

Inventory and Production Management in Supply Chains

Fourth Edition



Edward A. Silver • David F. Pyke • Douglas J. Thomas

Inventory and Production Management in Supply Chains

Fourth Edition



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Inventory and Production Management in Supply Chains

Fourth Edition

Edward A. Silver

University of Calgary (retired), Alberta, Canada

David F. Pyke

University of San Diego, California, USA

Douglas J. Thomas

Penn State University, Pennsylvania, USA



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2017 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed on acid-free paper
Version Date: 20160830

International Standard Book Number-13: 978-1-4665-5861-8 (Hardback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

Names: Silver, Edward A. (Edward Allen), 1937- author. | Pyke, D. F. (David F.) author. | Silver, Edward A. (Edward Allen), 1937- Decision systems for inventory management and production and planning. | Silver, Edward A. (Edward Allen), 1937- Inventory management and production planning and scheduling.
Title: Inventory and production management in supply chains / Edward A. Silver, David F. Pyke, Douglas J. Thomas.
Description: Fourth Edition. | Boca Raton : Taylor & Francis, 2017. | Revised edition of Inventory management and production planning and scheduling. | Includes index.
Identifiers: LCCN 2016022678 | ISBN 9781466558618 (hardback : alk. paper)
Subjects: LCSH: Inventory control--Decision making. | Production planning--Decision making.
Classification: LCC HD40 .S55 2017 | DDC 658.7/87--dc23
LC record available at <https://lccn.loc.gov/2016022678>

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

and the CRC Press Web site at
<http://www.crcpress.com>

Edward A. Silver dedicates this work to Maxine, Michelle, Norman, and Heidi

David F. Pyke dedicates this work to Susan, James, Daniel, and Cory Ad majorem Dei gloriam

Douglas J. Thomas dedicates this work to Traci, Alison, Kate, and Maya



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Contents

Preface	xix
Acknowledgments	xxiii
Authors	xxv

SECTION I THE CONTEXT AND IMPORTANCE OF INVENTORY MANAGEMENT AND PRODUCTION PLANNING

1 The Importance of Inventory Management and Production Planning and Scheduling	3
1.1 Why Aggregate Inventory Investment Fluctuates: The Business Cycle	7
1.2 Corporate Strategy and the Role of Top Management	8
1.3 The Relationship of Finance and Marketing to Inventory Management and Production Planning and Scheduling	10
1.3.1 Finance	10
1.3.2 Marketing	11
1.4 Operations Strategy	12
1.4.1 Mission	13
1.4.2 Objectives	13
1.4.3 Management Levers	15
1.4.4 General Comments	16
1.5 Measures of Effectiveness for Inventory Management and Production Planning and Scheduling Decisions	17
1.6 Summary	18
Problems	18
References	20
2 Frameworks for Inventory Management and Production Planning and Scheduling	23
2.1 The Diversity of Stock-Keeping Units	23
2.2 The Bounded Rationality of a Human Being	24
2.3 Decision Aids for Managing Diverse Individual Items	25
2.3.1 Conceptual Aids	25
2.3.2 Physical Aids	25
2.4 Frameworks for Inventory Management	26
2.4.1 Functional Classifications of Inventories	26
2.4.2 The A–B–C Classification as a Basis for Designing Individual Item Decision Models	28

2.5	A Framework for Production Planning and Scheduling	31
2.5.1	A Key Marketing Concept: The Product Life Cycle.....	31
2.5.2	Different Types of Production Processes	33
2.5.3	The Product-Process Matrix	37
2.6	Costs and Other Important Factors	40
2.6.1	Cost Factors.....	40
2.6.2	Other Key Variables	44
2.7	Three Types of Modeling Strategies	46
2.7.1	Detailed Modeling and Analytic Selection of the Values of a Limited Number of Decision Variables	47
2.7.2	Broader-Scope Modeling with Less Optimization	47
2.7.3	Minimization of Inventories with Little Modeling.....	47
2.8	The Art of Modeling	47
2.9	Explicit Measurement of Costs.....	49
2.10	Implicit Cost Measurement and Exchange Curves.....	52
2.11	The Phases of a Major Study of an Inventory Management or Production Planning and Scheduling System	53
2.11.1	Consideration	54
2.11.2	Analysis	55
2.11.3	Synthesis.....	57
2.11.4	Choosing among Alternatives	57
2.11.5	Control	58
2.11.6	Evaluation	58
2.11.7	General Comments.....	58
2.11.8	Transient Effects	59
2.11.9	Physical Stock Counts	59
2.12	Summary	61
	Problems	61
	Appendix 2A: The Lognormal Distribution	68
	References.....	70
3	Forecasting Models and Techniques	73
3.1	The Components of Time-Series Analysis	75
3.2	The Three Steps Involved in Statistically Forecasting a Time Series	77
3.3	Some Aggregate Medium-Range Forecasting Methods.....	78
3.3.1	Regression Procedures	79
3.4	Individual-Item, Short-Term Forecasting: Models and Procedures	81
3.4.1	The Simple Moving Average	82
3.4.2	Simple Exponential Smoothing.....	84
3.4.3	Exponential Smoothing for a Trend Model	88
3.4.4	Winters Exponential Smoothing Procedure for a Seasonal Model.....	92
3.4.5	Selection of Smoothing Constants	101
3.5	Measuring the Performance of a Forecasting Process	104
3.5.1	Measures of Forecast Accuracy	105
3.5.2	Estimating the Standard Deviation of Forecast Errors over a Lead Time	109
3.5.3	Monitoring Bias.....	111

3.5.4	Corrective Actions in Statistical Forecasting.....	115
3.5.5	Probability Distributions of Forecast Errors.....	117
3.6	Handling Anomalous Demand	117
3.7	Incorporation of Human Judgment.....	118
3.7.1	Factors Where Judgment Input Is Needed.....	118
3.7.2	Guidelines for the Input and Monitoring of Judgment	119
3.8	Dealing with Special Classes of Individual Items	120
3.8.1	Items with Limited History.....	120
3.8.2	Intermittent and Erratic Demand	122
3.8.3	Replacement or Service Parts	123
3.8.4	Terminal Demand	124
3.9	Assessing Forecasting Procedures: Tactics and Strategy.....	125
3.9.1	Statistical Accuracy of Forecasts.....	125
3.9.2	Some Issues of a More Strategic Nature	126
	Problems	128
	Appendix 3A: Derivations	135
	References.....	137

SECTION II REPLENISHMENT SYSTEMS FOR MANAGING INDIVIDUAL ITEM INVENTORIES WITHIN A FIRM

4	Order Quantities When Demand Is Approximately Level	145
4.1	Assumptions Leading to the Basic EOQ.....	146
4.2	Derivation of the EOQ.....	147
4.2.1	Numerical Illustration	151
4.3	Sensitivity Analysis.....	152
4.4	Implementation Aids.....	154
4.4.1	Numerical Illustration	155
4.5	Quantity Discounts	155
4.5.1	Numerical Illustrations	158
4.5.2	Item A (An Illustration of Case a of Figure 4.5).....	159
4.5.3	Item B (An Illustration of Case b of Figure 4.5).....	159
4.5.4	Item C (An Illustration of Case c of Figure 4.5).....	160
4.6	Accounting for inflation	160
4.6.1	Price Established Independent of Ordering Policy.....	161
4.6.2	Price Set as a Fixed Fractional Markup on Unit Variable Cost	163
4.7	Limits on order sizes.....	164
4.7.1	Maximum Time Supply or Capacity Restriction.....	164
4.7.2	Minimum Order Quantity	165
4.7.3	Discrete Units	165
4.8	Finite Replenishment Rate: The Economic Production Quantity	166
4.9	Incorporation of Other Factors	168
4.9.1	Nonzero Constant Lead Time That Is Known with Certainty	168
4.9.2	Nonzero Payment Period.....	169
4.9.3	Different Types of Carrying Charge	169

4.9.4	Multiple Setup Costs: Freight Discounts	170
4.9.5	A Special Opportunity to Procure	172
4.10	Selection of the Carrying Charge (r), the Fixed Cost per Replenishment (A), or the Ratio A/r Based on Aggregate Considerations: The Exchange Curve	176
4.10.1	Exchange Curve Illustration	177
4.11	Summary	179
	Problems	179
	Appendix 4A: Derivations	187
	References	193
5	Lot Sizing for Individual Items with Time-Varying Demand	199
5.1	The Complexity of Time-Varying Demand	200
5.2	The Choice of Approaches	201
5.3	General Assumptions and a Numerical Example	202
5.3.1	The Assumptions	202
5.3.2	A Numerical Example	203
5.4	Use of a Fixed EOQ	204
5.5	The Wagner-Whitin Method: An “Optimal” Solution under an Additional Assumption	205
5.5.1	The Algorithm	206
5.5.2	Potential Drawbacks of the Algorithm	209
5.6	Heuristic Approaches for a Significantly Variable Demand Pattern	212
5.6.1	The Silver–Meal, or Least Period Cost, Heuristic	212
5.6.2	The EOQ Expressed as a Time Supply (POQ)	216
5.6.3	Lot-for-Lot	216
5.6.4	Least Unit Cost	216
5.6.5	Part-Period Balancing	216
5.6.6	Performance of the Heuristics	218
5.6.7	When to Use Heuristics	219
5.6.8	Sensitivity to Errors in Parameters	220
5.6.9	Reducing System Nervousness	221
5.7	Handling of Quantity Discounts	221
5.8	Aggregate Exchange Curves	223
5.9	Summary	223
	Problems	223
	Appendix 5A: Dynamic Programming and Linear Programming Formulations	232
	References	233
6	Individual Items with Probabilistic Demand	237
6.1	Some Important Issues and Terminology	238
6.1.1	Different Definitions of Stock Level	238
6.1.2	Backorders versus Lost Sales	239
6.1.3	Three Key Issues to Be Resolved by a Control System under Probabilistic Demand	239
6.2	The Importance of the Item: A, B, and C Classification	240
6.3	Continuous versus Periodic Review	240
6.4	The Form of the Inventory Policy: Four Types of Control Systems	241

