODS The Passenger Origin-Destination Simulator, a competitive network simulation used by MIT and a research consortium of airlines to test revenue management methods (Chapters 6 and 20).

POS Point of sale: the country or channel where a ticket was sold. It is not necessarily where the traveller lives or where the journey starts.

PRASK, PRASM Passenger revenue per available seat kilometre or mile. Equal to yield times load factor (Chapter 2).

Propagated delay Delay passed on from one flight to the next flight of the same aircraft, crew or passengers when the slack between them is too small to absorb it (Chapter 16).

Prorate The allocation of a ticket's revenue among the legs or carriers of its itinerary, for example in proportion to distance. A prorated leg revenue is an accounting allocation, not a marginal value (Chapter 12).

Protection level The number of seats reserved for a booking class and the classes above it, which lower classes may not use (Chapter 18).

Q

QSI Quality of service index: a market share model in which each itinerary receives a score based on its number of stops, frequency, elapsed time and aircraft, and shares are proportional to scores (Chapters 4 and 6).

R

RASK, RASM Total operating revenue, including cargo and other revenue, per available seat kilometre or mile.

Recapture Spilled demand that the airline retains on another of its own flights or itineraries (Chapter 9).

ReFuelEU Aviation The European Union regulation that requires fuel suppliers at EU airports to blend rising minimum shares of sustainable aviation fuel and requires airlines to refuel there rather than carry extra fuel from elsewhere (Chapter 23).

Revenue management The practice of controlling which fare products are available on each flight over the booking period to maximize revenue from fixed, perishable capacity (Chapters 18 and 20).

Rostering Assigning pairings, reserve days and days off to individual crew members, usually monthly, by bidding or optimization (Chapter 15).

Rotation A sequence of flights flown by one aircraft that returns it to its starting point, such as YUL–SVQ–YUL. The rotation time, including turns, determines how many aircraft a schedule needs (Chapter 13).

RPK, RPM Revenue passenger kilometre and mile: one paying passenger carried one kilometre or one statute mile. The basic unit of traffic.

S

S-curve The more-than-proportional relationship between frequency share and market share in a competitive market (Chapter 4).

SAF Sustainable aviation fuel: fuel from biomass, waste or synthetic sources that can be blended with conventional jet fuel. It costs several times more than conventional fuel, which matters for route economics where blending is mandated (Chapter 23).

Schedule delay The difference between the time a traveller would like to depart and the time an available flight departs. It falls as frequency rises, which is why frequency has value beyond seats (Chapter 13).

Sell-up A passenger buying a more expensive fare product when the cheaper one he or she wanted is unavailable (Chapter 21).

Sixth-freedom traffic Passengers travelling between two foreign countries through an airline's home hub, such as Boston to Paris through Montreal on a Canadian airline (Chapter 10).

Slack The time between when an aircraft, crew or passenger is ready for the next flight and the scheduled departure of that flight. Positive slack absorbs delay (Chapter 16).

Slot A permission to use the full range of airport infrastructure to land or take off on a specific date and at a specific time at a coordinated airport. Slots are allocated twice a year under the IATA Worldwide Airport Slot Guidelines (Chapter 13).

Spill Demand that cannot be accommodated on a flight because it is full; the unconstrained demand above capacity (Chapter 9).

Spiral down The self-reinforcing fall in protection for high fares that occurs when a revenue management system forecasts demand by class from history without accounting for buy-down (Chapter 21).

SSIM IATA's Standard Schedules Information Manual, the industry standard for exchanging schedule data (Chapter 13).

Stage length The distance flown per departure. Average stage length is total distance divided by departures. Unit costs and yields fall as stage length rises (Chapter 3).

Stimulation New traffic created by a better product, typically a new nonstop service, as opposed to traffic diverted from existing itineraries (Chapter 4).

T

T-100 The United States Department of Transportation's monthly operated-traffic data. The segment table reports nonstop flights, seats and passengers; the market table reports passengers between enplanement and deplanement points on one flight number, which is not necessarily their full journey (Network Planning Handbook 2026).

Tail An individual aircraft, identified by its registration. Tail assignment is the operational version of aircraft routing.

Time-space network A network whose nodes are events (an arrival or departure at an airport at a time) and whose arcs are flights and ground waits. The standard representation for fleet assignment and aircraft routing (Chapter 14).

Turn time The ground time between an aircraft's arrival and its next departure. The minimum turn time depends on the aircraft, the airport and whether the next flight is domestic or international (Chapter 13).

U

Unconstraining Estimating true demand from booking data that capacity and booking limits have capped, for example with the expectation-maximization algorithm (Chapter 9).

Utilization Average block hours flown per aircraft per day. Higher utilization spreads ownership and other fixed costs over more output (Chapter 3).

V

Value of time The money a traveller would pay to save one hour of travel time, estimated in a choice model as the ratio of the time coefficient to the fare coefficient (Chapter 5).

Virtual nesting A form of network revenue management that maps each origin-destination fare product into a value bucket on every leg it uses, according to its network value, and controls the buckets with leg-based nesting (Chapter 20).

W

WASG The Worldwide Airport Slot Guidelines, the industry standard published by Airports Council International, IATA and the Worldwide Airport Coordinators Group that sets the rules for slot allocation at coordinated airports (Chapter 13).

Willingness to pay The highest price a traveller would pay for a given itinerary and fare product. Its distribution across travellers determines how much revenue price discrimination can extract (Chapter 21).

Y

Yield Passenger revenue per revenue passenger kilometre or mile. Yield falls with distance and with load factor, so it must be compared on like-for-like stage lengths (Chapter 2).

APPENDIX

D

Notation

The symbols below keep the same meaning throughout the book. Symbols used in a single chapter are defined where they appear and are not listed here. A few letters carry two meanings in different parts of the book; the last table lists them, so that a symbol met in a new chapter can be read in its local sense. Vectors and matrices are not set in bold; their dimensions are stated where they are introduced.

D.1 General mathematics

Table D.1 General mathematical notation.

Symbol	Meaning
$\mathbb{E}[X]$, $\mathrm{Var}(X)$, $\mathrm{Cov}(X,Y)$	expectation, variance, covariance
$\Pr(A)$	probability of event A
$\mathbb{1}\{A\}$	indicator: 1 if A holds, 0 otherwise
$[x]^+$	positive part, $\max(x,0)$
$\mathcal{N}(\mu,\sigma^2)$	normal distribution with mean μ and variance σ^2
φ, Φ	standard normal density and distribution function
$L(z)$	standard normal loss function, $\mathbb{E}\,[Z-z]^+ = \varphi(z) - z\,(1-\Phi(z))$
$\mathrm{Poisson}(\lambda)$, $\mathrm{Bin}(n,p)$	Poisson and binomial distributions
$\mathbb{R}$, $\mathbb{N}$	real numbers, natural numbers
$\arg\max$, $\arg\min$	maximizing and minimizing arguments
$\ln$	natural logarithm

D.2 Measures, costs and capacity

Table D.2 Notation for airline measures, costs and capacity.

Symbol	Meaning
ASK, ASM	available seat kilometres, available seat miles
RPK, RPM	revenue passenger kilometres, revenue passenger miles
LF, LF^*	load factor, RPK/ASK; break-even load factor
p	yield, passenger revenue per RPK
PRASK, RASK, CASK	passenger revenue, total revenue and cost per ASK
TC	trip cost of one flight
BH	block hours
d	stage length (distance flown)
F, v	distance-independent cost per departure and cost per kilometre in $\text{TC}(d) = F + v\,d$
C, C_l	capacity in seats; capacity of leg l
r	contribution per passenger
c	cost per seat

D.3 Demand, frequency and spill

Table D.3 Notation for demand, frequency and spill.

Symbol	Meaning
D	random (unconstrained) demand
μ, σ	mean and standard deviation of demand
$k = \sigma/\mu$	k-factor (coefficient of variation)
$Y = \min(D, C)$	load (constrained demand)
$S = [D - C]^+$	spill
$z = (C - \mu)/\sigma$	standardized capacity
$\text{DF} = \mu/C$	demand factor
f_j, $\bar{f}$	fare of product j, average fare
η	elasticity
f (schedule chapters)	frequency, departures per day or week
FS, MS	frequency share, market share

D.4 Choice models

Table D.4 Notation for choice models.

Symbol	Meaning
n, i, j	decision-maker, alternatives
$U_{ni} = V_{ni} + \varepsilon_{ni}$	utility, its representative part and its random part
β	vector of utility coefficients
P_{ni}	probability that n chooses i
λ_k	dissimilarity (log-sum) parameter of nest k
I_k	inclusive value of nest k, $\ln \sum_{j \in B_k} e^{V_{nj}/\lambda_k}$
s_j, s_0	market share of product j and of the outside option
CS_n	consumer surplus of decision-maker n

D.5 Schedule and fleet

Table D.5 Notation for schedule design and fleet assignment.

Symbol	Meaning
$k \in \mathcal{K}$	fleet types
$i \in \mathcal{F}$	flights (legs) of the schedule
x_{ik}	1 if flight i is flown by fleet type k
c_{ik}	cost of flying flight i with type k, including spill cost
N_k	number of aircraft of type k available
p, r (recapture)	itineraries; b_p^r is the share of passengers spilled from p who accept r

D.6 Revenue management

Table D.6 Notation for revenue management.

Symbol	Meaning
$i = 1, \dots, n$	booking classes, numbered from the highest fare, $f_1 > f_2 > \cdots > f_n$
y_i	protection level for classes $1, \dots, i$
b_i	booking limit for class i
$V_t(x)$, $\Delta V_t(x)$	value of x remaining seats with t periods to go, and the marginal value of a seat
q	show-up probability
AU	authorization level
$l \in \mathcal{L}$	flight legs of the network
$j \in \mathcal{J}$	origin-destination fare products (ODFs)
a_{lj}	1 if product j uses leg l
π_l	bid price of leg l

D.7 Letters with two meanings

Table D.7 Letters used in two senses, with the chapters where the second sense applies.

Symbol	Meanings
α	exponent of the S-curve (Chapters 4 and 13); marginal utility of money, the fare coefficient of a choice model (Chapters 5 and 7)
r	contribution per passenger; an alternative itinerary in recapture rates b^r_p (Chapters 6 and 14)
k	k-factor of demand; index of a fleet type (Chapter 14) or of a nest (Chapter 6)
λ	nest parameter of a nested logit; mean of a Poisson distribution; $\lambda(z)$, the inverse Mills ratio (Chapter 9 and Appendix A)
p	yield (Chapter 2); an itinerary in recapture rates; a probability where stated
s	market share of a product; block speed in the cost model of Chapter 3
f	fare of a product; frequency in the schedule chapters

APPENDIX

E

Sources and Reading List

This book was written from the Network Planning Reading Library, a collection of course material, industry documents and company filings assembled on 27 October 2026, supplemented by current public sources on the industry in 2026. This appendix lists every library document, says what it contains and how old it is, and points to the chapters that use it. It closes with the books and papers to read next.

E.1 The reading library

E.1.1 MIT 16.75J Airline Management, Spring 2006

A graduate course at the Massachusetts Institute of Technology taught by Peter P. Belobaba, with lectures by Cynthia Barnhart, published on MIT OpenCourseWare under a Creative Commons BY-NC-SA licence (Belobaba 2006c). The slides are terse; the book supplies the derivations behind them. Several decks in the library also appear as separate files under descriptive names (route planning, demand and spill, network revenue management, airline economics, operating costs, the passenger choice model, schedule development); they are the same lectures.

Table E.1 Lectures and readings of MIT 16.75J used in this book.