6.4.1	Order-Point, Order-Quantity (s, Q) System	242
6.4.2	Order-Point, Order-Up-to-Level (s, S) System	242
6.4.3	Periodic-Review, Order-Up-to-Level (R, S) System.....	243
6.4.4	(R, s, S) System.....	244
6.5	Specific Cost and Service Objectives	245
6.5.1	Choosing the Best Approach	246
6.5.2	SSs Established through the Use of a Simple-Minded Approach	246
6.5.3	SSs Based on Minimizing Cost	248
6.5.4	SSs Based on Customer Service	248
6.5.5	SSs Based on Aggregate Considerations	250
6.6	Two Examples of Finding the Reorder Point s in a Continuous-Review, Order-Point, Order-Quantity (s, Q) System	250
6.6.1	Protection over the Replenishment Lead Time	251
6.6.2	An Example Using a Discrete Distribution.....	252
6.7	Decision Rules for Continuous-Review, Order-Point, Order-Quantity (s, Q) Control Systems	256
6.7.1	Common Assumptions and Notation	257
6.7.2	General Approach to Establishing the Value of s	259
6.7.3	Common Derivation.....	260
6.7.4	Decision Rule for a Specified Safety Factor (k)	263
6.7.5	Decision Rule for a Specified Cost (B_1) per Stockout Occasion.....	263
6.7.6	Decision Rule for a Specified Fractional Charge (B_2) per Unit Short	266
6.7.7	Decision Rule for a Specified Fractional Charge (B_3) per Unit Short per Unit Time	268
6.7.8	Decision Rule for a Specified Charge (B_4) per Customer Line Item Short	269
6.7.9	Decision Rule for a Specified Probability (P_1) of No Stockout per Replenishment Cycle.....	269
6.7.10	Decision Rule for a Specified Fraction (P_2) of Demand Satisfied Directly from Shelf	271
6.7.11	Decision Rule for a Specified Average Time (TBS) between Stockout Occasions.....	273
6.7.12	Decision Rule for the Allocation of a TSS to Minimize the ETSOPY	274
6.7.13	Decision Rule for the Allocation of a TSS to Minimize the ETVSPY	274
6.7.14	Nonnormal Lead Time Demand Distributions	275
6.8	Implied Costs and Performance Measures	277
6.9	Decision Rules for Periodic-Review, Order-Up-to-Level (R, S) Control Systems.....	277
6.9.1	The Review Interval (R)	278
6.9.2	The Order-Up-to-Level (S)	278
6.9.3	Common Assumptions and Notation	280
6.9.4	Common Derivation.....	280
6.10	Variability in the Replenishment Lead Time Itself	282
6.10.1	Approach 1: Use of the Total Demand over the Full Lead Time.....	283

6.10.2	Approach 2: Use of the Distribution of Demand Rate per Unit Time Combined with the Lead Time Distribution	284
6.10.3	Nonnormal Distributions	285
6.11	Exchange Curves Involving SSs for (s, Q) Systems	286
6.11.1	Single Item Exchange Curve: Inventory versus Service.....	287
6.11.2	An Illustration of the Impact of Moving Away from Setting Reorder Points as Equal Time Supplies	288
6.11.3	Derivation of the SS Exchange Curves	290
6.11.4	Composite Exchange Curves	293
6.12	Summary	294
	Problems	295
	Appendix 6A: Some Illustrative Derivations and Approximations	304
	References.....	312

SECTION III SPECIAL CLASSES OF ITEMS

7	Managing the Most Important Inventories.....	319
7.1	Nature of Class A Items.....	319
7.2	Guidelines for Control of A Items.....	320
7.3	Simultaneous Determination of s and Q for Fast-Moving Items	322
7.3.1	Decision Rules.....	323
7.3.2	Cost Penalties.....	325
7.3.3	Further Comments	325
7.4	Decision Rules for (s, S) Systems	327
7.4.1	Simple Sequential Determination of s and S	328
7.4.2	Simultaneous Selection of s and S Using the Undershoot Distribution.....	328
7.4.3	Comparison of the Methods.....	331
7.5	Decision Rules for (R, s, S) Systems	332
7.5.1	Decision Rule for a Specified Fractional Charge (B_3) per Unit Short at the End of Each Period.....	332
7.5.2	Decision Rule for a Specified Fraction (P_2) of Demand Satisfied Directly from Shelf	334
7.6	Coping with Nonstationary Demand.....	337
7.7	Comments on Multiple Sources of Supply and Expediting	339
7.8	Summary	341
	Problems	341
	Appendix 7A: Simultaneous Solutions for Two Control Parameters.....	345
	References.....	346
8	Managing Slow-Moving and Low-Value (Class C) Inventories	351
8.1	Order-Point, Order-Quantity (s, Q) Systems for Slow-Moving A Items	351
8.1.1	B_2 Cost Measure for Very-Slow-Moving, Expensive Items ($Q = 1$).....	353
8.1.2	Case of $Q \geq 1$ and a B_1 Cost Structure.....	356
8.1.3	Simultaneous Determination of s and Q for Slow-Moving Items.....	356
8.2	Controlling the Inventories of Intermittent Demand Items.....	357

8.3	Nature of C Items	358
8.4	Control of C Items Having Steady Demand	359
8.4.1	Inventory Records	359
8.4.2	Selecting the Reorder Quantity (or Reorder Interval)	359
8.4.3	Selecting the Reorder Point (or Order-up-to Level)	360
8.4.4	Two-Bin System Revisited	361
8.4.5	Simple Form of the (R, S) System	362
8.4.6	Grouping of Items	363
8.5	Control of Items with Declining Demand Patterns	363
8.5.1	Establishing the Timing and Sizes of Replenishments under Deterministic Demand	363
8.5.2	Sizing of the Final Replenishment under Probabilistic Demand	364
8.6	Reducing Excess Inventories	365
8.6.1	Review of the Distribution by Value	366
8.6.2	Rule for the Disposal Decision	368
8.6.3	Options for Disposing of Excess Stock	370
8.7	Stocking versus Not Stocking an Item	371
8.7.1	Relevant Factors	371
8.7.2	Simple Decision Rule	372
8.7.3	Some Extensions	373
8.8	Summary	374
	Problems	374
	Appendix 8A: Poisson Distribution and Some Derivations	379
	References	384
9	Style Goods and Perishable Items	387
9.1	Style Goods Problem	388
9.2	Simplest Case: Unconstrained, Single-Item, Newsvendor Problem	389
9.2.1	Determination of the Order Quantity by Marginal Analysis	389
9.2.2	An Equivalent Result Obtained through Profit Maximization	391
9.2.3	Case of Normally Distributed Demand	392
9.2.4	Case of a Fixed Charge to Place the Order	394
9.2.5	Case of Discrete Demand	395
9.3	Single-Period, Constrained, Multi-Item Situation	397
9.3.1	Numerical Illustration	399
9.4	Postponed Product Differentiation	401
9.4.1	Value of Delayed Financial Commitment	402
9.4.2	Value of Flexibility	403
9.5	More than One Period in Which to Prepare for the Selling Season	408
9.6	Multiperiod Newsvendor Problem	408
9.7	Other Issues Relevant to the Control of Style Goods	409
9.7.1	Updating of Forecasts	409
9.7.2	Reorders and Markdowns	410
9.7.3	Reserving Capacity Ahead of Time	411
9.7.4	Inventory Policies for Common Components	411
9.7.5	Other Research	412

9.8	Inventory Control of Perishable Items	413
9.9	Summary	414
	Problems	414
	Appendix 9A: Derivations	422
	References.....	427

SECTION IV MANAGING INVENTORY ACROSS MULTIPLE LOCATIONS AND MULTIPLE FIRMS

10	Coordinated Replenishments at a Single Stocking Point.....	437
10.1	Advantages and Disadvantages of Coordination.....	438
10.2	Deterministic Case: Selection of Replenishment Quantities in a Family of Items	439
10.2.1	Assumptions	439
10.2.2	Decision Rule.....	440
10.2.3	A Bound on the Cost Penalty of the Heuristic Solution	443
10.3	Deterministic Case with Group Discounts	443
10.3.1	Numerical Illustration	446
10.4	Case of Probabilistic Demand and No Quantity Discounts.....	447
10.4.1	(S, c, s) , or Can-Order, Systems.....	448
10.4.2	Periodic Review System.....	448
10.5	Probabilistic Demand and Quantity Discounts.....	451
10.5.1	A Full Truckload Application	453
10.5.2	Numerical Illustration	454
10.6	Production Environment	456
10.6.1	Case of Constant Demand and Capacity: Economic Lot Scheduling Problem	456
10.6.2	Case of Time-Varying Demand and Capacity: Capacitated Lot Sizing.....	461
10.6.3	Probabilistic Demand: The Stochastic Economic Lot Scheduling Problem.....	463
10.7	Shipping Consolidation	464
10.8	Summary	465
	Problems	465
	Appendix 10A: Derivation of Results in Section 10.2	474
	References.....	477
11	Multiechelon Inventory Management	487
11.1	Multiechelon Inventory Management	487
11.2	Structure and Coordination	489
11.3	Deterministic Demand	491
11.3.1	Sequential Stocking Points with Level Demand	491
11.3.2	Other Results for the Case of Level Demand	495
11.3.3	Multiechelon Stocking Points with Time-Varying Demand	496
11.4	Probabilistic Demand	498
11.4.1	Base Stock Control System	501
11.4.2	Serial Situation	503

11.4.3	Arborescent Situation	506
11.5	Remanufacturing and Product Recovery	513
11.5.1	Multiechelon Situation with Probabilistic Usage and One-for-One Ordering	515
11.5.2	Some Extensions of the Multiechelon Repair Situation.....	520
11.5.3	Some Insights and Results for the More General Context of Remanufacturing and Product Recovery.....	521
11.6	Additional Insights	523
11.6.1	Economic Incentives to Centralize Stocks	523
11.6.2	Where to Deploy Stock.....	525
11.6.3	Lateral Transshipments	526
11.7	Summary	526
	Problems	526
	Appendix 11A: Derivation of the Logic for Computing the Best Replenishment Quantities in a Deterministic, Two-Stage Process.....	530
	References.....	531
12	Coordinating Inventory Management in the Supply Chain	543
12.1	Information Distortion in a Supply Chain	544
12.2	Collaboration and Information Sharing.....	546
12.2.1	Sales and Operations Planning.....	546
12.2.2	Collaborative Forecasting	547
12.3	Vendor-Managed Inventory	548
12.4	Aligning Incentives	548
12.4.1	Wholesale Price Contract	549
12.4.2	Buyback Contract	551
12.4.3	Revenue-Sharing Contract.....	553
12.4.4	Service-Level Agreements	554
12.4.5	Challenges Implementing Coordinating Agreements	554
12.5	Summary	555
	Problems	555
	References.....	556

SECTION V PRODUCTION MANAGEMENT

13	An Overall Framework for Production Planning and Scheduling	561
13.1	Characteristics of Different Production Processes	561
13.2	A Framework for Production Decision Making	564
13.2.1	A Review of Anthony's Hierarchy of Managerial Decisions.....	564
13.2.2	Integration at the Operational Level.....	565
13.2.3	The Framework	565
13.3	Options in Dealing with the Hierarchy of Decisions.....	571
13.3.1	Monolithic Modeling Approach	571
13.3.2	Implicit Hierarchical Planning.....	572
13.3.3	Explicit Hierarchical Planning	572
13.3.4	The Hax–Meal Hierarchical Planning System	573
13.4	Summary	576

Problems	577
References.....	577
14 Medium-Range Aggregate Production Planning.....	581
14.1 The Aggregate Planning Problem.....	581
14.2 The Costs Involved	585
14.2.1 Costs of Regular-Time Production	585
14.2.2 Overtime Costs	587
14.2.3 Costs of Changing the Production Rate	587
14.2.4 Inventory Associated Costs	588
14.2.5 Costs of Insufficient Capacity in the Short Run.....	589
14.3 The Planning Horizon	590
14.4 Two Pure Strategies: Level and Chase.....	591
14.5 Feasible Solution Methods	592
14.5.1 General Comments.....	592
14.5.2 An Example of a Graphic–Tabular Method	593
14.6 Linear Programming Models	599
14.6.1 Strengths and Weaknesses.....	601
14.6.2 The Inclusion of Integer Variables in LP Formulations	602
14.6.3 The Land Algorithm	603
14.7 Simulation Search Procedures	603
14.8 Modeling the Behavior of Managers	605
14.8.1 Management Coefficients Models.....	605
14.8.2 Manpower Decision Framework.....	607
14.9 Planning for Adjustments Recognizing Uncertainty	607
14.9.1 The Production-Switching Heuristic	608
14.10 Summary	609
Problems	610
References.....	617
15 Material Requirements Planning and Its Extensions.....	621
15.1 The Complexity of Multistage Assembly Manufacturing.....	622
15.2 The Weaknesses of Traditional Replenishment Systems in a Manufacturing Setting.....	623
15.3 Closed-Loop MRP.....	624
15.4 Material Requirements Planning.....	626
15.4.1 Some Important Terminology.....	626
15.4.2 Information Required for MRP	630
15.4.3 The General Approach of MRP	630
15.4.4 A Numerical Illustration of the MRP Procedure	633
15.4.5 The Material Requirements Plan and Its Uses	639
15.4.6 Low-Value, Common-Usage Items.....	639
15.4.7 Pegging.....	639
15.4.8 Handling Requirements Updates.....	640
15.4.9 Coping with Uncertainty in MRP	641
15.5 Capacity Requirements Planning.....	642
15.6 Distribution Requirements Planning	644

15.7	Weaknesses of MRP	645
15.8	ERP Systems	647
15.8.1	Enhancements to ERP Systems	649
15.9	Summary	650
	Problems	650
	References.....	656
16	Just-in-Time, Optimized Production Technology and Short-Range Production	
	Scheduling.....	661
16.1	Production Planning and Scheduling in Repetitive Situations: Just-in-Time.....	662
16.1.1	Philosophy of JIT.....	662
16.1.2	Kanban Control System.....	664
16.1.3	Benefits and Weaknesses of JIT.....	669
16.2	Planning and Scheduling in Situations with Bottlenecks: Optimized Production Technology.....	671
16.2.1	Philosophy of OPT.....	671
16.2.2	Drum-Buffer-Rope Scheduling.....	676
16.2.3	A Related System: CONWIP.....	677
16.2.4	Benefits and Weaknesses of OPT.....	678
16.3	Short-Range Production Scheduling.....	679
16.3.1	Issues in Short-Term Scheduling.....	680
16.3.2	Techniques for Short-Term Scheduling.....	684
16.3.3	Deterministic Scheduling of a Single Machine: Priority Sequencing Rules.....	688
16.3.4	General Job Shop Scheduling.....	692
16.4	Summary.....	699
	Problems.....	699
	Appendix 16A: Proof that SPT Minimizes Total Flowtime.....	703
	References.....	704
17	Summary.....	713
17.1	Operations Strategy.....	713
17.2	Changing the Givens.....	714
17.3	Future Developments.....	715
	Appendix I: Elements of Lagrangian Optimization	717
	Appendix II: The Normal Probability Distribution	723
	Appendix III: Approximations and Excel Functions	743
	Author Index	749
	Subject Index.....	767



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Preface

Dramatic advances in information and communication technologies have allowed firms to be more closely connected to a broad and global network of suppliers providing components, finished goods, and services. Customers want their products quickly and reliably, even if they are coming from another continent; and they are often willing to look for other suppliers if products are frequently delivered late. As a result, inventory management and production planning and scheduling have become even more vital to competitive success.

Inventory management and production planning and scheduling have been studied in considerable depth from a theoretical perspective. Yet, the application of these theories is still somewhat limited in practice. A major gap has existed between the theoretical solutions, on the one hand, and the real-world problems, on the other.

Our primary objective in the first three editions of the book was to bridge this gap through the development of *operational* inventory management and production planning decision systems that would allow management to capitalize on readily implementable improvements to current practices. Extensive feedback from both academicians and practitioners has been very gratifying in this regard. Our primary objective is unchanged in this revised edition. In the third edition, we placed an emphasis on how computing tools, such as spreadsheets, can be used to apply the models presented in this book. While sophisticated supply chain planning software is now widely available, we still find that spreadsheets are ubiquitous in practice. As such, we have retained and expanded content on the spreadsheet application of the models in this book.

Advances in information technology have made it easier for firms to exchange information with their trading partners. Indeed, much of the academic and practitioner literature on supply chain management has focused on how firms can collaborate and share information to reduce costs in their shared supply chains. Generally speaking, these initiatives have had a positive economic impact. As discussed in Chapter 1, inventory as a percent of sales, and logistics costs as a percent of gross domestic product, have been consistently declining. Our opinion is that these savings are just the tip of the iceberg. It is now commonplace for an employee in an inventory or manufacturing planning role to (a) have powerful tools on their personal computer, and (b) have access to demand and supply-related data. This means it has never been easier to implement rigorous planning models such as those presented in this book. This is not to say that it is easy, just *easier* than before. There are new challenges to planning in modern supply chains where operations are distributed throughout the globe and across many companies. As such, we have included new material in this edition addressing how to collaborate from a planning perspective across departments within an organization and with external trading partners.

As with all editions of this book, we have drawn on the writings of literally hundreds of scholars who have extended theory and who have implemented theory in practice. We have also incorporated many helpful suggestions from our colleagues.

Where appropriate, new modeling approaches and an updated discussion of the literature have been added throughout the book. Here, we briefly highlight some of the major changes in this edition.

1. The first two chapters from the previous edition have been updated and streamlined into one chapter. This chapter highlights recent macroeconomic trends in inventory and logistics costs, and includes a discussion of how these changes affect operations and corporate strategy.
2. Chapters from the previous edition on Just-in-Time and short-range production scheduling have been updated and consolidated into a single chapter.
3. A new chapter on coordinating inventory management in the supply chain has been added. This chapter addresses collaboration and coordination within a firm through processes such as sales and operations planning, as well as external collaboration through initiatives such as collaborative planning, forecasting and replenishment, and vendor managed inventory. Contractual approaches for coordinating inventory decisions between firms are also covered.
4. The treatment of how to manage inventory for items with low-volume, erratic, or non-stationary demand has been expanded. This includes the addition of derivations and spreadsheet applications for the Poisson distribution.
5. The discussion of supply chain planning software tools including Enterprise Resource Planning (ERP) systems and advanced planning tools has been updated and expanded.
6. The treatment of managing multiple items through the use of composite exchange curves has been expanded. To accompany this material, data sets and associated problems are available as supplemental material. These data sets provide students with the opportunity to apply the concepts in the book to realistic, practical settings.

We have continued to attempt to provide a deep and rigorous treatment of the material without presenting complicated mathematics in the main text. Additional derivations have been added in this edition, but this content generally appears in an appendix. We have chosen this style of presentation deliberately so that sufficient material of interest is made available to the analytically inclined reader, while at the same time providing a meaningful text for a less analytically oriented audience.

Section I of the book presents a discussion on inventory management and production planning decisions as important components of total business strategy. The topics covered include operations strategy, the diverse nature of inventories, the complexity of production/inventory decision making, cost measurement, and an introduction to the important concept of exchange curves. A separate chapter is devoted to the determination of forecasting strategy and the selection of forecasting methods.

Section II is concerned with traditional decision systems for the inventory control of individual items. The use of approximate decision rules, *based on sound logic*, permits realistic treatment of time-varying (e.g., seasonal) demand, probabilistic demand, and the many different attitudes of management toward costing the risk of insufficient capacity in the short run.

Section III deals with special classes of items, including the most important (Class A) and the large group of low-activity items (Class C). Also discussed are procedures for dealing with items that can be maintained in inventory for only relatively short periods of time—for example, style goods and perishable items (the newsvendor problem).

Section IV addresses three types of coordination of groups of items primarily in nonproduction contexts. First, there is the situation of a family of items, at a single stocking location, that share a common supplier, a common mode of transport, or common production equipment. The

second type of coordination is where a single firm is managing inventory across several of their locations in a multiechelon network. For such a setting, replenishment requests from one stocking point become part of the demand at another location. The third context is where items are managed across a network where locations in the network are managed by different firms. This setting requires new approaches to share information and align incentives among trading partners to effectively manage inventory throughout the supply chain.

Section V is concerned with decision making in a production environment. A general framework for such decision making is presented. It covers aggregate production planning, material requirements planning, enterprise resource planning systems, JIT, OPT[®], and short-range production scheduling.