Lecture	Title	Used in
1	Introduction to the airline planning process	Chapters 1 and 10
2	Airline economics review	Chapters 2, 4, 10 and 17
3	Operating costs and productivity measures	Chapter 3
4a	Boeing, Load factor analysis (1978)	Chapter 9

Lecture	Title	Used in
4b	Airline demand analysis and spill modelling	Chapter 9
5	Airline schedule development	Chapter 13
6a	Introduction to the PODS passenger choice model	Chapters 4 and 6
6b	Boeing, Decision window path preference methodology (1993)	Chapter 6
7	The ePODS airline management game	Chapters 6 and 17
9	Airline fleet assignment (Barnhart)	Chapter 14
11	Route planning and evaluation	Chapter 12
13a	Airline operations control	Chapter 16
16	Airline fleet planning models	Chapter 11
18	Introduction to revenue management: flight leg optimization	Chapter 18
19	Flight overbooking: models and practice	Chapter 19
20	Network revenue management: origin-destination control	Chapter 20
23	Pricing challenges: ePODS and reality	Chapters 17 and 21
24	O&D control: what have we learned?	Chapters 20 and 21
26	Willingness to pay and competitive revenue management	Chapter 21
Reading	Boeing, Demand driven dispatch	Chapter 14

E.1.2 MIT 1.206J Airline Schedule Planning, Spring 2003

A graduate course taught by Cynthia Barnhart on the optimization models behind schedule planning, also on MIT OpenCourseWare (Barnhart 2003). It is the source for the network flow, fleet assignment, crew and robustness material. The problem sets supply several of the small instances solved in Chapters 14 and 15.

Table E.2 Lectures and assignments of MIT 1.206J used in this book.

Lecture	Title	Used in
1	Schedule planning and airline networks	Chapters 1, 10 and 14
2–4	Multicommodity flows: formulations, key paths, column and row generation, linear and integer solutions	Chapter 14 and Appendix B
5	The passenger mix model	Chapter 14
6	Fleet assignment	Chapter 14
7	Crew scheduling	Chapter 15
8	Aircraft maintenance routing	Chapter 15
9	Crew pairing and aircraft routing	Chapter 15
10	Schedule design	Chapter 13
11–13	Airline operations and recovery	Chapter 16
14	Robustness in airline schedules (Barnhart and Clarke)	Chapter 16
Problem sets 1–3	Express network design, fleet assignment, crew pairing	Chapters 14 and 15

E.1.3 Discrete choice

Kenneth Train's Discrete Choice Methods with Simulation, second edition (2009), chapters 2 (properties of discrete choice models), 3 (logit), 4 (generalized extreme value models), 6 (mixed logit) and 13 (endogeneity). It is the standard graduate text on the models of Chapters 5 to 7. The book paraphrases and rederives its results; it does not reproduce its text.

E.1.4 Industry, regulatory and company documents

Table E.3 Industry, regulatory and company documents in the library.

Document	Date	Used in
Network planning preparation handbook	27 October 2026	throughout
Air Canada, second-quarter 2026 MD&A	11 August 2026	Chapters 2, 3 and 23
Competition Bureau, Cleared for take-off, executive summary	2025	Chapters 10 and 23
IATA, ACI and WWACG, Worldwide Airport Slot Guidelines, edition 4	effective 1 August 2025	Chapter 13
EUROCONTROL, Arriving on time, Think Paper 6	January 2020	Chapter 16
Airbus, A321 aircraft characteristics, revision 33	1 June 2024	Chapters 11 and 13
Lufthansa Systems, Air Canada and NetLine/Ops ++	21 January 2025	Chapters 16 and 23

The handbook deserves a word. It is a thirty-page preparation guide written for a new network planning analyst: the vocabulary, the unit discipline, the economics of a route decision, the data sources and their traps, a worked Montreal–Seville case and a plan for the first ninety days. Every Air Canada fact it states is cited to a public source. This book develops its arguments from first principles and reproduces its exercises with attribution.

E.1.5 Current public sources

Facts about the industry in 2026 (fleet programs, network announcements, regulation, technology and market structure) come from public sources consulted on 27 October 2026: company releases and filings, regulators, industry associations and vendors. Each is cited where it is used, mostly in Chapter 23, and listed in the references. Public facts change quickly; recheck any of them before relying on it at work.

E.2 Reading the older material

The 2003 and 2006 courses predate branded fares, the New Distribution Capability, continuous pricing, the A321XLR and the current Canadian passenger protection rules. Their methods are not dated: the spill model, the S-curve, Littlewood's rule, EMSR, the fleet assignment model, crew pairing by column generation and bid prices

are what planning and revenue management systems still implement, sometimes in refined form. Their numbers are dated. Fares, market shares, cost levels, reported revenue gains and the parameter values used in simulations describe the industry of their time, and this book labels them historical whenever it uses them. When a lecture states a rule of thumb (a spill rate above 5 percent is too high, revenue management adds 4 to 6 percent to revenue), treat it as an order of magnitude to test against current data, not as a constant.

E.3 Where to go next

Textbooks Belobaba, Odoni and Barnhart's The Global Airline Industry (2015) is the closest single book to the scope of this one and is written by the authors of the two MIT courses. Doganis (2019) gives the commercial and strategic view. For revenue management, Talluri and van Ryzin (2004b) remains the reference for the theory, and Gallego and Topaloglu (2019) covers pricing and choice-based methods in more depth. For operations research applications across the planning chain, see the handbook edited by Barnhart and Smith (2012) and Bazargan (2010). For demand modelling, read Train (2009) and Garrow (2010); for forecasting, Hyndman and Athanasopoulos (2021).

Papers by topic

- Spill and unconstraining: Swan (2002); Weatherford and Pölt (2002); Abramovich (2013) on spill under revenue management.
- Itinerary choice: Coldren and coauthors (2003); Coldren and Koppelman (2005).
- Airline demand and competition: Berry, Carnall and Spiller (1996); Berry, Levinsohn and Pakes (1995).
- Schedule design and fleet assignment: Hane and coauthors (1995); Barnhart, Kniker and Lohatepanont (2002); Lohatepanont and Barnhart (2004); Barnhart and Cohn (2004) for a survey.
- Crew and routing: Vance and coauthors (1997); Gopalan and Talluri (1998); Cohn and Barnhart (2003); Barnhart and coauthors on branch-and-price (1998).
- Robustness and recovery: Lan, Clarke and Barnhart (2006); Clausen and coauthors (2010).
- Revenue management: Littlewood (1972); Belobaba (1987, 1989); Brumelle and McGill (1993); Talluri and van Ryzin (1998, 2004a); Gallego and van Ryzin (1994); Cooper, Homem-de-Mello and Kleywegt (2006); Fiig and coauthors (2010); McGill and van Ryzin (1999) for a survey.

References

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." Journal of the American Statistical Association 105 (490): 493–505.

Abara, Jeph. 1989. "Applying Integer Linear Programming to the Fleet Assignment Problem." Interfaces 19 (4): 20–28.

Abramovich, Michael. 2013. "Impacts of Revenue Management on Estimates of Spilled Passenger Demand." S.M. thesis, Massachusetts Institute of Technology. https://dspace.mit.edu/entities/publication/32bc4894-1f16-4269-8354-75ec04ce27d6.

Adams, William James, and Janet L. Yellen. 1976. "Commodity Bundling and the Burden of Monopoly." Quarterly Journal of Economics 90 (3): 475–98.

AGIFORS. 2026. "SSP 2026 Presentation Abstracts (Tentative)." AGIFORS Scheduling and Strategic Planning Study Group. 2026. https://agifors.org/ssp2026_abstracts.

Ahuja, Ravindra K., Thomas L. Magnanti, and James B. Orlin. 1993. Network Flows: Theory, Algorithms, and Applications. Englewood Cliffs, NJ: Prentice Hall.

Air Canada. 2025. "Mark Galardo." Executive biography. Air Canada Media Centre. February 27, 2025. https://www.aircanada.com/media/mark-galardo/.

———. 2026a. "A New Era in Customer Comfort and Network Growth Begins as Air Canada Receives Its First Airbus A321XLR." Air Canada Media Centre. April 24, 2026. https://www.aircanada.com/media/a-new-era-in-customer-comfort-and-network-growth-begins-as-air-canada-receives-its-first-airbus-a321xlr/.

———. 2026b. "Accueil Chaleureux : Une Nouvelle Ère s'amorce Pour Air Canada Avec Le Décollage Du Premier Vol Commercial de l'A321XLR d'Airbus." Air Canada, Salle de Presse. June 9, 2026. https://www.aircanada.com/medias/accueil-chaleureux—une-nouvelle-ere-samorce-pour-air-canada-avec-le-decollage-du-premier-vol-commercial-de-la321xlr-dairbus/.

———. 2026c. "Air Canada Announces Appointment of Anko van Der Werff as President and Chief Executive Officer." Air Canada Media Centre. July 8, 2026. https://www.aircanada.com/media/air-canada-announces-appointment-of-anko-van-der-werff-as-president-and-chief-executive-officer/.

———. 2026d. "Air Canada Reports Second Quarter 2026 Financial Results." Air Canada Media Centre. August 11, 2026.

https://www.aircanada.com/media/air-canada-reports-second-quarter-2026-financial-results/.

———. 2026e. “Second Quarter 2026 Management’s Discussion and Analysis of Results of Operations and Financial Condition.” Air Canada.

———. 2026f. “Air Canada Expands Toronto Global Hub with New International Routes and Destinations in Summer 2027.” Air Canada Media Centre. September 3, 2026. https://www.aircanada.com/media/air-canada-expands-toronto-global-hub-with-new-international-routes-and-destinations-in-summer-2027/.

———. 2026g. “More Flamenco and Tapas: Air Canada Adds Seville to Growing European Network.” News release. https://www.aircanada.com/media/more-flamenco-and-tapas-air-canada-adds-seville-to-growing-european-network.

———. 2026h. “Air Canada Expands Canada-U.S. Connectivity with New Summer 2027 Routes, Increasing Options for Travellers Under Its Transborder Joint Business Agreement with United Airlines.” News release. https://www.aircanada.com/media/air-canada-expands-canada-us-connectivity-with-new-summer-2027-routes-increasing-options-for-travellers-under-its-transborder-joint-business-agreement-with-united-airlines/.

———. n.d.a. “Airbus A220-300 Fact Sheet.” Air Canada. Accessed October 27, 2026. https://aircanada.com/content/dam/aircanada/portal/documents/PDF/en/a220-fact-sheet-en.pdf.

———. n.d.b. “Our Star Alliance and Other Airline Partners.” Aircanada.com. Accessed October 27, 2026. https://www.aircanada.com/ca/en/aco/home/book/routes-and-partners/star-alliance-and-other-airline-partners.html.

Airbus. 2024. “A321 Aircraft Characteristics: Airport and Maintenance Planning.” Airbus. https://www.aircraft.airbus.com/sites/g/files/jlcbta126/files/2024-06/AC_A321_0624.pdf.

Airlines Reporting Corporation. 2026a. “Annual U.S. Travel Agency Air Ticket Sales Surpass $100 Billion for the First Time.” ARC Newsroom. January 15, 2026. https://www2.arccorp.com/about-us/newsroom/2026-news-releases/december-2025-air-ticket-sales/.

———. 2026b. “May U.S. Travel Agency Air Ticket Sales Near $10 Billion.” ARC Newsroom. June 18, 2026. https://www2.arccorp.com/about-us/newsroom/2026-news-releases/may-air-ticket-sales/.

Angrist, Joshua D., and Jörn-Steffen Pischke. 2009. Mostly Harmless Econometrics: An Empiricist’s Companion. Princeton: Princeton University Press.

Athanasopoulos, George, Roman A. Ahmed, and Rob J. Hyndman. 2009. “Hierarchical Forecasts for Australian Domestic Tourism.” International Journal of Forecasting 25 (1): 146–66.

ATPCO. 2022. "ATPCO Leads Airline Industry Transformation from Static Fare Filing to the Future of Airline Offers; Commits to Move Industry to 80% of All Airline Offers Dynamically Created by 2026." ATPCO Press Release. October 11, 2022. https://atpco.net/single-press/atpco-leads-industry-transformation-from-static-fare-filing-to-the-future-of-airline-offers-commits-to-move-industry-to-80-of-all-airline-offers-dynamically-created-by-2026/.

———. 2023. "Transformation Under Way from Low-Fare Search to Modern Airline Retailing." ATPCO Blog. July 24, 2023. https://atpco.net/single-blog/transformation-under-way-from-low-fare-search-to-modern-airline-retailing/.

Aviacionline. 2022. "Air Canada: Airbus A321XLR Fleet Will Use Pratt & Whitney Engines." Aviacionline, April 26, 2022. https://www.aviacionline.com/air-canada-airbus-a321xlr-fleet-will-use-pratt-whitney-engines.

Aviation A2Z. 2026. "WestJet Cancels Flights to These 13 International Destinations in 2026." Aviation A2Z. October 9, 2026. https://aviationa2z.com/index.php/2026/10/09/westjet-13-international-destinations-disappear-from-network/.

Barnhart, Cynthia. 2003. "Airline Schedule Planning." Course lecture notes, MIT 1.206J/16.77J/ESD.215J. https://ocw.mit.edu/courses/1-206j-airline-schedule-planning-spring-2003/pages/lecture-notes/.

———. 2006. "Airline Fleet Assignment." Guest lecture slides, MIT 16.75J/1.234J Airline Management. https://ocw.mit.edu/courses/16-75j-airline-management-spring-2006/pages/lecture-notes/.

Barnhart, Cynthia, Natashia L. Boland, Lloyd W. Clarke, Ellis L. Johnson, George L. Nemhauser, and Rajesh G. Shenoi. 1998. "Flight String Models for Aircraft Fleeting and Routing." Transportation Science 32 (3): 208–20.

Barnhart, Cynthia, and John-Paul Clarke. 2003. "New Approaches to Add Robustness into Airline Schedules." Lecture slides, MIT 1.206J Airline Schedule Planning. https://ocw.mit.edu/courses/1-206j-airline-schedule-planning-spring-2003/pages/lecture-notes/.

Barnhart, Cynthia, and Amy Cohn. 2004. "Airline Schedule Planning: Accomplishments and Opportunities." Manufacturing & Service Operations Management 6 (1): 3–22. https://doi.org/10.1287/msom.1030.0018.

Barnhart, Cynthia, Ellis L. Johnson, George L. Nemhauser, Martin W. P. Savelsbergh, and Pamela H. Vance. 1998. "Branch-and-Price: Column Generation for Solving Huge Integer Programs." Operations Research 46 (3): 316–29.

Barnhart, Cynthia, Timothy S. Kniker, and Manoj Lohatepanont. 2002. "Itinerary-Based Airline Fleet Assignment." Transportation Science 36 (2): 199–217. https://doi.org/10.1287/trsc.36.2.199.566.

Barnhart, Cynthia, and Barry Smith, eds. 2012. Quantitative Problem Solving Methods in the Airline Industry: A Modeling Methodology Handbook. Vol. 169. International Series in Operations Research and Management Science. New York: Springer.

Bazargan, Massoud. 2010. Airline Operations and Scheduling. 2nd ed. Farnham: Ashgate.

Bellman, Richard. 1957. Dynamic Programming. Princeton, NJ: Princeton University Press.

Belobaba, Peter P. 1987. "Airline Yield Management: An Overview of Seat Inventory Control." Transportation Science 21 (2): 63–73.

———. 1989. "Application of a Probabilistic Decision Model to Airline Seat Inventory Control." Operations Research 37 (2): 183–97.

———. 2002. "O&D Control: What Have We Learned?" Presentation to the IATA Revenue Management and Pricing Conference, Toronto; reprinted in MIT 16.75J/1.234J Airline Management lecture notes, 2006, October. https://ocw.mit.edu/courses/16-75j-airline-management-spring-2006/pages/lecture-notes/.

———. 2006a. "Willingness to Pay and Competitive Revenue Management." Lecture slides, MIT 16.75J/1.234J Airline Management. https://ocw.mit.edu/courses/16-75j-airline-management-spring-2006/pages/lecture-notes/.

———. 2006b. "Airline Demand Analysis and Spill Modeling." Lecture notes, MIT 16.75J/1.234J Airline Management, February. https://ocw.mit.edu/courses/16-75j-airline-management-spring-2006/pages/lecture-notes/.

———. 2006c. "Introduction to the Airline Planning Process." Lecture slides, MIT 16.75J/1.234J Airline Management, February 8. https://ocw.mit.edu/courses/16-75j-airline-management-spring-2006/pages/lecture-notes/.

———. 2006d. "Airline Economics Review." Lecture slides, MIT 16.75J/1.234J Airline Management, February 13. https://ocw.mit.edu/courses/16-75j-airline-management-spring-2006/pages/lecture-notes/.

———. 2006e. "Operating Costs and Productivity Measures." Lecture slides, MIT 16.75J/1.234J Airline Management, February 15. https://ocw.mit.edu/courses/16-75j-airline-management-spring-2006/pages/lecture-notes/.

———. 2006f. "Airline Schedule Development." Lecture slides, MIT 16.75J/1.234J Airline Management, February 22. https://ocw.mit.edu/courses/16-75j-airline-management-spring-2006/pages/lecture-notes/.

———. 2006g. "Introduction to PODS Passenger Choice Model." Lecture slides, MIT 16.75J/1.234J Airline Management, February 27. https://ocw.mit.edu/courses/16-75j-airline-management-spring-2006/pages/lecture-notes/.

———. 2006h. “ePODS Airline Management Educational Game.” Lecture slides, MIT 16.75J/1.234J Airline Management, March 1. https://ocw.mit.edu/courses/16-75j-airline-management-spring-2006/pages/lecture-notes/.

———. 2006i. “Route Planning and Evaluation.” Lecture slides, MIT 16.75J/1.234J Airline Management, March 15. https://ocw.mit.edu/courses/16-75j-airline-management-spring-2006/pages/lecture-notes/.

———. 2006j. “Airline Operations Control.” Lecture slides, MIT 16.75J/1.234J Airline Management, March 22. https://ocw.mit.edu/courses/16-75j-airline-management-spring-2006/pages/lecture-notes/.

———. 2006k. “Airline Fleet Planning Models.” Lecture slides, MIT 16.75J/1.234J Airline Management, April 10. https://ocw.mit.edu/courses/16-75j-airline-management-spring-2006/pages/lecture-notes/.

———. 2006l. “Introduction to Revenue Management: Flight Leg Revenue Optimization.” Lecture slides, MIT 16.75J/1.234J Airline Management, April 19. https://ocw.mit.edu/courses/16-75j-airline-management-spring-2006/pages/lecture-notes/.

———. 2006m. “Flight Overbooking: Models and Practice.” Lecture slides, MIT 16.75J/1.234J Airline Management, April 24. https://ocw.mit.edu/courses/16-75j-airline-management-spring-2006/pages/lecture-notes/.

———. 2006n. “Network Revenue Management: Origin-Destination Control.” Lecture slides, MIT 16.75J/1.234J Airline Management, April 26. https://ocw.mit.edu/courses/16-75j-airline-management-spring-2006/pages/lecture-notes/.

———. 2006o. “Pricing Challenges: ePODS and Reality.” Lecture slides, MIT 16.75J/1.234J Airline Management, May 8. https://ocw.mit.edu/courses/16-75j-airline-management-spring-2006/pages/lecture-notes/.

Belobaba, Peter P., Amedeo Odoni, and Cynthia Barnhart, eds. 2015. The Global Airline Industry. 2nd ed. Chichester: Wiley.

Ben-Akiva, Moshe, and Steven R. Lerman. 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. Cambridge, MA: MIT Press.

Berge, Matthew E., and Craig A. Hopperstad. 1993. “Demand Driven Dispatch: A Method for Dynamic Aircraft Capacity Assignment, Models and Algorithms.” Operations Research 41 (1): 153–68. https://doi.org/10.1287/opre.41.1.153.

Berry, Steven T. 1994. “Estimating Discrete-Choice Models of Product Differentiation.” RAND Journal of Economics 25 (2): 242–62.

Berry, Steven T., Michael Carnall, and Pablo T. Spiller. 1996. “Airline Hubs: Costs, Markups and the Implications of Customer Heterogeneity.”

NBER Working Paper 5561. National Bureau of Economic Research. https://www.nber.org/papers/w5561.

Berry, Steven T., James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." Econometrica 63 (4): 841–90.

Berry, Steven, and Panle Jia. 2010. "Tracing the Woes: An Empirical Analysis of the Airline Industry." American Economic Journal: Microeconomics 2 (3): 1–43.

Bertsimas, Dimitris, and Ioana Popescu. 2003. "Revenue Management in a Dynamic Network Environment." Transportation Science 37 (3): 257–77.

Bertsimas, Dimitris, and John N. Tsitsiklis. 1997. Introduction to Linear Optimization. Belmont, MA: Athena Scientific.

Boeing. 2011. "Airport Noise and Emissions Regulations: Montreal Dorval (YUL)." Boeing Commercial Airplanes, Airport Noise Regulations. March 2011. https://www.boeing.com/resources/boeingdotcom/commercial/noise/dorval.html.

———. 2026. "FAA Certifies Boeing 737-10." Boeing Features. October 20, 2026. https://www.boeing.com/features/2026/10/faa-certifies-boeing-737-10.

Boeing Commercial Airplane Company. 1978. "Load Factor Analysis: The Relationship Between Flight Load and Passenger Turnaway." Seattle: Boeing.

Boeing Commercial Airplane Group. 1993. "Decision Window Path Preference Methodology: Time Mode." Methodology presentation, reading in MIT 16.75J Airline Management. Seattle: Boeing.

———. n.d. "Demand Driven Dispatch." Brochure. Seattle: Boeing Commercial Airplane Group.

Borenstein, Severin. 1989. "Hubs and High Fares: Dominance and Market Power in the U.S. Airline Industry." RAND Journal of Economics 20 (3): 344–65.

Box, George E. P., and Gwilym M. Jenkins. 1970. Time Series Analysis: Forecasting and Control. San Francisco: Holden-Day.

Boynton, Christine, and Helen Massy-Beresford. 2025. "Generative AI 'Super Analyst' May Change How Airlines Set Prices." Aviation Week, January 7, 2025. https://aviationweek.com/air-transport/airlines-lessors/generative-ai-super-analyst-may-change-how-airlines-set-prices.

Bratu, Stephane, and Cynthia Barnhart. 2006. "Flight Operations Recovery: New Approaches Considering Passenger Recovery." Journal of Scheduling 9 (3): 279–98.

Brueckner, Jan K. 2001. "The Economics of International Codesharing: An Analysis of Airline Alliances." International Journal of Industrial Organization 19 (10): 1475–98.

Brumelle, Shelby L., and Jeffrey I. McGill. 1993. "Airline Seat Allocation with Multiple Nested Fare Classes." Operations Research 41 (1): 127–37.