This book should be of interest to faculty and students in programs of business administration, industrial/systems engineering, and management sciences/operations research. Although *the presentation is geared to a foundational course in production planning, scheduling, and inventory management*, the inclusion of extensive references permits its use in advanced elective courses and as a starting point for research activities. At the same time, the book should continue to have a broad appeal to practicing analysts and managers.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Acknowledgments

The manuscript, as is often the case, had its origin in the teaching notes used by the authors for several years in courses they taught at the Amos Tuck School at Dartmouth College, The University of Auckland, Boston University, The University of Calgary, The University of Canterbury (New Zealand), The Smeal College of Business at Penn State, The Swiss Federal Polytechnique Institute, and the University of Waterloo. A large number of students (including many part-time students holding employment in industry) have provided excellent critiques.

A significant portion of the book has developed out of research supported by the Natural Sciences and Engineering Research Council of Canada, the Defence Research Board of Canada, and the Ford Foundation. Support was also provided by the Tuck School, the School of Business at the University of San Diego, and the Center for Supply Chain Research in the Smeal College of Business. The authors gratefully acknowledge all the aforementioned support. Numerous consulting assignments involving the authors have also had substantial influence on the contents of the book.

Special mention must be made of two authorities in the field who, early in our careers, encouraged us to work in the general area of inventory and production management—namely, Robert G. Brown, as a colleague at Arthur D. Little, Inc., and Morris A. Cohen, as a PhD advisor at the Wharton School, University of Pennsylvania. Also, we wish to express our sincere thanks to Rein Peterson for his contributions as coauthor of earlier editions of this book.

Many of our professional colleagues have provided helpful comments concerning our research papers, the earlier editions of the book, and the drafts of the current edition. In addition, numerous other colleagues have made available drafts of papers in the more advanced topic areas. Many of the publications of these contributors are referenced in this book. We appreciate all these important contributions. The list of all these colleagues would be too long to include, but we do wish to specifically mention Daniel Costa (Nestle S.A.), Robert Lamarre (Gestion Conseil Robert Lamarre), and David Robb (University of Auckland).

Finally, on such a major task, a special word of thanks is necessary for those individuals who have edited and proofread the many drafts of the manuscript. While a student at Penn State, Jimmy Chen (Bucknell University) carefully reviewed the manuscript, providing very helpful suggestions and edits. Lauren Bechtel, Sharon Cox, Rachel Gimuriman, and Denis Harp were extremely helpful in supporting editing and formatting of tables and figures.

Edward A. Silver
David F. Pyke
Douglas J. Thomas



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Authors



Edward A. Silver is a professor emeritus of operations and supply chain management in the Haskayne School of Business at the University of Calgary. Until his retirement he held the Carma Chair at the University of Calgary. Prior to his appointment at the University of Calgary, he was a professor of management sciences in the Faculty of Engineering at the University of Waterloo. He also previously taught at Boston University and, as a visiting professor, at the Swiss Federal Polytechnique Institute (in Lausanne, Switzerland), the University of Canterbury (in Christchurch, New Zealand), and the University of Auckland (New Zealand).

A native of Montreal, Professor Silver completed a bachelor of civil engineering (applied mechanics) at McGill University and a science doctorate in operations research at the Massachusetts Institute of Technology. He is a licensed professional engineer and has been a member of a number of professional societies, including the American Production and Inventory Control Society, the Canadian Operational Research Society (of which he was the president in 1980–1981), the Institute of Industrial Engineers, the Institute for Operations Research and the Management Sciences, the International Society for Inventory Research (of which he was the president in 1994–1996), the Production and Operations Management Society, and the Operational Research Society (UK).

Professor Silver has presented seminars and talks at national and international meetings of a number of professional societies as well as educational institutions throughout North America and in parts of Europe, Asia, and New Zealand. He has published close to 170 articles in a broad range of professional journals. Dr. Silver has also served in an editorial capacity for several journals.

Professor Silver spent four years as a member of the Operations Research Group of the international consulting firm, Arthur D. Little Inc. Subsequently he has done independent consulting for a wide range of industrial and government organizations throughout North America and elsewhere. These consulting activities have addressed both tactical and strategic problems arising in the management of operations. Specific areas of application have included inventory management, supply chain management, process improvement, production planning, and logistics management. An additional important activity has been his involvement in several executive development programs and other workshops related to the inventory/production field.

Dr. Silver has received extensive professional recognition. This has included being named as a Fellow of five organizations, specifically the Institute of Industrial Engineers (1995), the Manufacturing and Services Operations Management Society (2000), the International Society for Inventory Research (2000), the Institute for Operations Research and the Management Sciences

(2003), and the Production and Operations Management Society (2010). He also was the recipient of two of the highest awards of the Canadian Operational Research Society, namely the Award of Merit (1990) and the Harold Larnder Memorial Prize (2007). Finally, he was awarded a Visiting Erskine Fellowship at the University of Canterbury, Christchurch, New Zealand (1998) and a Visiting Research Fellowship by the Japan Society for the Promotion of Science (2002).



David F. Pyke, PhD, is professor of operations and supply chain management in the School of Business at the University of San Diego (USD). He was dean of the business school from 2008 until 2015, leading the school to a significant presence in national and international business school rankings. Formerly, he was the Benjamin Ames Kimball Professor of Operations Management, and associate dean at the Amos Tuck School of Business Administration at Dartmouth College. He obtained his BA from Haverford College, MBA from Drexel University, and his MA and PhD from the Wharton School of the University of Pennsylvania. He was awarded an honorary MA from Dartmouth College in 1999.

He has taught executive programs at USD, Tuck, Wharton, and in other environments. He has taught globally at the International University of Japan, the Helsinki School of Economics, and the WHU-Otto-Beisheim-Hochschule, Vallendar, Germany. Professor Pyke has consulted for The Rand Corporation, Accenture, Corning, DHL, Nixon, Eaton, Home Depot, Lemmon Company and Markem, among others. He has also served as an expert witness on supply chain management issues for securities cases. He serves on the Board of Directors of GW Plastics, Concepts NREC, and until recently the Lwala Community Alliance, a nonprofit focused on community development in Kenya. He is an advisor to GuardHat, Inc., Analytics Ventures, and Moore Venture Partners, an operating partner of Tuckerman Capital LLC, and partner with San Diego Social Venture Partners.

Professor Pyke's research interests include operations management, supply chain management, integrated enterprise risk management, pricing, inventory systems, manufacturing in China, production management, and manufacturing strategy. He has published numerous papers in journals such as *Management Science*, *Operations Research*, *Production and Operations Management*, *Sloan Management Review*, *Journal of Operations Management*, *Naval Research Logistics*, *European Journal of Operational Research*, and *Interfaces*. He coedited a book, *Supply Chain Management: Innovations for Education*, with M. Eric Johnson, in the *POMS Series in Technology and Operations Management*. He has served on the editorial boards for *Management Science* and *Naval Research Logistics*, among others. He is a member of the Institute for Operations Research and Management Sciences, and the Production and Operations Management Society.



Douglas J. Thomas, PhD, is a professor of supply chain management in the Smeal College of Business at Pennsylvania State where he was the faculty director of the MBA program from 2011 to 2014. He also serves as chief scientist for Plan2Execute, LLC, a firm that provides supply chain software and consulting solutions in warehouse management, transportation management, and advanced planning. Doug earned his MS and PhD from Georgia Institute of Technology in industrial engineering and has a BS in operations research from Cornell University. Prior to returning to graduate school, he worked for C-Way Systems, a software company specializing in manufacturing scheduling. In addition to his years on the faculty at Penn State, Doug has had the pleasure of serving as a visiting faculty member

at INSEAD (in Fontainebleau, France), the Johnson Graduate School of Management at Cornell University and The Darden School at the University of Virginia.

Doug currently teaches courses in the areas of supply chain management and quantitative modeling in MBA, executive MBA, and PhD programs. A frequent faculty leader in executive development programs, Doug has led numerous executive education sessions in Africa, Asia, Europe, and North America, including programs at Penn State, INSEAD, and Georgia Institute of Technology as well as custom programs for Accenture, DuPont, ExxonMobil, IBM, Ingersoll-Rand, Mars, Office Depot, Parker-Hannifin, Pfizer, Schlumberger, and the U.S. Marine Corps. He has testified as an expert witness and consulted for several large organizations on supply chain strategy, including Accenture, CSL Behring, Dell, ExxonMobil, and Lockheed Martin Aerospace Corporation.

His research interests include coordinating production and inventory planning across the extended enterprise and connecting decision models to logistics performance measurement. His work has appeared in several academic and practitioner journals in the areas of logistics and operations management, including *Management Science*, *Manufacturing and Service Operations Management*, and *Production and Operations Management*. He serves as a senior editor for *Production and Operations Management*. He is a member of the Institute for Operations Research and Management Sciences, and the Production and Operations Management Society.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

THE CONTEXT AND IMPORTANCE OF INVENTORY MANAGEMENT AND PRODUCTION PLANNING

I

Since the mid-1980s, the strategic benefits of inventory management and production planning and scheduling have become obvious. The business press has highlighted the success of Japanese, European, and North American firms that have achieved unparalleled efficiency in manufacturing and distribution. In recent years, many of these firms have “raised the bar” yet again by coordinating with other firms in their supply chains. For instance, instead of responding to unknown and highly variable demand, they share information so that the variability of the demand they observe is significantly lower.

In spite of the media attention devoted to these developments, two mistakes are far too common. First, many managers assume that new levels of efficiency can be attained simply by sharing information and forming “strategic alliances” with their supply chain partners. These managers do not understand that “the devil is in the details,” and that knowing what to do with the data is as important as getting the data in the first place. Developing sound inventory management and production planning and scheduling methods may seem mundane next to strategy formulation, but these methods are a critical element of long-term survival and competitive advantage. Second, many analysts assume that implementing sophisticated inventory and production methods will solve all the problems. These analysts may achieve a high level of efficiency by optimizing given the lead times and demand variability the firm observes. But they do not understand that this efficiency is insignificant compared to the efficiencies available by *changing the givens*.

2 ■ *Inventory and Production Management in Supply Chains*

In Chapter 1, we discuss the importance of inventories and production planning at an aggregate level. This includes a discussion of the role of aggregate inventories in the business cycle. We also present an overview of corporate strategy, and discuss the linkages of finance and marketing strategies with inventory management and production planning and scheduling. We then present a framework for formulating an operations strategy, and we describe how inventory management and production planning and scheduling fit in that context.