Bureau of Transportation Statistics. 2026a. "First Quarter of OD40 Data Now Live." BTS Newsroom. February 17, 2026. https://www.bts.gov/newsroom/first-quarter-od40-data-live.

———. 2026b. "Origin-Destination 40 Percent Survey of Airline Passenger Traffic." Bureau of Transportation Statistics. March 6, 2026. https://www.bts.gov/OD-40.

Business Travel News Europe. n.d. "Lufthansa Launches NDC Continuous Pricing for Lower Fares." Business Travel News Europe. Accessed October 27, 2026. https://www.businesstravelnewseurope.com/TMC-Distribution/Lufthansa-launches-NDC-continuous-pricing-for-lower-fares-.

Canadian Airports Council. 2024. "Letter to the House of Commons Standing Committee on Transport, Infrastructure and Communities." Brief, 31 May 2024. House of Commons of Canada. https://www.ourcommons.ca/Content/Committee/441/TRAN/Brief/BR13228039/br-external/CanadianAirportsCouncil-e.pdf.

Canadian Transportation Agency. 2023. "Flight Delays and Cancellations: A Guide." Canadian Transportation Agency Publications. January 13, 2023. https://www.otc-cta.gc.ca/eng/publication/flight-delays-and-cancellations-a-guide.

———. 2024. "Regulations Amending the Air Passenger Protection Regulations (Proposed)." Canada Gazette, Part I 158 (51). https://gazette.gc.ca/rp-pr/p1/2024/2024-12-21/html/reg1-eng.html.

———. 2026a. "Bumping (Denied Boarding): Rebooking, Refunds and Compensation." Air Passenger Protection. July 22, 2026. https://protection-passager-passenger.otc-cta.gc.ca/en/refunds-and-compensation/bumping-denied-boarding-rebooking-refunds-compensation.

———. 2026b. "Air Passenger Protection." Transition Book: Chair 2026. September 23, 2026. https://otc-cta.gc.ca/eng/transition-book-chair-2026/air-passenger-protection.

Card, David, and Alan B. Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania." American Economic Review 84 (4): 772–93.

Casella, George, and Roger L. Berger. 2002. Statistical Inference. 2nd ed. Pacific Grove, CA: Duxbury.

Caves, Douglas W., Laurits R. Christensen, and Michael W. Tretheway. 1984. "Economies of Density Versus Economies of Scale: Why Trunk and Local Service Airline Costs Differ." RAND Journal of Economics 15 (4): 471–89.

CBC News. 2025. "Is It Lights Out for Overnight Flights at Montreal's Trudeau Airport?" Yahoo News Canada (Republished from CBC), December 23, 2025. https://ca.news.yahoo.com/lights-overnight-flights-montreal-trudeau-001459537.html.

Chvátal, Václav. 1973. “Edmonds Polytopes and a Hierarchy of Combinatorial Problems.” Discrete Mathematics 4 (4): 305–37. https://doi.org/10.1016/0012-365X(73)90167-2.

Cirium. 2026a. “Traffic and Fares Data Sheet.” Cirium. https://assets.fta.cirium.com/wp-content/uploads/2026/04/10160130/Traffic-Fares-Data-Sheet_March-2026.pdf.

———. 2026b. “How Airlines Are Deploying the Airbus A321XLR.” Ascend Consultancy Viewpoint 92. Cirium ThoughtCloud. October 7, 2026. https://www.cirium.com/thoughtcloud/ascend-consultancy-how-airlines-are-deploying-the-airbus-a321xlr/.

Clarke, L. W., C. A. Hane, E. L. Johnson, and G. L. Nemhauser. 1996. “Maintenance and Crew Considerations in Fleet Assignment.” Transportation Science 30 (3): 249–60. https://doi.org/10.1287/trsc.30.3.249.

Clausen, Jens, Allan Larsen, Jesper Larsen, and Natalia J. Rezanova. 2010. “Disruption Management in the Airline Industry: Concepts, Models and Methods.” Computers & Operations Research 37 (5): 809–21.

Cléaz-Savoyen, Richard L. 2005. “Airline Revenue Management Methods for Less Restricted Fare Structures.” S.M. thesis, Massachusetts Institute of Technology. https://dspace.mit.edu/handle/1721.1/32439.

Cohn, Amy M., and Cynthia Barnhart. 2003. “Improving Crew Scheduling by Incorporating Key Maintenance Routing Decisions.” Operations Research 51 (3): 387–96.

Coldren, Gregory M., and Frank S. Koppelman. 2005. “Modeling the Competition Among Air-Travel Itinerary Shares: GEV Model Development.” Transportation Research Part A: Policy and Practice 39 (4): 345–65.

Coldren, Gregory M., Frank S. Koppelman, Krishnan Kasturirangan, and Amit Mukherjee. 2003. “Modeling Aggregate Air-Travel Itinerary Shares: Logit Model Development at a Major US Airline.” Journal of Air Transport Management 9 (6): 361–69.

Competition Bureau Canada. 2025. “Cleared for Take-Off: Elevating Airline Competition. A Market Study of Canada’s Airline Industry.” Market study. Government of Canada. https://competition-bureau.canada.ca/en/how-we-foster-competition/education-and-outreach/publications/cleared-take-elevating-airline-competition.

Conlon, Christopher, and Jeff Gortmaker. 2020. “Best Practices for Differentiated Products Demand Estimation with PyBLP.” RAND Journal of Economics 51 (4): 1108–61.

Cooper, William L., Tito Homem-de-Mello, and Anton J. Kleywegt. 2006. “Models of the Spiral-down Effect in Revenue Management.” Operations Research 54 (5): 968–87.

Cordeau, Jean-François, Goran Stojković, François Soumis, and Jacques Desrosiers. 2001. "Benders Decomposition for Simultaneous Aircraft Routing and Crew Scheduling." Transportation Science 35 (4): 375–88.

Curry, Renwick E. 1990. "Optimal Airline Seat Allocation with Fare Classes Nested by Origins and Destinations." Transportation Science 24 (3): 193–204.

Daily Hive. 2026. "WestJet Criticizes $376M Federal Loans for Flair, Porter, Air Transat." Daily Hive, September 15, 2026. https://dailyhive.com/canada/westjet-flair-porter-transat-government-loans.

Dantzig, George B. 1963. Linear Programming and Extensions. Princeton, NJ: Princeton University Press.

Dantzig, George B., and Philip Wolfe. 1960. "Decomposition Principle for Linear Programs." Operations Research 8 (1): 101–11. https://doi.org/10.1287/opre.8.1.101.

Delta Air Lines. 2026. "Delta Responds to Rep. Pallone on Misinformation on Consumer Surveillance Pricing." Delta News Hub. September 18, 2026. https://news.delta.com/delta-responds-rep-pallone-misinformation-consumer-surveillance-pricing.

Dempster, Arthur P., Nan M. Laird, and Donald B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." Journal of the Royal Statistical Society, Series B 39 (1): 1–38.

Deneckere, Raymond J., and R. Preston McAfee. 1996. "Damaged Goods." Journal of Economics and Management Strategy 5 (2): 149–74.

Department of Finance Canada. 2026. "Government of Canada Introduces Targeted Support to Help Canada's Airline Sector Weather Global Fuel Market Volatility." Department of Finance Canada News. June 8, 2026. https://www.canada.ca/en/department-finance/news/2026/06/government-of-canada-introduces-targeted-support-to-help-canadas-airline-sector-weather-global-fuel-market-volatility.html.

Desrosiers, Jacques, and Marco E. Lübbecke. 2005. "A Primer in Column Generation." In Column Generation, edited by Guy Desaulniers, Jacques Desrosiers, and Marius M. Solomon, 1–32. New York: Springer.

Dijkstra, E. W. 1959. "A Note on Two Problems in Connexion with Graphs." Numerische Mathematik 1: 269–71. https://doi.org/10.1007/BF01386390.

Doganis, Rigas. 2019. Flying Off Course: Airline Economics and Marketing. 5th ed. London: Routledge.

Douglas, George W., and James C. Miller III. 1974. Economic Regulation of Domestic Air Transport: Theory and Policy. Washington, DC: Brookings Institution.

Dubé, Jean-Pierre, Jeremy T. Fox, and Che-Lin Su. 2012. "Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation." Econometrica 80 (5): 2231–67.

Dunning, Iain, Joey Huchette, and Miles Lubin. 2017. "JuMP: A Modeling Language for Mathematical Optimization." SIAM Review 59 (2): 295–320. https://doi.org/10.1137/15M1020575.

EASA. 2024. "Fit for 55 and ReFuelEU Aviation." European Union Aviation Safety Agency. July 4, 2024. https://www.easa.europa.eu/en/light/topics/fit-55-and-refueleu-aviation.

Efron, Bradley, and Robert J. Tibshirani. 1993. An Introduction to the Bootstrap. New York: Chapman & Hall.

ePlane AI. 2025. "Sabre Introduces AI-Driven Revenue Engine Aiming to Transform Airline Economics." ePlane AI. October 14, 2025. https://www.eplaneai.com/news/sabre-introduces-ai-driven-revenue-engine-aiming-to-transform-airline-economics.

EUROCONTROL. 2020. "Arriving on Time: The Passenger Priority." Think Paper 6. EUROCONTROL. https://www.eurocontrol.int/publication/arriving-time-passenger-priority.

European Commission. 2026. "Commission Welcomes Landmark Agreement on Revised Air Passenger Rights." Mobility and Transport News. June 15, 2026. https://transport.ec.europa.eu/news-events/news/commission-welcomes-landmark-agreement-revised-air-passenger-rights-2026-06-15_en.

———. n.d. "Reducing Emissions from Aviation." Directorate-General for Climate Action. Accessed October 27, 2026. https://climate.ec.europa.eu/areas-action/transport-decarbonisation/reducing-emissions-aviation_en.

European Parliament and Council of the European Union. 2001. Directive 2000/84/EC of 19 January 2001 on Summer-Time Arrangements. Official Journal of the European Communities. Vol. L 31. https://eur-lex.europa.eu/eli/dir/2000/84/oj/eng.

———. 2023. Regulation (EU) 2023/2405 of 18 October 2023 on Ensuring a Level Playing Field for Sustainable Air Transport (ReFuelEU Aviation). Official Journal of the European Union. https://eur-lex.europa.eu/eli/reg/2023/2405/oj.

European Union. 2026. "Air Passenger Rights." Your Europe. September 21, 2026. https://europa.eu/youreurope/citizens/travel/passenger-rights/air/index_en.htm.

Fiig, Thomas, Karl Isler, Craig Hopperstad, and Peter P. Belobaba. 2010. "Optimization of Mixed Fare Structures: Theory and Applications." Journal of Revenue and Pricing Management 9 (1-2): 152–70.

Fisher, Marshall L. 1981. "The Lagrangian Relaxation Method for Solving Integer Programming Problems." Management Science 27 (1): 1–18. https://doi.org/10.1287/mnsc.27.1.1.

FlightGlobal. 2025. "Air Transat Now Expects A321XLR Deliveries to Start in 2027." FlightGlobal, December 18, 2025.

https://www.flightglobal.com/fleets/2025/12/air-transat-now-expects-a321xlr-deliveries-to-start-in-2027/.

Fourer, Robert, David M. Gay, and Brian W. Kernighan. 2003. AMPL: A Modeling Language for Mathematical Programming. 2nd ed. Pacific Grove, CA: Thomson/Brooks/Cole.

Gallego, Guillermo, and Garrett van Ryzin. 1994. "Optimal Dynamic Pricing of Inventories with Stochastic Demand over Finite Horizons." Management Science 40 (8): 999–1020.

Gallego, Guillermo, and Huseyin Topaloglu. 2019. Revenue Management and Pricing Analytics. Vol. 279. International Series in Operations Research and Management Science. New York: Springer.

Garrow, Laurie A. 2010. Discrete Choice Modelling and Air Travel Demand: Theory and Applications. Farnham: Ashgate.

Geoffrion, A. M. 1974. "Lagrangean Relaxation for Integer Programming." Mathematical Programming Study 2: 82–114. https://doi.org/10.1007/BFb0120690.

Gillen, David W., William G. Morrison, and Christopher Stewart. 2004. "Air Travel Demand Elasticities: Concepts, Issues and Measurement." Final report. Ottawa: Department of Finance Canada.

Gilmore, P. C., and R. E. Gomory. 1961. "A Linear Programming Approach to the Cutting-Stock Problem." Operations Research 9 (6): 849–59. https://doi.org/10.1287/opre.9.6.849.

Gneiting, Tilmann. 2011. "Making and Evaluating Point Forecasts." Journal of the American Statistical Association 106 (494): 746–62.

Gomory, Ralph E. 1958. "Outline of an Algorithm for Integer Solutions to Linear Programs." Bulletin of the American Mathematical Society 64 (5): 275–78. https://doi.org/10.1090/S0002-9904-1958-10224-4.

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." Journal of Econometrics 225 (2): 254–77.

Goolsbee, Austan, and Chad Syverson. 2008. "How Do Incumbents Respond to the Threat of Entry? Evidence from the Major Airlines." Quarterly Journal of Economics 123 (4): 1611–33.

Gopalan, Ram, and Kalyan T. Talluri. 1998. "The Aircraft Maintenance Routing Problem." Operations Research 46 (2): 260–71.

Government of British Columbia. 2026. "Permanent Daylight Saving Time." Government of British Columbia. 2026. https://www2.gov.bc.ca/gov/content/governments/celebrating-british-columbia/daylight-saving-time.

Government of Canada. 2019. "Air Passenger Protection Regulations (SOR/2019-150)." Justice Laws Website. 2019. https://laws-lois.justice.gc.ca/eng/regulations/SOR-2019-150/.

Hamilton, Scott. 2026. "Pratt & Whitney Expands MROs in Push to Resolve Geared Turbofan Issues." Leeham News and Analysis. July 15, 2026. https://leehamnews.com/2026/07/15/pratt-whitney-expands-mros-in-push-to-wrap-up-gtf-problems/.

Hane, Christopher A., Cynthia Barnhart, Ellis L. Johnson, Roy E. Marsten, George L. Nemhauser, and Gabriele Sigismondi. 1995. "The Fleet Assignment Problem: Solving a Large-Scale Integer Program." Mathematical Programming 70: 211–32.

Hart, William E., Jean-Paul Watson, and David L. Woodruff. 2011. "Pyomo: Modeling and Solving Mathematical Programs in Python." Mathematical Programming Computation 3 (3): 219–60. https://doi.org/10.1007/s12532-011-0026-8.

Hausman, Jerry A. 1978. "Specification Tests in Econometrics." Econometrica 46 (6): 1251–71.

———. 1997. "Valuation of New Goods Under Perfect and Imperfect Competition." In The Economics of New Goods, edited by Timothy F. Bresnahan and Robert J. Gordon, 207–48. Chicago: University of Chicago Press.

Hausman, Jerry, and Daniel McFadden. 1984. "Specification Tests for the Multinomial Logit Model." Econometrica 52 (5): 1219–40.

Hellenic Civil Aviation Authority. n.d. "How Are CO2 Emissions Calculated from the Fuel Used?" HCAA Frequently Asked Questions. Accessed October 27, 2026. https://hcaa.gov.gr/en/faq/18-pos-ypologizontai-oi-ekpompes-co2-me-basi-kaysimo-poy-hrisimopoieitai.

Hoffman, A. J., and J. B. Kruskal. 1956. "Integral Boundary Points of Convex Polyhedra." In Linear Inequalities and Related Systems, edited by H. W. Kuhn and A. W. Tucker, 223–46. Annals of Mathematics Studies 38. Princeton, NJ: Princeton University Press.

Huangfu, Q., and J. A. J. Hall. 2018. "Parallelizing the Dual Revised Simplex Method." Mathematical Programming Computation 10 (1): 119–42. https://doi.org/10.1007/s12532-017-0130-5.

Hyndman, Rob J., and George Athanasopoulos. 2021. Forecasting: Principles and Practice. 3rd ed. Melbourne: OTexts. https://otexts.com/fpp3/.

Hyndman, Rob J., and Anne B. Koehler. 2006. "Another Look at Measures of Forecast Accuracy." International Journal of Forecasting 22 (4): 679–88.

IATA. 2026a. "Fact Sheet: Airport Slots." International Air Transport Association, Pressroom Fact Sheets. June 2026. https://www.iata.org/en/iata-repository/pressroom/fact-sheets/fact-sheet-airport-slots/.

———. 2026b. "Fact Sheet: CORSIA." International Air Transport Association, Pressroom Fact Sheets. June 2026. https://www.iata.org/en/iata-repository/pressroom/fact-sheets/fact-sheet-corsia/.

———. 2026c. "Fulfilling with Orders: ONE Order (Fact Sheet)." International Air Transport Association, Pressroom Fact Sheets. June 2026. https://www.iata.org/en/iata-repository/pressroom/fact-sheets/fact-sheet-one-order/.

———. 2026d. “Global Outlook for Air Transport: Energy in Crisis.” International Air Transport Association. https://www.iata.org/en/iata-repository/publications/economic-reports/global-outlook-for-air-transport-june-2026/.

———. 2026e. “IATA Highlights Priorities to Create Value from Digitalization.” Press release No. 50. International Air Transport Association Pressroom. October 28, 2026. https://www.iata.org/en/pressroom/2026-releases/10-28-iata-highlights-priorities-create-value-digitalization.

IATA, ACI, and WWACG. 2025. “Worldwide Airport Slot Guidelines.” 4th ed. International Air Transport Association. https://www.iata.org/en/programs/ops-infra/slots/slot-guidelines/.

ICAO. 2023. “CORSIA Offsetting Calculations: Examples.” Workshop presentation, ICAO-SASO Environment Workshop, 24 to 27 October 2023. https://www.icao.int/ESAF/Documents/meetings/2023/ICAO-SASO%20ENV%20Workshop%2024-27%20October%202023/PPTs/PPT03_E_adendum_Offseting%20calculations_Examples.pdf.

IdeaWorksCompany. 2025. “Mission Possible: Airlines Earn Record Ancillary Revenue While Consumers Enjoy Lower Fares.” Press release 197. IdeaWorksCompany. https://ideaworkscompany.com/wp-content/uploads/2025/11/Press-Release-197-Global-Estimate-2025.pdf.

International Air Transport Association. 2021. “Dynamic Offers: The Road to Customer Centric Airline Retailing. Understanding the Difference Between Dynamic Pricing, Continuous Pricing and Dynamic Bundling.” Factsheet published by IATA; examples credited to Lufthansa Group. IATA. https://www.iata.org/contentassets/0688c780d9ad4a4fadb461b479d64e0d/dynamic-pricing_continuous-pricing_dynamic-bundling.pdf.

International Civil Aviation Organization. n.d. “Freedoms of the Air.” ICAO. Accessed October 28, 2026. https://icao.int/pages/freedomsair.aspx.

InterVISTAS Consulting. 2007. “Estimating Air Travel Demand Elasticities: Final Report.” Report prepared for IATA. International Air Transport Association. https://www.iata.org/en/iata-repository/publications/economic-reports/estimating-air-travel-demand-elasticities—by-intervistas.

Investing.com. 2026. “Air Canada at CIBC Eastern Institutional Investor Conference: Fuel Pain, Growth Plan.” Conference transcript. Investing.com, September 24, 2026. https://www.investing.com/news/transcripts/air-canada-at-cibc-eastern-institutional-investor-conference-fuel-pain-growth-plan-93CH-4915468.

Irnich, Stefan, and Guy Desaulniers. 2005. “Shortest Path Problems with Resource Constraints.” In Column Generation, edited by Guy

Desaulniers, Jacques Desrosiers, and Marius M. Solomon, 33–65. New York: Springer.

Jiang, Hai. 2006. "Dynamic Airline Scheduling and Robust Airline Schedule de-Peaking." Ph.D. thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology. https://dspace.mit.edu/handle/1721.1/37976.

Koppelman, Frank S., Gregory M. Coldren, and Roger A. Parker. 2008. "Schedule Delay Impacts on Air-Travel Itinerary Demand." Transportation Research Part B: Methodological 42 (3): 263–73.

Lan, Shan, John-Paul Clarke, and Cynthia Barnhart. 2006. "Planning for Robust Airline Operations: Optimizing Aircraft Routings and Flight Departure Times to Minimize Passenger Disruptions." Transportation Science 40 (1): 15–28. https://doi.org/10.1287/trsc.1050.0134.

Land, A. H., and A. G. Doig. 1960. "An Automatic Method of Solving Discrete Programming Problems." Econometrica 28 (3): 497–520. https://doi.org/10.2307/1910129.

Lee, Tak C., and Marvin Hersh. 1993. "A Model for Dynamic Airline Seat Inventory Control with Multiple Seat Bookings." Transportation Science 27 (3): 252–65.

Leeham News. 2026. "A220 Engine Problems Nearly over, Say Airbus and Pratt & Whitney." Leeham News and Analysis. July 7, 2026. https://leehamnews.com/2026/07/07/a220-engine-problems-nearly-over-say-airbus-and-pratt-whitney/.

Legal 500. 2026. "European Commission Proposes to Extend EU ETS to International Flights Beyond Europe, Puts CORSIA on Trial." Legal 500 Intelligence. August 13, 2026. https://www.legal500.com/intelligence/european-union/environment/european-commission-proposes-to-extend-eu-ets-to-international-flights-beyond-europe-puts-corsia-on-trial.

Lexology. n.d. "Competition Bureau Reaches Settlement with Air Canada and United Continental." Lexology. Accessed October 27, 2026. https://www.lexology.com/library/detail.aspx?g=d1f7f62d-e91b-4040-9a35-f6cf192c5f39.

Lhéritier, Alix, Michael Bocamazo, Thierry Delahaye, and Rodrigo Acuna-Agost. 2019. "Airline Itinerary Choice Modeling Using Machine Learning." Journal of Choice Modelling 31: 198–209. https://doi.org/10.1016/j.jocm.2018.02.002.

Lindley, Dennis V. 1952. "The Theory of Queues with a Single Server." Mathematical Proceedings of the Cambridge Philosophical Society 48 (2): 277–89.

Littlewood, Kenneth. 1972. "Forecasting and Control of Passenger Bookings." In AGIFORS Symposium Proceedings, 12:95–117.

Lohatepanont, Manoj, and Cynthia Barnhart. 2004. "Airline Schedule Planning: Integrated Models and Algorithms for Schedule Design and Fleet Assignment." Transportation Science 38 (1): 19–32. https://doi.org/10.1287/trsc.1030.0026.

Lohr, Sharon L. 2021. Sampling: Design and Analysis. 3rd ed. Boca Raton: Chapman and Hall/CRC.

Long, Yanbin, and Peter Belobaba. 2023. "Airline Revenue Management with Segmented Continuous Pricing: Methods and Competitive Effects." Journal of Revenue and Pricing Management, December. https://doi.org/10.1057/s41272-023-00462-6.