Chapter 2 presents several frameworks that will be helpful to operating managers, and to general managers. We describe various statistical properties of inventories, and we discuss how inventories can be classified. A useful framework for understanding different production processes is presented. Finally, we discuss the costs involved, and briefly note where they can be found.

Since managerial expectations about the future have such a tremendous impact on inventory management and production planning and scheduling decisions, forecasting the demand variable is given special treatment in a separate chapter. In Chapter 3, we present a number of strategies and techniques that could be adopted to cope with the unknown future.

Chapter 1

The Importance of Inventory Management and Production Planning and Scheduling

Some of the strongest and fastest-growing industries throughout the world are in the service sector. The consulting and financial services industries, in particular, hire thousands of college graduates each year. This fact has prompted many to suggest that manufacturing in developed countries has not only declined in importance, but is on a path toward extinction. It seems clear, however, that nations care deeply about manufacturing. This is particularly true in certain industries. Our observations suggest that, although they receive little attention in the press, the textile/apparel, machine tool,* and food industries are considered vital. Why is this so? We think it is due to at least two reasons. The first is that food and machine tools are fundamental to national security. Without an industry that manufactures and distributes food, a nation is very vulnerable if it is isolated due to a conflict; and in such situations, without machine tools, the hardware that drives the economy will grind to a halt. The second reason is employment. The textile/apparel industry employs millions of people worldwide, many in low-skill, entry-level, jobs. Without this industry a nation can face high levels of unemployment and has reduced capacity to bring people, such as immigrants, into the workforce.

Many other industries including microprocessors, computers, and automobiles are also considered vital in today's world. These industries are often the source of innovation, productivity improvements, and high-skill jobs. Consider, for example, the productivity improvements that resulted from the introduction of computer numerically controlled machine tools. So, while services have been growing in importance, we submit that manufacturing is still fundamental to the health of most modern economies. Adoption of the methods discussed in this book will ultimately help to strengthen both developed and developing economies, to the extent that they are successfully implemented in manufacturing and logistics firms. Many of the methods in this book also

* Machine tools are the machines that make machines, and therefore are the fundamental building blocks for any manufacturing operation.

4 ■ *Inventory and Production Management in Supply Chains*

apply to service industries such as banking, hospitals, hotels, restaurants, schools, and so on, where firms hold inventories of supplies. These stocks must be managed carefully, and in some cases, careful management of these stocks can be critical to the performance of the firm. Two recent studies highlight the importance of inventory management to firm performance. Hendricks and Singhal (2009) show that excess inventory announcements are followed by a negative stock market reaction, and Chen et al. (2007) relate abnormal inventory levels to poor stock returns.

Costs associated with production, inventory, and logistics are quite economically significant. Table 1.1 illustrates logistics costs in the United States. Note that the value of inventories in U.S. firms exceeded \$2 trillion in 2007, and increased to close to \$2.5 trillion by 2014. Total logistics costs, expressed as a percentage of gross domestic product (GDP), have slowly decreased in recent years, from 9.9% in 2000 to 8.2% in 2014. Several key factors continue to make effective inventory management challenging, including increasing product variety, shortening product lifecycles, and an increase in global sourcing. Despite these pressures, inventory levels for manufacturers have been dropping. Figure 1.1 shows inventory-to-sales ratios over time for U.S. retailers, manufacturers, and wholesalers. The figure suggests that inventory levels were decreasing up until 2009, perhaps starting to increase since then. In a rigorous examination of U.S. firms between 1981 and 2000, Chen et al. (2005) report that inventory levels dropped at an average annual rate of 2%. The majority of this improvement came from a reduction in work-in-process (WIP) inventory, with finished goods inventories not declining significantly. In a related study, Rajagopalan and Malhotra (2001) also find that raw material and WIP inventories declined for U.S. firms between 1961 and 1994.

While inventory levels as a percentage of sales have been declining, inventories have shifted downstream, closer to the customer. Figure 1.2 shows the fraction of inventory held by manufacturers compared to the fraction held by wholesalers and retailers. In 1992, manufacturers held 46% of total inventory compared to 37% in 2014. There are several factors that may be causing this shift. In recent years, firms have sourced their products from all over the globe in an effort to find cost efficiencies. Global sourcing initiatives will invariably increase lead times leading to higher inventory levels for wholesalers and retailers. A recent empirical study by Jain et al. (2013) reported that firms with more global sourcing have higher inventory levels. In addition to global sourcing initiatives, product variety offered to customers continues to increase. Firms with many products and short product lifecycles often have high gross margin and high demand uncertainty (Fisher 1997). An empirical study by Rumyantsev and Netessine (2007) confirms that higher gross margin and more demand uncertainty translate to higher inventory levels. An empirical study of U.S. retailers by Kesavan et al. (2016) observes that retailers with high inventory turnover are able to more effectively adjust purchase quantities to react to demand fluctuations, and the negative consequences of excess and shortage of inventory are far less severe for retailers with high inventory turnover.

For analysts, managers, consultants, and entrepreneurs, the opportunities to add value to manufacturing or logistics firms are enormous. On the basis of research and our own experience, it is evident that most firms do not fully understand the complexities of managing production and inventory throughout their supply chain. While there are certainly opportunities for the creative manager in the realms of finance and marketing, our focus is on the benefits that can be won by careful and competent management of the flow of goods throughout the supply chain. One of the authors regularly requires students in an inventory management class to work on consulting projects with local firms. Over the years, we have seen that in more than 90% of the cases, improved inventory or production management would lead to cost savings of at least 20%, without sacrificing customer service. This figure has been replicated in our consulting experience as well. One of our

Table 1.1 U.S. Logistics Costs (\$ Billion)

<i>Year</i>	<i>GDP</i>	<i>Value of Business Inventory (a)</i>	<i>Inventory Carrying Rate (%) (b)</i>	<i>Inventory Carrying Costs (c) = (a × b)</i>	<i>Transportation Costs (d)</i>	<i>Administrative Costs (e)</i>	<i>Total U.S. Logistics Costs (f) = (c + d + e)</i>	<i>Total Logistics Costs as a % of GDP</i>
2000	10,280	1,525	25.3	386	594	39	1,019	9.9
2001	10,620	1,448	22.8	330	609	37	976	9.2
2002	10,980	1,495	20.7	309	582	35	926	8.4
2003	11,510	1,557	20.1	313	607	36	956	8.3
2004	12,270	1,698	20.4	346	652	39	1,037	8.5
2005	13,090	1,842	22.3	411	739	46	1,196	9.1
2006	13,860	1,994	24.0	479	809	50	1,338	9.7
2007	14,480	2,119	24.1	511	855	54	1,420	9.8
2008	14,720	2,050	21.4	439	872	52	1,363	9.3
2009	14,420	1,927	19.3	372	705	43	1,120	7.8
2010	14,960	2,130	19.2	409	770	47	1,226	8.2
2011	15,520	2,301	19.1	439	819	50	1,308	8.4
2012	16,160	2,392	19.1	457	847	52	1,356	8.4
2013	16,770	2,444	19.1	467	885	54	1,406	8.4
2014	17,420	2,496	19.1	477	901	55	1,433	8.2

Source: Council of Supply Chain Management Professionals Annual State of Logistics Report® 2015.

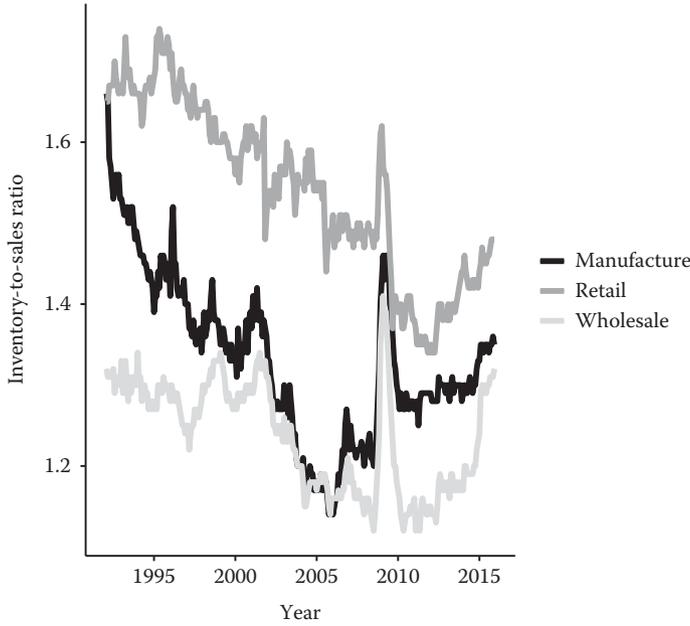


Figure 1.1 Inventory-to-sales ratios (seasonally adjusted) for U.S. retailers, manufacturers, and wholesalers. (From U.S. Census Bureau. November 2015 Manufacturing and Trade Inventories and Sales report.)

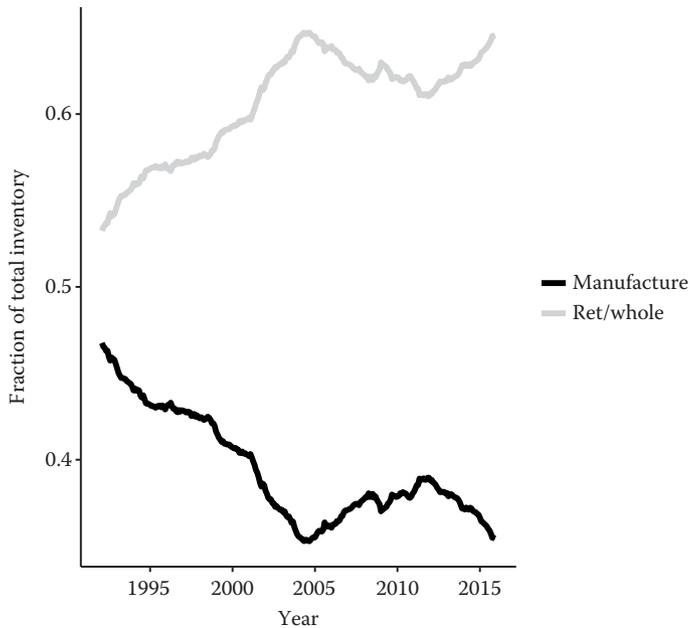


Figure 1.2 Distribution of inventory between manufacturing and combined wholesale and retail (seasonally adjusted). (From U.S. Census Bureau. November 2015 Manufacturing and Trade Inventories and Sales report.)

goals in this book is to provide the reader with knowledge and skill so that he or she can bring real value to employers and, as a result, to the economy as a whole.