Lübbecke, Marco E., and Jacques Desrosiers. 2005. "Selected Topics in Column Generation." Operations Research 53 (6): 1007–23.

Lufthansa Systems. 2025a. "Air Canada Enhances Operational Efficiency with NetLine/Ops ++." Press release. https://cdn.lhsystems.com/2025-01/PressNews_Lufthansa-Systems_Air-Canada.pdf.

———. 2025b. "NetLine/FleetAssigner: Product Information." Product brochure. Lufthansa Systems. https://cdn.lhsystems.com/2025-02/2025_02_Product_information_NetLine_FleetAssigner.pdf.

———. 2025c. "NetLine/Plan: Product Information." Product brochure. Lufthansa Systems. https://cdn.lhsystems.com/2025-02/2025_02_Product_information_NetLine_Plan.pdf.

———. 2025d. "Schedule Optimization: Product Information." Product brochure. Lufthansa Systems. https://cdn.lhsystems.com/2025-02/2025_02_Product_information_Schedule_Optimization.pdf.

———. 2026. "NetLine/Plan." Lufthansa Systems. October 12, 2026. https://www.lhsystems.com/solutions/network-plan-schedule/netlineplan.

Lurkin, Virginie, Laurie A. Garrow, Matthew J. Higgins, Jeffrey P. Newman, and Michael Schyns. 2017. "Accounting for Price Endogeneity in Airline Itinerary Choice Models: An Application to Continental U.S. Markets." Transportation Research Part A: Policy and Practice 100: 228–46. https://doi.org/10.1016/j.tra.2017.04.007.

McCarthy Tétrault. 2026. "Updates to the Legislative Framework Governing Air Carriers Complaints: An Overview of Bill C-31." Mondaq. July 13, 2026. https://www.mondaq.com/canada/aviation/1815564/updates-to-the-legislative-framework-governing-air-carriers-complaints-an-overview-of-bill-c-31.

McFadden, Daniel. 1974. "Conditional Logit Analysis of Qualitative Choice Behavior." In Frontiers in Econometrics, edited by Paul Zarembka, 105–42. New York: Academic Press.

———. 1978. "Modelling the Choice of Residential Location." In Spatial Interaction Theory and Planning Models, edited by Anders Karlqvist, Lars Lundqvist, Folke Snickars, and Jörgen W. Weibull, 75–96. Amsterdam: North-Holland.

McFadden, Daniel, and Kenneth Train. 2000. "Mixed MNL Models for Discrete Response." Journal of Applied Econometrics 15 (5): 447–70.

McGill, Jeffrey I., and Garrett J. van Ryzin. 1999. “Revenue Management: Research Overview and Prospects.” Transportation Science 33 (2): 233–56.

Michell, Simon. 2026. “Unfriendly Skies: Why Canada Needs Real Airline Competition.” C2C Journal, March 2, 2026. https://c2cjournal.ca/2026/03/unfriendly-skies-why-canada-needs-real-airline-competition/.

Ministère de la Justice du Québec. 2024. “Daylight Savings: For or Against? (Public Consultation, 22 October to 1 December 2024).” Consultation Québec. October 22, 2024. https://consultation.quebec.ca/processes/changement-heure/embed.html?locale=en.

National Research Council Canada. 2025. “Time Zones and Daylight Saving Time.” Canada’s Official Time. November 1, 2025. https://nrc.canada.ca/en/certifications-evaluations-standards/canadas-official-time/time-zones-daylight-saving-time.

NAV CANADA. 2023. “Montréal / Pierre Elliott Trudeau Intl (CYUL): Noise Abatement Procedures.” Canada Air Pilot, page CYUL-NOR-2 effective 30 November 2023, in the cycle effective 29 October 2026 (reproduced by FltPlan). https://imageserver.fltplan.com/merge/Canada/merge261029/Single/CYUL_NOISE_ABATEMENT_PROCEDURES_PG_02.pdf.

Nemhauser, George L., and Laurence A. Wolsey. 1988. Integer and Combinatorial Optimization. New York: Wiley.

Network Planning Handbook. 2026. “Network Planning: Economics, Schedules and the Decisions Behind an Airline Network.” Independent preparation handbook prepared for Enzo Simier.

Nevo, Aviv. 2000. “A Practitioner’s Guide to Estimation of Random-Coefficients Logit Models of Demand.” Journal of Economics & Management Strategy 9 (4): 513–48.

OAG. n.d. “Air Traffic Analyser.” OAG. Accessed October 27, 2026. https://www.oag.com/traffic-analyser.

Open Jaw. 2026a. “Air Canada Launches Basic Fares for Premium Cabins.” Open Jaw, July 28, 2026. https://openjaw.com/newsroom/airline/2026/07/28/air-canada-launches-basic-fares-for-premium-cabins/.

———. 2026b. “Cracks in Demand as Jet Fuel Hits Record US$206 a Barrel.” Open Jaw, October 26, 2026. https://openjaw.com/newsroom/airline/2026/10/26/cracks-in-demand-as-jet-fuel-hits-record-us206-a-barrel.

openparliament.ca. 2026. “Bill C-31 (45th Parliament, 1st Session): Budget 2025 Implementation Act, No. 2.” Openparliament.ca. June 10, 2026. https://openparliament.ca/bills/45-1/C-31/.

Panzar, John C., and Robert D. Willig. 1981. “Economies of Scope.” American Economic Review 71 (2): 268–72.

Parliament of Canada. 2026. “C-31, 45th Parliament, 1st Session: Budget 2025 Implementation Act, No. 2.” LEGISinfo. 2026. https://www.parl.ca/legisinfo/en/bill/45-1/C-31.

PAX News. 2026. “Oslo, Shannon, Nice, Basel & Dubrovnik: Air Canada Unveils Major Summer 2027 Expansion.” PAX News. September 3, 2026. https://www.paxnews.com/news/airline/oslo-shannon-nice-basel-dubrovnik-air-canada-unveils-major-summer-2027-expansion.

Petrin, Amil, and Kenneth Train. 2010. “A Control Function Approach to Endogeneity in Consumer Choice Models.” Journal of Marketing Research 47 (1): 3–13.

PhocusWire. 2025. “Delta Encouraged by Early Results of AI Pricing with Fetcherr.” PhocusWire. July 14, 2025. https://www.phocuswire.com/delta-fetcherr-ai-pricing-deployment.

———. n.d. “The Smart AI Behind Effective Airline Revenue Management.” PhocusWire (Sponsored Content from PROS). Accessed October 27, 2026. https://www.phocuswire.com/pros-smart-science-effective-airline-revenue-management.

Prince of Travel. 2026. “How to Choose the Right Air Canada Fare.” Prince of Travel. July 30, 2026. https://princeoftravel.com/guides/how-to-choose-the-right-air-canada-fare.

PROS. 2025. “Air Canada Elevates the Customer Experience with Pricing Science.” PROS Case Studies. June 5, 2025. https://pros.com/learn/case-studies-testimonials/air-canada-elevates-customer-experience/.

PROS Holdings. 2021. “PROS Willingness-to-Pay (WTP) Forecasting and Optimization Technology Powers the Future of Airline Revenue Management.” Press release. Business Wire. https://www.businesswire.com/news/home/20210824005135/en.

Puterman, Martin L. 1994. Markov Decision Processes: Discrete Stochastic Dynamic Programming. New York: Wiley.

PYMNTS. 2026. “Air Canada, United Seek Wider Antitrust Immunity for Cross-Border Flights.” PYMNTS, CPI Posts. October 26, 2026. https://www.pymnts.com/cpi-posts/air-canada-united-seek-wider-antitrust-immunity-for-cross-border-flights/.

Québec. 2007. Legal Time Act, CQLR c T-5.1. https://www.legisquebec.gouv.qc.ca/en/document/cs/T-5.1.

ResumeGeni. 2026. “Schedule Optimization Analyst, Air Canada, Dorval, Quebec.” Job posting reproduced by a job aggregator. June 6, 2026. https://resumegeni.com/jobs/schedule-optimization-analyst-at-air-canada-dorval-quebec.

Rexing, Brian, Cynthia Barnhart, Tim Kniker, Ahmad Jarrah, and Nirup Krishnamurthy. 2000. “Airline Fleet Assignment with Time Windows.” Transportation Science 34 (1): 1–20.

Rivers, Douglas, and Quang H. Vuong. 1988. "Limited Information Estimators and Exogeneity Tests for Simultaneous Probit Models." Journal of Econometrics 39 (3): 347–66.

Rothstein, Marvin. 1985. "OR and the Airline Overbooking Problem." Operations Research 33 (2): 237–48.

Rushmeier, Russell A., and Spyridon A. Kontogiorgis. 1997. "Advances in the Optimization of Airline Fleet Assignment." Transportation Science 31 (2): 159–69. https://doi.org/10.1287/trsc.31.2.159.

Ryan, David M., and Brian A. Foster. 1981. "An Integer Programming Approach to Scheduling." In Computer Scheduling of Public Transport: Urban Passenger Vehicle and Crew Scheduling, edited by Anthony Wren, 269–80. Amsterdam: North-Holland.

SAF Investor. 2026. "EASA Publishes 2025 Reference Prices for ReFuelEU Aviation." SAF Investor. March 4, 2026. https://www.safinvestor.com/news/149690/easa-2025/.

Sherali, Hanif D., Ebru K. Bish, and Xiaomei Zhu. 2006. "Airline Fleet Assignment Concepts, Models, and Algorithms." European Journal of Operational Research 172 (1): 1–30. https://doi.org/10.1016/j.ejor.2005.01.056.

Simon, Julian L. 1968. "An Almost Practical Solution to Airline Overbooking." Journal of Transport Economics and Policy 2 (2): 201–2.

Skies Magazine. 2024. "Air Canada Plans to Acquire 90 New Aircraft in Push for $30B Revenue by 2028." Skies Magazine, December 19, 2024. https://skiesmag.com/news/air-canada-plans-to-acquire-90-new-aircraft-in-push-for-30b-revenue-by-2028/.

Small, Kenneth A., and Harvey S. Rosen. 1981. "Applied Welfare Economics with Discrete Choice Models." Econometrica 49 (1): 105–30.

Smith, Barry C., John F. Leimkuhler, and Ross M. Darrow. 1992. "Yield Management at American Airlines." Interfaces 22 (1): 8–31.

Smyth, Mark, and Brian Pearce. 2008. "Air Travel Demand: Measuring the Responsiveness of Air Travel Demand to Changes in Prices and Incomes." IATA Economics Briefing No. 9. International Air Transport Association. https://www.iata.org/en/iata-repository/publications/economic-reports/air-travel-demand.

Staiger, Douglas, and James H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." Econometrica 65 (3): 557–86.

Subramanian, Radhika, Richard P. Scheff, John D. Quillinan, D. Steve Wiper, and Roy E. Marsten. 1994. "Coldstart: Fleet Assignment at Delta Air Lines." Interfaces 24 (1): 104–20. https://doi.org/10.1287/inte.24.1.104.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." Journal of Econometrics 225 (2): 175–99.

Swan, William M. 2002. "Airline Demand Distributions: Passenger Revenue Management and Spill." Transportation Research Part E: Logistics and Transportation Review 38 (3-4): 253–63.

Sylvera. 2025. "CORSIA Phase 1 Pricing: What Airlines Need to Know about Costs." Sylvera Blog. July 25, 2025. https://www.sylvera.com/blog/corsia-phase-1-pricing-what-airlines-need-to-know-about-costs.