In Section 1.1, we discuss how business cycles affect aggregate inventory investments. We turn our attention to the connection between corporate strategy and inventory management and production planning in Section 1.2. In Section 1.3, we extend the strategy discussion to the areas of marketing and finance. Section 1.4 contains a framework for operations strategy. Section 1.5 deals with an important ingredient of decision making—namely, the specification of appropriate measures of effectiveness.

1.1 Why Aggregate Inventory Investment Fluctuates: The Business Cycle

It is an unfortunate fact that economies go through cycles—periods of expansion when employment rates are high and the general mood is one of unending prosperity, followed by periods of contraction when unemployment grows and there is a general feeling of malaise about the economy (Mack 1967; Reagan and Sheehan 1985; Blinder and Maccini 1991). As we shall see, inventories play an important role in these cycles. Many economists have tried to compile comprehensive models of the business cycle to explain all patterns of fluctuations that have occurred historically (e.g., Schmitt-Grohé and Uribe 2012). To date, no one has succeeded in building an all-purpose model. Nevertheless, it is apparent that although each cycle is somewhat different, especially with regard to its exact timing and relative magnitude, the cycles continue and there are several common underlying factors. These are illustrated in Figure 1.3.

Prior to the peak labeled in Figure 1.3, the economy is expanding and managers are optimistic about future sales. However, managers can be overzealous in their optimism. In fact, at the peak, because of overoptimistic expectations, too many products are manufactured by the economy and cannot be sold. These surplus goods increase aggregate inventories, and so producers start to decrease production levels. Eventually, the rate of sale exceeds the rate of production of goods.

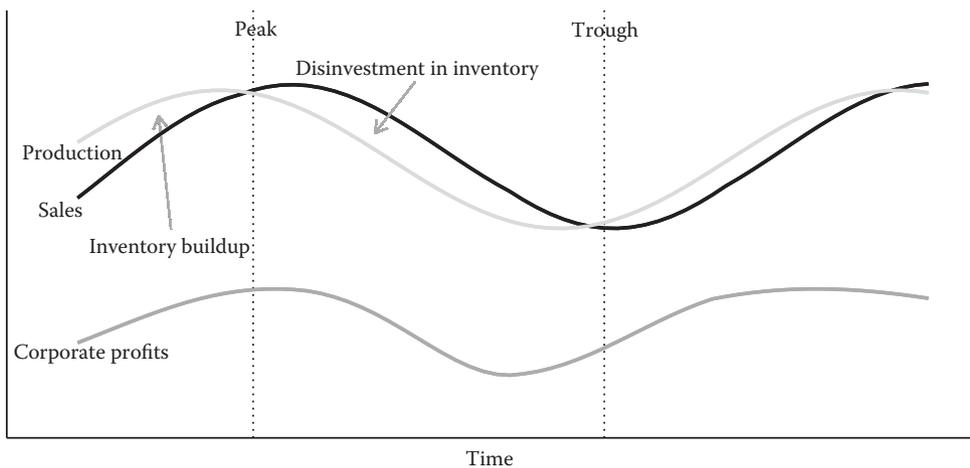


Figure 1.3 The business cycle.

The resulting disinvestment in inventories creates a recession during which prices, production, and profits fall and unemployment is prevalent. The financial crisis of 2007–2008 led to a rapid drop in sales and a subsequent buildup of inventory. This can be seen in the inventory-to-sales ratio of that time period in Figure 1.1.

After a time, an economic recovery is generated by a slowing in the rate of inventory liquidation. Some top managers, expecting that prices will recover, start to slowly expand their operations while costs are low. More and more firms slowly start to hire additional labor, purchase more raw materials, and thereby infuse more money into circulation in the economy. Consumers, with money to spend once again, start to bid up prices of available goods. Once prices of goods start to rise, more and more executives get on the bandwagon by expanding their operations and thereby accentuate the expansionary phase of the cycle. The boom that results eventually is brought to an end when costs of materials are bid up once again by competing firms, when the labor force begins to demand higher wages, and when the scarcity of money for further expansion causes the banks to raise their interest rates. (At this writing, interest rates remain quite low.) The crisis phase that follows is a period of uncertainty and hesitation on the part of consumers and business. Executives find that their warehouses are restocked with excess inventory that, once again, cannot be sold. The business cycle then is ready to repeat itself as explained above.

In Figure 1.3, the following key relationships are illustrated by the data. The peaks and troughs in corporate profits before taxes precede the peaks and troughs in production. Inventory investment lags slightly and thereby contributes to higher cyclical amplitudes in production than are really necessary.

This explanation is, of course, highly simplified, but it does illustrate the main forces, especially the role of inventories, at work during each cycle. Note that the expectations of business and consumers, as well as the ability of decision makers to react quickly and correctly to change, are important determinants of the length and severity of a cycle. Expectations about the future have been shown to depend on the following variables: the trend of recent sales and new orders, the volume of unfilled orders, price pressures, the level of inventories in the recent past, the ratio of sales to inventories (the turnover ratio to be discussed in Section 1.3.1), interest rates on business loans, the current level of employment, and the types of decision-making systems used by management. It is precisely this last point that we address in this book. We want to provide managers with sophisticated, yet understandable approaches for managing production and inventory in the supply chain, enabling them to make good decisions in a timely manner.

Recent research suggests that business cycles are less volatile than they were 50 years ago due, in part, to advances in inventory management and production planning and scheduling (Bloom et al. 2014). Many of the topics we will discuss in this book serve to more closely match production with demand. Therefore, managers can have a more timely perspective on the market, and are less surprised by economic downturns. Other models, such as the probabilistic inventory models discussed in Chapters 6 through 9, have been shown to decrease inventory volatility when applied correctly. Still, there is much work to be done, and there are many opportunities for knowledgeable and creative operations managers to help their firms weather the volatility of economic cycles.

1.2 Corporate Strategy and the Role of Top Management

Earlier in this chapter, we saw that, although management of production and inventory throughout the supply chain can be critical to the ability of a firm to achieve and maintain competitive advantage, top managers often do not recognize the importance of these issues. Operations

managers, for their part, often neglect the role their activities play in the strategic direction of the firm, and even in its operations strategy. Both senior and operations managers need to understand the nature of corporate and operations strategy, and they need to understand how inventory management and production planning and supply chain issues impact other functions. In this chapter, we briefly discuss corporate strategy and the functional strategies that derive from it, including marketing, finance, and operations.

It is important to note that some firms have reorganized, or reengineered, to eliminate functional areas and replace them with *business processes*. For instance, a major multinational food and beverage firm recently reengineered in a way that redefined roles to be more responsive to the customer. A common reengineering approach is to replace the operations, logistics, and marketing functions with teams that are process focused. For instance, one team may be devoted to *generating demand*, while another focuses on *fulfilling demand*. The members of the generate demand team perform many of the tasks that traditionally have been done by the marketing department but they may carry out other functions as well, including new product development. The fulfill-demand team often looks like the manufacturing or operations department but may include other functions such as supply chain, logistics, sales, and marketing. The idea is to align the organizational structure with the processes the firm uses to satisfy its customers. Barriers between functions that created delays and tension are removed, enabling the firm to meet customer orders in a more seamless way. For our purposes, the management of inventory, production planning, and supply chains can be thought of as part of the operations function or the fulfill-demand process. In this book, we shall refer to marketing and operations functions rather than to generate and fulfill-demand processes. Nevertheless, the procedures and insights we present apply in either case.

It is not our intention to cover the topic of strategic planning in any depth. There are a number of texts on this topic (e.g., Porter 1980; Mintzberg and Quinn 1991; Grant and Jordan 2015). Instead, our objective is to show that decision making in the production and inventory areas must not be done in a vacuum, as is too often the case in practice, but rather must be coordinated with decisions in other functional areas by means of corporate strategic planning.

The key organizational role of top managers is strategic business planning. Senior management has the responsibility for defining in broad outline what needs to be done, and how and when it should be done. Top management also must act as the final arbiter of conflicts among operating divisions and has the ultimate responsibility for seeing that the general competitive environment is monitored and adapted to effectively. In most business organizations, four levels of strategy can be delineated (listed from the highest to the lowest level):

1. *Enterprise Strategy*: What role does the organization play in the economy and in society? What should be its legal form and how should it maintain its moral legitimacy?
2. *Corporate Strategy*: What set of businesses or markets should the corporation serve? How should resources be deployed among the businesses?
3. *Business Strategy*: How should the organization compete in each particular industry or product/market segment? On the basis of price, service, or what other factor?
4. *Functional Area Strategy*: At this level, the principal focus of strategy is on the maximization of resource productivity and the development of distinctive competencies.

While each level of strategy can be seen as being distinct, they must fit together to form a coherent and consistent whole. Typically, each level is constrained by the next higher one.

We will have more to say on strategic planning in Chapter 13 where a specific framework for decision making in a production environment will be developed.

1.3 The Relationship of Finance and Marketing to Inventory Management and Production Planning and Scheduling

1.3.1 Finance

Inventories have an important impact on the usual aggregate scorecards of management performance—namely, on the balance sheet and the income statement.* First, inventories are classified as one of the current assets of an organization. Thus, *all other things being equal*, a reduction in inventories lowers assets relative to liabilities. However, the funds freed by a reduction in inventories normally would be used to acquire other types of assets or to reduce liabilities. Such actions directly influence the so-called *current ratio*, the ratio of current assets to current liabilities, which is the most commonly used measure of liquidity.

Income statements represent the flow of revenues and expenses for a given period (e.g., 1 year). Specifically,

$$\text{Operating profit} = \text{Revenue} - \text{Operating expenses} \quad (1.1)$$

Changes in inventory levels can affect both of the terms on the right side of Equation 1.1. Sales revenue can increase if inventories are allocated among different items in an improved way. Also, because inventory carrying charges represent a significant component of operating expenses, this term can be reduced if aggregate inventory levels decrease. The labor component of operating expenses can also be reduced by more effective production scheduling and inventory control.