Szymański, Bazyli, Peter P. Belobaba, and Alexander Papen. 2021. "Continuous Pricing Algorithms for Airline RM: Revenue Gains and Competitive Impacts." Journal of Revenue and Pricing Management 20 (6): 669–88. https://doi.org/10.1057/s41272-021-00350-x.

Talluri, Kalyan T., and Garrett J. van Ryzin. 1998. "An Analysis of Bid-Price Controls for Network Revenue Management." Management Science 44 (11): 1577–93.

———. 1999. "A Randomized Linear Programming Method for Computing Network Bid Prices." Transportation Science 33 (2): 207–16.

———. 2004a. "Revenue Management Under a General Discrete Choice Model of Consumer Behavior." Management Science 50 (1): 15–33.

———. 2004b. The Theory and Practice of Revenue Management. New York: Springer.

The Canadian Press. 2026. "Air Canada Suspends 2026 Full-Year Guidance Amid Uncertain Jet Fuel Costs." BNN Bloomberg, April 30, 2026. https://www.bnnbloomberg.ca/business/company-news/2026/04/30/air-canada-suspends-2026-full-year-guidance-amid-uncertain-jet-fuel-costs/.

Tirole, Jean. 1988. The Theory of Industrial Organization. Cambridge, MA: MIT Press.

Tobin, James. 1958. "Estimation of Relationships for Limited Dependent Variables." Econometrica 26 (1): 24–36.

Train, Kenneth E. 2009. Discrete Choice Methods with Simulation. 2nd ed. Cambridge: Cambridge University Press.

Transat A.T. 2026. "Transat A.T. Inc. Reports Results for the Third Quarter of Fiscal 2026." Transat News Releases. September 10, 2026. https://www.transat.com/en-CA/corporate/media/news-releases/124622.

Transport Canada. 2024. "Flight Crew Member Fatigue Management: Prescriptive Regulations." Advisory Circular AC 700-047, issue 05. Transport Canada. https://tc.canada.ca/en/aviation/reference-centre/advisory-circulars/advisory-circular-ac-no-700-047.

———. 2026. "Sustainable Aviation Fuels Blueprint for Canada." Transport Canada. https://tc.canada.ca/sites/default/files/2026-07/sustainable-aviation-fuels-blueprint-canada.pdf.

———. n.d. "The 2006 Blue Sky Policy: Objectives, Principles and Approach." Transport Canada. Accessed October 27, 2026. https://tc.canada.ca/en/corporate-services/policies/2006-blue-sky-policy-objectives-principles-approach.

Travel Daily. 2020. "Lufthansa Dynamic Pricing." Travel Daily. October 9, 2020. https://traveldaily.com.au/news/lufthansa-dynamic-pricing/352725.

Travel Distribution News. 2026. “ARC CEO Lauri Reishus: NDC Hits 21.6% as Orders-Based Settlement Goes Live.” Travel Distribution News, June 17, 2026. https://traveldistributionnews.com/arc-ceo-lauri-reishus-ndc-hits-21-6-as-orders-based-settlement-goes-live/.

Travelweek. 2018. “Air Canada Revamps Economy Options with New Comfort Fare.” Travelweek, April 4, 2018. https://www.travelweek.ca/news/air-canada-revamps-economy-options-with-new-comfort-fare/.

———. 2024. “Air Canada Makes Changes to DCR Fee Policy, and Other Updates After a Year of NDC.” Travelweek, June 17, 2024. https://www.travelweek.ca/news/air-canada-makes-changes-to-dcr-fee-policy-and-other-updates-after-a-year-of-ndc/.

———. 2026a. “Transport Canada Opens Public Consultation on Air Canada-United Joint Venture Expansion.” Travelweek, October 19, 2026. https://www.travelweek.ca/news/lead-story/transport-canada-opens-public-consultation-on-air-canada-united-joint-venture-expansion/.

———. 2026b. “Jackson, Eldebs Talk Future Growth and the 5-Year Plan as Porter Airlines Marks 20th Anniversary.” Travelweek, October 23, 2026. https://travelweek.ca/news/airlines/jackson-eldebs-talk-future-growth-and-the-5-year-plan-as-porter-airlines-marks-20th-anniversary.

U.S. Department of Transportation. 2024. “Alliances and Codeshares.” U.S. Department of Transportation, Office of Aviation Analysis. December 16, 2024. https://www.transportation.gov/policy/aviation-policy/competition-data-analysis/alliance-codeshares.

Vance, Pamela H., Cynthia Barnhart, Ellis L. Johnson, and George L. Nemhauser. 1997. “Airline Crew Scheduling: A New Formulation and Decomposition Algorithm.” Operations Research 45 (2): 188–200.

Varian, Hal R. 1989. “Price Discrimination.” In Handbook of Industrial Organization, edited by Richard Schmalensee and Robert D. Willig, 1:597–654. Amsterdam: North-Holland.

Virtanen, Pauli, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, et al. 2020. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python.” Nature Methods 17 (3): 261–72. https://doi.org/10.1038/s41592-019-0686-2.

Weatherford, Larry R., and Stefan Pölt. 2002. “Better Unconstraining of Airline Demand Data in Revenue Management Systems for Improved Forecast Accuracy and Greater Revenues.” Journal of Revenue and Pricing Management 1 (3): 234–54.

WestJet. 2025. “Onex Partners and Its Affiliated Funds and Co-Investors Close Sale of 25 Per Cent Equity Stake in WestJet to Delta, Korean Air and Air France-KLM.” WestJet Mediaroom. October 22, 2025. https://westjet.mediaroom.com/2025-10-22-Onex-Partners-and-its-affiliated-funds-and-co-investors-close-sale-of-25-per-

cent-equity-stake-in-WestJet-to-Delta,-Korean-Air-and-Air-France-KLM.

Wickramasuriya, Shanika L., George Athanasopoulos, and Rob J. Hyndman. 2019. "Optimal Forecast Reconciliation for Hierarchical and Grouped Time Series Through Trace Minimization." Journal of the American Statistical Association 114 (526): 804–19.

Williams, Kevin R. 2022. "The Welfare Effects of Dynamic Pricing: Evidence from Airline Markets." Econometrica 90 (2): 831–58.

Williamson, Elizabeth L. 1992. "Airline Network Seat Inventory Control: Methodologies and Revenue Impacts." Ph.D. thesis, Massachusetts Institute of Technology.

Wittman, Michael D., and Peter P. Belobaba. 2019. "Dynamic Pricing Mechanisms for the Airline Industry: A Definitional Framework." Journal of Revenue and Pricing Management 18 (2): 100–106.

Wolsey, Laurence A. 1998. Integer Programming. New York: Wiley.

Wooldridge, Jeffrey M. 2010. Econometric Analysis of Cross Section and Panel Data. 2nd ed. Cambridge, MA: MIT Press.

Working, E. J. 1927. "What Do Statistical 'Demand Curves' Show?" Quarterly Journal of Economics 41 (2): 212–35.

Yeoman, Ian. 2025. "Expanding the Frontiers of Revenue and Pricing Management." Journal of Revenue and Pricing Management 24 (6): 503–5. https://link.springer.com/article/10.1057/s41272-025-00563-4.

Index

Symbols

80 percent test, 351

A

acyclic, 698
additive pickup, 209
adjusted CASM, 37
adjusted demand, 567
adjusted fare, 567
advance purchase, 462
aircraft and crew, 357
aircraft movement, 30, 638
aircraft operating, 54
aircraft productivity, 64
aircraft rotation planning, 4, 330
aircraft routing, 10
aircraft servicing, 54
aircraft time, 348
aircraft utilization, 64
airline model, 201
airline O&D, 196
airport slot, 350
airport-pair market, 82
alternative, 15
alternative-specific constants, 111
ancillary, 467
anticipation, 189
arc elasticity, 86
arrival node, 363
assignment cost, 363
assumption register, 645
authorization level, 505
available seat-kilometres, 34
available seat-miles (ASM), 34
average fare, 369
average load factor, 228
average passenger trip length, 39
average stage length, 39
average-traveller error, 132
avoidable, 18
avoidable cost, 68

B

backward induction, 700
bank, 23, 253, 340
bank peak, 357
base forecasts, 206
basic feasible solution, 686
basic solution, 686
basis, 686
behind, 255
behind the gateway, 252
Bellman equation, 700
Bellman's equations, 698
beta-binomial, 507
between two spokes, 252
beyond, 255
bid price, 239, 534
bid price times seats, 586
bidline, 409
bid-price control, 534
binding resources, 15
block hours, 56
block time, 398
Blue Sky policy, 265
Boeing's examples, 152
booking class, 461
booking curve, 208, 495
booking limit, 482
booking profile, 233
bottom-up, 206, 276
branch and bound, 694
branch and cut, 696
branch and price, 706
branch-and-price, 407
branded fare, 467
break errors down, 204
break-even, 44, 216
break-even load factor, 71
breguet range equation, 282
bundle constraint, 699
business, 85
buy down, 238
buy-down, 556

C

cabotage, 264
calendar effects, 201
calendar time, 56
calibration, 143, 160
capacity, 222
capacity changes, 178
capacity cost, 57
capacity gap, 277
capital and strategic case, 303

censored, 223, 681
censored maximum likelihood, 231
central limit theorem, 676
check the data's support, 192
check the intervals, 205
choice probability, 111
choice set, 160
circuity, 248
circulation, 698
city-pair market, 82
codeshare duplicate, 117
coefficient of variation, 666
coherent, 205, 206
coherent scenarios, 320
column generation, 404, 704
committed, 68
company, 25
competition, 12
competitor-change note, 657
concave, 226
concentration, 266
conditional expectation, 666
confidence interval, 682
connecting passengers displace locals, 529
connection arc, 364
connection window, 340
connections, 356
consequences, 181
consistency, 161
constant market share, 353
constrained demand, 598
consumer surplus, 131
continuity correction, 676
continuous pricing, 610
contribution, 596
contribution value, 584
control, 480
convexity constraint, 704
coordinator, 350
correlation, 666
CORSIA, 625
cost driver, 56
Cost per ASK (CASK), 37
cost per passenger, 61
cost shifters, 177
count line, 363
count time, 395
counterfactual, 15
counting markets, 257
counting test passes, 397
coupon, 30
covariance, 666
coverage state, 151
credit, 400
crew hours, 368
crew scheduling, 6, 11
crews, 414
critical path, 292
cumulative flight time limits, 412
curve, 286
cut, 695
cutting planes, 695

D

damaged goods, 459
data, 171, 199, 210, 212, 479
data contract, 645
dates, 348
decision threshold, 654
decision window, 151
decompose, 7
delay, 439
Delta-T, 151
demand, 170, 336, 356
demand capture, 321
demand driven dispatch, 384
demand factor, 228
demand shock, 164
denied boarding, 505
departure node, 363
departures, 56
deterministic linear program, 536
deterministic LP, 588
deterministic LP re-solved, 587
deterministic rule, 508
dichotomy of demand and supply, 83
difference in differences, 601
difference-in-differences, 188
difficult 2026, 26
direction, 318
displacement cost, 307, 532, 584
displacement-adjusted virtual nesting, 532
disrupted, 439
dissimilarity parameter, 141
distance, 56, 286
disutility, 458
disutility cost, 462
diverted, 103
dual, 688
duty 1, 401
duty 2, 401
duty period, 398
dynamic bundling, 610
dynamic pricing, 610
dynamic programming, 700
dynamic regression, 201