A primary aggregate performance measure for inventory management is inventory turnover, or *stockturns*. In this book, we define inventory turnover as

$$\text{Inventory turnover} = \frac{\text{Annual sales or usage (at cost)}}{\text{Average inventory (in \$)}} \quad (1.2)$$

Turnover can be a very useful measure, especially when comparing divisions of a firm or firms in an industry. See Gaur et al. (2005) for a detailed analysis of inventory turnover in the retail sector. From Equation 1.2, we can see that an increase in sales without a corresponding increase in inventory will increase the inventory turnover, as will a decrease in inventory without a decline in sales.

There are several dangers with turnover as a performance measure, however. Too frequently we have seen leaders of manufacturing firms respond to turnover less than the industry average by issuing an edict to increase turnover to the industry average. In one company, while competitive firms purchased some component parts, this company manufactured them because of a special competence in this type of manufacturing. WIP inventory was necessary to keep the production process flowing smoothly. To increase inventory turnover, all they would have to do is to buy more and make less. When the edict came down, the controller recalculated the decision rules for inventory ordering so that lot sizes were cut and raw material inventory would be reduced. Shipments to customers continued for a few months as warehouse inventories were depleted, but then customer service degenerated when finished goods and component inventories bottomed out. Rampant stockouts created massive, expensive expediting. The smaller lot sizes drove up the product cost, and profits plummeted. In short, strategic competitive advantage can have a significant effect on the determination of the appropriate inventory turnover number.

* Droms (1979) provides a nontechnical description of financial reports.

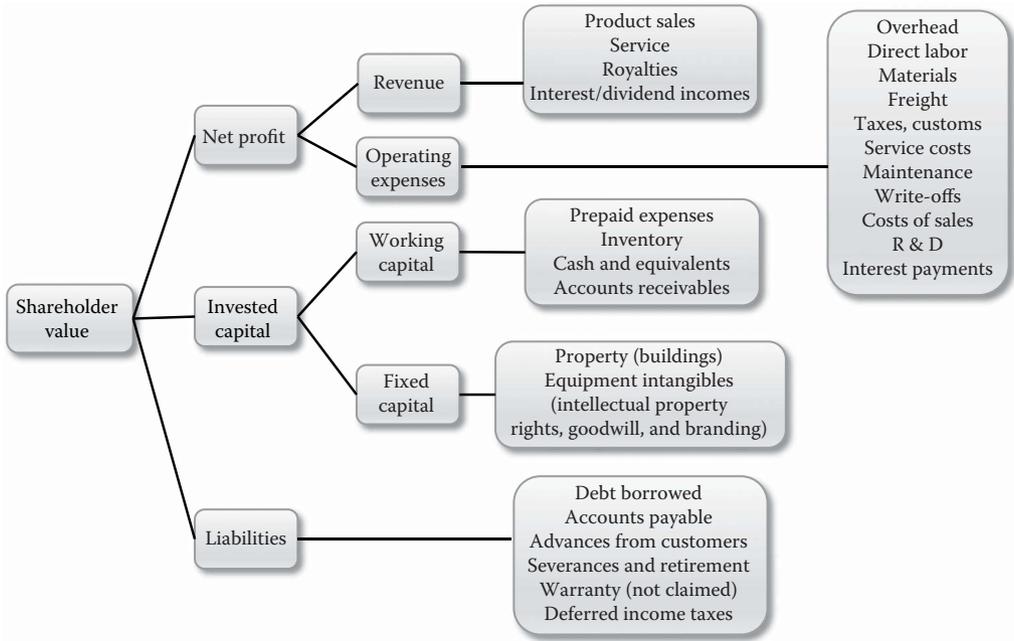


Figure 1.4 Components of shareholder value.

One of the most common measures of managerial performance is the *return on investment* (ROI), which represents the profit (after taxes) divided by the average investment (or level of assets). From the above discussion, it is clear that inventories have an important impact on the ROI (also see Problem 1.18).

Finally we note that many of these relationships can be summarized in Figure 1.4, which is due to Lee, H. and S. Whang (1997, Personal communication). Inventory is a component of working capital, and the interest payments to finance that inventory are a component of operating expenses. Improvements in inventory management, therefore, can have a significant impact on shareholder value. Likewise, improvements in production planning and scheduling can increase customer service and, therefore, product sales, and can decrease expenses related to direct labor, materials, and freight. Thus, net profit can increase, again increasing shareholder value. It is useful to consider other potential impacts of supply chain management, inventory management, and production planning and scheduling on the components of Figure 1.4 (see Problem 1.14).

1.3.2 Marketing

The most common complaint we hear from operations managers about marketing is that marketing managers simply do not understand how difficult it is to manufacture and distribute a wide variety of products. Because of low-volume production of many products, productive capacity is lost to setups, and product demand is more difficult to forecast. Workers and equipment must be more flexible, and inventories must be higher. Marketing managers prefer a wide variety of products because they are listening to customers and trying to respond to their needs and desires. High inventories are desirable because demand can be met without delay. Furthermore, sales and

marketing managers may make pricing decisions, such as steep promotional discounts, that increase unit sales, but send a shock through the factory and supply chain. These fundamental conflicts can be mitigated by a process orientation as discussed in the introduction to this chapter. Even with a process orientation, however, conflicts are bound to arise. This again highlights the need for senior managers to be involved in setting policy for supply chain management, inventory management, and production planning and scheduling.

1.4 Operations Strategy

In this section, we develop a framework for operations strategy and we emphasize the place of the management of inventory, production planning, and supply chain in the broader context of the operations function.

We begin by stressing again that the operations strategy of the firm should be in concert with other functional area strategies, including marketing, finance, and human resources. Sometimes, however, one function should take precedence over the others. One firm went through a painful period of not responding to customer needs in a long-term effort to improve customer satisfaction. The operations group needed to develop the capability to manufacture high-quality items in high volume. To avoid disruptions during the improvement process, customer desires were neglected in the short term, and marketing was frustrated. In the long term, however, customers were delighted by the level of quality and by the responsiveness of the firm.

The operations strategy itself is made up of three levels: mission, objectives, and management levels. (See Figure 1.5.)

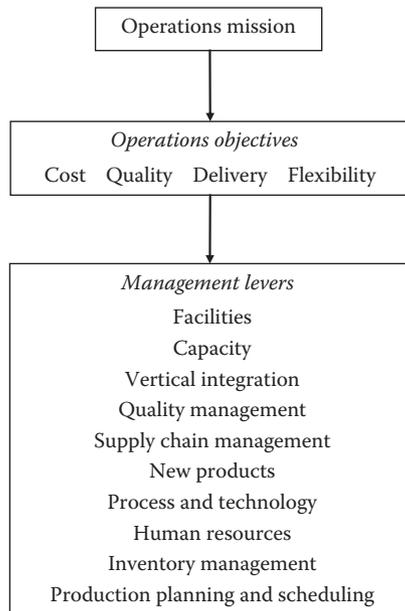


Figure 1.5 A framework for operations strategy.

1.4.1 Mission

The operations mission defines a direction for the operations function. McDonald's, for instance, uses four terms to describe its operational mission: quality, cleanliness, service, and value. For many years, the annual report has prominently highlighted these terms, most recently in the 2014 shareholder letter from the CEO. Because the mission should not change significantly over time, the mission statement is often somewhat vague. Otherwise, it is likely that it would have to be reworded frequently as changes occur in the environment, competition, or product and process technology. Employees need to know that there is a consistent direction in which their company is moving.

It is important however that the mission be attractive to excellent people. If it is a sleepy statement of direction that resembles the mission of most other firms in the industry, it will be difficult to attract the best employees, not to mention investors. Therefore, a mission statement should incorporate some of the excitement of top management and should communicate to employees, investors, and customers that this is an excellent firm.

1.4.2 Objectives

Because the mission statement is vague, it is difficult to know whether or not it has been achieved. How can one know, for instance, whether "quality" has been achieved? The second level of an operations strategy, operations objectives, provides carefully defined, measurable goals that help the firm achieve its mission. For many years, firms have used four operational objectives: cost, quality, delivery, and flexibility.

It is critical that the objectives are *defined* carefully, clearly *measurable*, and *ranked*. They must be defined carefully because these terms are often used loosely. Quality at a McDonald's restaurant is very different from quality at a five-star restaurant, which is very different from quality in a hospital operating room. It is also important that the objectives be measurable so that management knows whether they are meeting their goals or not. It is quite possible, and perhaps desirable, to use more than one measure for each objective (for instance, warranty cost and parts-per-million defective for the quality objective). The objectives should be rank ordered as well, so that managers can give priority to the correct objective when it is necessary to make tradeoffs. A manager of a high volume manufacturing line made it quite clear to one of the authors that, in his opinion, cost was more important than delivery. When questioned further, however, he noted that in fact he had occasionally gone over budget by using overtime in order to meet a due date. In other words, he said he ranked cost above delivery, but his behavior indicated that delivery was more important. Subsequent discussions with senior managers helped him understand that delivery was, in fact, primary.

In the 1970s, many people in operations thought that cost and quality were opposites, as were delivery and flexibility. For instance, one could aim for low-cost production, but then it would be impossible to achieve high quality. Likewise, if one had rapid delivery, it would be impossible to be flexible. More recent experience, however, suggests that cost and quality are complements rather than opposites. Warranty, prevention, and detection costs decrease as quality improves. Rework and congestion on the factory floor also decrease significantly, thereby reducing the cost of products. In addition, recent experience would indicate that rapid delivery of customized products can be possible. This is especially true with technology such as flexible automation, electronic data interchange, supplier web portals, radio frequency identification (RFID), and scanning technology. Most firms have instances, however, in which tradeoffs among the objectives must be made. Therefore, although combined improvements are possible, the objectives should still be ranked.

The cost objective can be considered in one of the three categories: low, competitive, or relatively unimportant. In a *low-cost* environment, such as that of certain discount retailers, the goal is to have the lowest cost products or services in the industry. Firms that aim for *competitive* costs do not necessarily strive to have the lowest cost products, but rather want to be competitive with most other firms in the industry. Some firms produce prototypes or have a unique product for which they can charge a *premium*; hence, cost becomes less important. Cost measures include dollars per unit, inventory turns, and labor hours per unit (or other productivity measures). In the United States, low cost tended to be the primary objective of manufacturing firms from the 1950s to the mid-1970s.