E

easter, 200
economies of scope, 249
efficient, 571
efficient frontier, 571
eighth freedom, 264
elastic, 86
elasticities, 126
EM algorithm, 231
employee productivity, 64
EMSR-a, 486
EMSR-b, 486
endogenous, 165
entry, 148
entry and fares, 266
equivalent still-air distance, 285
error components, 146
E-step, 232
estimates, 213
estimating feed, 258
estimation, 160, 181
EU ETS, 625
event study, 188
ex post bid price, 585
ex post optimum, 587
excluded, 173
exogenous, 165
expected marginal seat revenue, 486, 541
expected schedule delay, 331
extended crew pairing, 410
extended diversion time operations, 288

F

facilitator, 350
fan chart, 216
fare basis, 461
fare class control, 528
fare product, 461
fare response, 148
fares, 170, 599
feasible, 15
feasible connection, 340
fence, 457
fenced, 591
first-come, first-served, 237
first-degree, 454
Fisher information, 677
fit, 199
five-hour weeks, 347
fleet assignment, 4, 10, 330
fleet assignment problem, 362
fleet assignment with time windows, 353
fleet commonality, 290
fleet plan, 274
fleet planning, 4, 5
fleet transition, 26
flexible fare, 88
flight, 4, 30, 330
flight arc, 363
flight contribution, 303
flight duty period limits, 412
flight leg, 4, 30, 244, 330, 638
flight operating costs, 55
flight selection, 354
flow, 255
follow-on branching, 408
forecast, 200, 211
forecasting, 132, 480
frequency and demand, 294
frequency elasticity, 294
frequency planning, 4, 330
frequency share, 96
fuel capacity, 282
fuel-limited, 283
full, 369
fully allocated cost, 68
functional, 54

G

gap, 695
gateway, 245
generalized cost, 93, 121, 462
GEV generating function, 139
go-show, 505
grain, 638
granularity, 583
gravity model, 212
greedy, 532
ground arcs, 363
ground operating costs, 55
grouped, 208
Gumbel, 112, 674

H

Hausman instruments, 178
historic precedence, 350
historical, 209
Historics Baseline Date, 350
Holt–Winters, 198
horizon, 599
hub premium, 268
hub-and-spoke network, 245
hubs, 245
hubs and network, 25

I

IATA scheduling season, 349
illustrative, 285
incidence, 529
inclusive value, 140
increment, 15
incremental, 15, 330
incremental cost, 68
incremental network contribution, 18, 303
incremental schedule design, 353
incumbent, 695
independence of irrelevant alternatives, 117
independent arrival delay, 422
independent departure delay, 422
indivisibility, 12
inelastic, 86
information, 583
inputs, 294
instrument, 171, 173
integer program, 692
integrality gap, 693
integrality property, 702
integrated image, 152
integration, 369
interior-point methods, 687
interval matrix, 697
inverse Mills ratio, 231, 668
inversion, 181
involuntary denied boarding, 519
islands, 372
itinerary, 4, 30, 83, 244, 330
itinerary and fare product, 196
itinerary elasticity and market elasticity, 120
itinerary-based fleet assignment model, 381
itinerary-based model, 382

J

Jensen's inequality, 666
journey, 15, 18, 30
judgment, 508

K

k-factor, 223

L

label, 405
Lagrangian bound, 405
Lagrangian dual, 701
Lagrangian function, 701
Lagrangian relaxation, 701
latent class, 146
Laurentian's assumptions, 317
Laurentian's market, 153
law of large numbers, 676
layover, 398
leg, 196, 529
leg arithmetic, 588
leg demand, 83
leg view, 595
leg-based model, 382
leisure, 85
level 1, 350
level 2, 350
level 3, 350
likelihood ratio, 124
linear program, 686
lines as built fail, 397
linking constraints, 703
load, 222, 239
load consolidation, 245
load factor, 34
local, 255
logistic, 114, 675
log-likelihood, 123
lognormal, 668
logsum, 129
longer route, 287
long-range narrowbody, 620
loss function, 668
LP relaxation, 693

M

maintenance, 368, 414
marginal, 537
marginal revenue transformation, 567
marginal utility of money, 121
marginal value of a seat, 490, 492
market, 196
market characteristics, 84
market elasticity, 120
market share, 96
market share equation, 152
market size, 180
master flight list, 353
master problem, 704
maximum landing weight, 282
maximum simulated likelihood, 147
maximum take-off weight, 282
maximum zero-fuel weight, 282
max-stability, 114
mean absolute error, 202
mean absolute percentage error, 202

mean absolute scaled error, 202
mean error, 202
metal neutrality, 260
mileage prorate, 305
minimum connecting time, 253, 340
minimum rest periods, 413
minimum stay, 462
minimum-cost flow problem, 698
missions, 286
mixed logit, 146
mixed-integer program, 692
mixing distribution, 146
model, 160, 375
modelling language, 707
money values, 126
Montreal to Seville, 26
M-step, 232
multicommodity flow, 699
multiplicative pickup, 209

N

naive, 198
naive estimators, 231
naive plan, 375
naive regression, 171
negative binomial, 673
nested logit, 140
network, 363, 601
network at mean demand, 596
network contribution, 18, 307
network displacement cost, 307
network effects, 12, 378
network value, 306, 532
New Distribution Capability (NDC), 613
new entrant, 350, 352
ninth freedom, 264
no fences, 556
node-arc formulation, 699
normally distributed, 224
no-show, 505
no-show rate, 505

O

objective, 54, 583
O&D and leg, 599
one 787-9, 336
one input at a time, 320
one market, one season, 295
ONE Order, 613
only differences in utility matter, 111
operating empty weight, 282
operations control, 6
opportunity cost, 68
optimization, 480
optimum, 375
origin and destination survey, 642
origin-destination fare product, 529
origin-destination (O&D) control, 531
origin-destination (O&D) market, 30, 83, 244
other, 54
outside option, 110
overbooking factor, 505
overdispersed, 224

P

pairing, 398, 401
parallel trends, 188
Paris, 92
partial, 369
partitioned, 481
partitioned forecast, 561
passenger connections, 414
passenger itinerary, 638
passenger mix model, 379, 380
Passenger revenue per ASK (PRASK), 37
passenger service, 54
passengers, 56
passengers rank the routings differently, 433
path formulation, 699
path quality index, 151
pattern, 319
payload, 282
peak, 315
per round trip, 318
perfect fences, 556
perishable, 11
perishable seats, 11
pickup, 208
pieces, 179
planned turn time, 422
point-to-point network, 245
Poisson process, 672
polyhedron, 686
pool, 350
post-launch review, 6, 657
prediction, 213
prediction interval, 215, 682
price, 85
price discrimination, 452
price elasticity, 86
price-and-branch, 408
price-oriented, 555
pricing, 4, 453
pricing and revenue management, 6

pricing problem, 404, 704
principle of optimality, 700
probabilistic rule, 508
product, 85
product differentiation, 453
product-oriented, 555
program, 295
propagated delay, 422
propagated delay barely moves, 433
prorate, 305
protection level, 482
pruning by bound, 695

Q

Q-forecasting, 560
quality of service index, 158
quality of service index (QSI), 101
quantile forecast, 215

R

ramp-up, 214
random coefficients, 146, 182
random demand and revenue management, 596
random variable, 665
randomized linear program, 542
range factor, 282
ratio test, 687
reading the eight numbers, 588
reading the result, 131
ready time, 363
real schedule, 316
recalibrating constants, 132
recapture, 378
recapture rate, 236, 369
recommendations, 267
reduced cost, 404, 687, 704
ReFuelEU, 625
ReFuelEU Aviation, 622
refund and change conditions, 462
region-pair market, 82
regression, 209
regression with calendar effects, 201
regulation and access, 12
release, 398
relevant, 173
reliability, 357
repair, 415
report, 398
representative fare, 369
representative utility, 111
required through flight, 368
reservations and sales, 54
resource-constrained shortest path, 699
resource-constrained shortest path problem, 405
rest, 401
restricted fare, 88
restricted master problem, 404, 704
restriction disutilities, 156
results, 144
returns to density, 249
returns to scale, 249
revenue management, 4, 11, 170, 453
revenue passenger-kilometres, 34
revenue passenger-miles (RPM), 34
Revenue per ASK (RASK), 37
revenue value, 584
reversal frontier, 654
rival products, 177
RM rejections, 598
robust, 421
rolling, 343
rolling origins, 203
rolling-origin evaluation, 203
root mean squared error, 202
rostering, 409
rotation, 345
route, 330
route evaluation, 4
route planning, 6

S

sample enumeration, 132
saturation frequency, 91, 333
scale parameter, 116
scaling up, 131
scenarios, 215, 216
schedule card, 331
schedule delay, 90, 331
schedule design, 352
schedule development, 4, 6, 330
schedule tolerance, 151
score, 103, 677
score test, 678
S-curve, 96, 334
season, 319
seasonal naive, 198
seat cost, 61
second frequency, 153
second-degree, 454
segmentation, 132
selection, 189
sell-up, 556
sell-up probability, 557
sensitivity, 104, 296

serially nested, 481
series of slots, 350
set partitioning, 401
shadow price, 310, 618, 690
share, 104, 131
share model, 158
shared, 369
short connect, 410
short haul blocks long haul, 529
shortest path, 698
shoulder, 315
show-up, 505
show-up rate, 505
simple exponential smoothing, 198
simplex method, 687
simulated controls, 588
simulation, 147, 149
sits, 398
sixth-freedom traffic, 255
size grows fast, 397
slack, 422
slack variable, 686
sources, 104
spill, 222, 356, 598
spill factor, 228
spill model, 586
spill rate, 228
spill tables, 228
spillovers, 190
spiral down, 466
spiral-down effect, 563
spoiled seat, 505
spoke, 245
stage-length adjustment, 41
staggered timing, 190
staging problem, 274
standard form, 686
standard normal loss function, 225
standardization, 668
standby, 505
station rules, 368
step 1: revenue by journey, 21
step 2: the flight itself, 21
step 4: diversion, 21
step 4: margin and break-even, 47
step 5: regions, 47
step 5: the alternative, 21
stimulated, 103
stimulation, 104
stochastic delay, 331
straight-rate prorate, 305
string, 395
structural, 283
subgradient method, 702
summer time, 347
summing matrix, 206
sustainable aviation fuel (SAF), 622
swap points decide, 397
system operating costs, 55

T

take-off weight, 282
temporal, 208
thin flow, 204
third-degree, 454
three estimates, 179
threshold, 321
ticket, 30, 638
time, 85
time away from base, 398
time free from duty, 413
time-space, 363
timetable and time zones, 10
timetable development, 4, 330
Tobit, 681
top-down, 206, 276
Toronto, 92
total arrival delay, 422
total cost, 155
total trip time, 90
totally unimodular, 696
track bias separately, 204
traffic servicing, 54
transportation, 698
trap, 214
trip cost, 56, 61
truncated, 681
Two A220s, 336
two logits, upper and lower, 141

U

uncertainty, 12
unconstrain, 229, 230
unconstrained demand, 222, 598
unforeseen operational circumstances, 413
unit cost falls with distance, 40
unit measures, 44
United transborder agreement, 26
unobserved quality, 168
upgauge, 230
upper bound, 376
use it or lose it, 350
utilization, 345

V

valid inequality, 695
validation, 161
value bucket, 531
value of an aircraft, 376
value of integration, 9
value of time, 121
variable market share, 354
variance, 666
vertices, 686
virtual classes, 531
visiting friends and relatives (VFR), 85
voluntary denied boarding, 519

W

Wald test, 678
weight-limited, 283
where the difference comes from, 295
willingness to pay, 155, 554
winter, 315, 348
without fences, 591
wrap-around arc, 363

Y

yield, 37
yield falls with distance, 40
YUL–SVQ, 23

Z

zero-fuel weight, 282