The quality objective rose to the fore in the mid-1970s to the mid-1980s with the advent of Japanese quality improvements and the sale of Japanese products in the United States. In particular, the automotive industry experienced the stunning effects of high-quality Japanese products. Quality can be defined by understanding which of its multiple dimensions are important. Garvin (1987) describes eight dimensions of quality: performance, conformance, reliability, durability, serviceability, features, aesthetics, and perceived quality. Quality measures include percent defective, percent returns, results from satisfaction surveys, warranty dollars, and so on.

Delivery can be defined on two dimensions, speed and reliability. For instance, some firms compete on delivering within 24 hours of the customer request. Others may take longer but will assure the customer that the products will be provided reliably within the quoted delivery time. Some companies rank delivery last among the objectives. Prototype manufacturing, for instance, sometimes involves complete customization and, therefore, may require long delivery times. Measures for delivery include percent on time, time from request to receipt, fill rate (the percent of demand met from the shelf), and so on.

The fourth objective is flexibility. Flexibility has three dimensions—volume, new product, and product mix. *Volume* flexibility is the ability to adjust for seasonal variations and fluctuations. *New product* flexibility is the speed with which new products are brought from concept to market. Automotive firms in recent years have made great strides in new product flexibility. A niche car allows a firm to jump in quickly to low-volume, profitable, markets. This flexibility is impossible if development time is eight to 10 years, as was traditionally the case with U.S. and European manufacturers. Most manufacturers have now reduced development time to roughly three-and-a-half years, a target first reached by several Japanese automotive firms. *Product mix* flexibility is the ability of a company to offer a wide range of products. This may simply mean that the catalog contains many items, or it may mean that the firm has the ability to develop customized products. Many machine tool companies produce a single product for a given customer, and then never produce that exact product again. Hence, they would rank product mix flexibility very high.

It has been argued that delivery and flexibility are the most important objectives in some industries because cost-cutting programs and quality improvement programs have “leveled the playing field” on cost and quality. These firms then compete on “time”—rapid introduction of new products and rapid delivery of existing products. The phrase *time-based competition* has been used to describe this phenomenon.

Finally, we note that the objectives are dynamic. For instance, as a new product begins full-scale production, the firm may emphasize flexibility (to design changes) and on-time delivery so that market share is not lost to competitors. Over time, as the product design stabilizes, the emphasis may change toward quality and cost.

1.4.3 Management Levers

Although the operations objectives provide measurable goals, they do not indicate how a firm should pursue those goals. The 10 management levers—facilities, capacity, vertical integration, quality management, supply chain management, new products, process and technology, human resources, inventory management, and production planning and scheduling—provide the tactical steps necessary to achieve the goals.*

It is interesting to note that in the late 1970s, researchers did not include quality management, supply chain relationships, new products, and human resources in the list of management levers (Wheelwright 1978). Today, these areas are considered critical to competitive success. It is also certain that new levers will be introduced in the next decade. The categories for the operations objectives, on the other hand, have not changed.

Facilities decisions concern the location and focus of factories and other facilities. Where does a company locate its factories and distribution centers to meet the needs of its customers? Is there a need for multiple facilities or is a single location sufficient? Does each facility perform all functions or is one facility focused on a particular market, a particular process, or a particular product? (See Sheu and Krajewski 1996). Many manufacturing companies have one plant that acts as a “parent plant” that is responsible for odd-ball parts and new product introduction. When products reach high-volume manufacture, they are moved to a satellite plant where efforts are focused on excellence in that particular product and/or process. It is critical to evaluate each plant manager based on the specific charter for his or her facility.

Capacity decisions involve magnitude and timing of capacity expansion. Capacity decisions interact with the facility location decisions. For instance, some firms set limits on the number of employees that may work at any one location; by remaining smaller than a given size, better communication and teamwork are possible. As demand grows, additional expansion at the same site would violate the limit, and a new plant site must be found. A Massachusetts textile manufacturer had rapidly expanding sales in Europe. When capacity in the Massachusetts factory was no longer able to meet total demand, they had to determine whether to expand in Massachusetts, elsewhere in the United States, or in Europe. The final decision was to build in eastern Germany to exploit lower transportation costs and import duties for European sales, and to take advantage of tax breaks offered by the German government.

Vertical integration concerns make/buy decisions. Some firms make components, perform final assembly, and distribute their products. Others focus only on product design and final assembly, relying on other firms to manufacture components and to distribute finished goods. Facilities, capacity, and vertical integration decisions are primarily made for the long term because they involve bricks and mortar and significant investment.

Quality management encompasses the tools, programs, and techniques used to achieve the quality goals. It is distinct from the quality objective in that the objective specifies the definition and measurable targets, and the lever specifies the means to achieve the targets. Quality management includes things such as statistical process control (SPC), Taguchi methods, quality circles, and so on. Many aspects of process improvement fall under this management lever, although process improvement leads to increased productivity as well as improved quality. It is important to note that different definitions of quality may dictate the use of different quality procedures.

* Portions of this discussion are drawn from Fine and Hax (1985).

The *supply chain management* lever focuses on relationships with suppliers and customers. Indeed, it is concerned with the entire network of firms and activities from, say, digging ore out of the ground, to getting the product into the hands of the customer. Furthermore, it is concerned with the handling of the product after it has finished its useful life. In general, supply chain management uses tools to efficiently move information, products, and funds across the entire supply chain. Chapter 12 contains an extensive discussion of supply chain management.

The *new products* category involves the procedures and organizational structures that facilitate new product introduction. The days of a focused team of engineers working without input from marketing or manufacturing are gone. Most firms now employ multifunctional teams, composed of design engineers, marketing personnel, manufacturing managers, and production line workers. The new product lever specifies the reporting relationships as well as any procedures for setting milestones in the development process.

The *process and technology* category encompasses the choice of a production process and the level of automation. The product-process matrix, to be discussed in Chapter 2, is a useful framework for analyzing process choices. Decisions about implementing new process technology are included here as well.

Human resources involves the selection, promotion, placement, reward, and motivation of the people that make the business work.

Inventory management encompasses decisions regarding purchasing, distribution, and logistics, and specifically addresses when and how much to order. This management lever is the topic of Chapters 4 through 11, although Chapter 6 alone provides a comprehensive discussion of the issues involved.

Finally, *production planning and scheduling* focuses on systems for controlling and planning production. Chapters 13 through 16 deal with specific production planning and scheduling decisions and systems.

1.4.4 General Comments

We conclude this section with several general comments. First, researchers and practitioners are continually introducing new approaches addressing some important aspect of management. Inventory management and production planning and scheduling methods that would have been considered impossible just a few years ago are commonplace today. The revolution in information technology and computer networking allows members of a supply chain to have access to manufacturers and to track consumer demand on a daily basis, and thereby more closely match production with demand. Many of these new approaches fall under the board area now widely known as *supply chain management*. The supply chain lever, of course, pertains to supply chain management, but so do inventory management (as we shall see in Chapters 11 and 12), production planning and scheduling, vertical integration, and new products. Occasionally it is necessary to introduce a new lever to focus attention on the important issues, but other times, the existing levers adequately serve the purpose. Even with new supply chain initiatives, however, inventory management and production planning and scheduling continue to be a headache for many managers. Evidence suggests that even though a massive amount of real-time data is available, many firms do not use it, either because they are overwhelmed with data, or because they do not know what to do with it. Although real-time information can significantly reduce the demand variability a firm observes, significant variability and the associated problems usually remain. Managers must understand the value of the information, and they must choose inventory policies and production planning and scheduling systems that are appropriate for their situations. One of our primary goals in this book

is to communicate the basic principles and procedures in this field so that managers, consultants, and analysts can provide insight and value to their firms in this dynamic world.

Second, as a firm audits its manufacturing strategy, it is important to recognize that the policies in place for each of the 10 levers should be consistent not only with the operations objectives, but also among themselves. If there are inconsistencies, the firm should initiate action programs to mitigate any negative effects. For instance, many companies historically pursued quality improvement programs without changing worker incentives. Workers were rewarded for the volume of their output, without regard for quality. In other words, the quality objectives and policies were inconsistent with the human resources policies. Reward systems had to be rearranged to fit with the quality goals. Likewise, as we discuss inventory management and production planning and scheduling in great depth throughout the remainder of this book, keep in mind the relationship of these management levers to the other eight, and to the operations objectives. Extensive effort to reduce inventory cost may not be warranted, for instance, if the firm can charge a premium price and really needs to focus on flexibility. In a similar vein, customer service levels, which will be discussed in Chapter 6, should be established with clear knowledge of the cost and delivery objectives. These issues will be developed further in Chapter 13.

Finally, in the process of auditing a manufacturing strategy, managers should understand distinctive competencies at the detailed level of the management levers. These distinctive competencies should influence the objectives, mission, and business strategy. Thus, information flows both ways, from business strategy to management levers and from levers to business strategy.

1.5 Measures of Effectiveness for Inventory Management and Production Planning and Scheduling Decisions

Analysts in the production/inventory area have tended to concentrate on a single quantifiable measure of effectiveness, such as cost, sometimes recognizing certain constraints, such as limited space, desired customer service, and so forth. In actual fact, of course, the impact of decisions is not restricted to a single measure of effectiveness. The appropriate measures of effectiveness should relate back to the underlying objectives of management. Unfortunately, some of the following objectives are very difficult to quantify:

1. Minimizing political conflicts (in terms of the competing interests) within the organization.
2. Maintaining a high level of flexibility to cope with an uncertain future.
3. Maximizing the chance of survival of the firm or the individual manager's position within the organization.
4. Keeping at an acceptable level the amount of human effort expended in the planning and operation of a decision system.

Moreover, as will be discussed in Chapter 2, even some of the more tangible cost factors are very difficult to estimate in practice.

From the perspective of top management, the measures of effectiveness must be aggregate in nature. Management is concerned with *overall* inventory levels and customer service levels (by broad classes of items) as opposed to, for example, minimization of costs on an individual item basis. Furthermore, rather than optimization in the absolute sense, a more appealing justification for a new or modified decision system is that it simply provides significant improvements (in terms of one or more measures of effectiveness) over an existing system.

One other complexity is that certain decisions in inventory management and production planning and scheduling should not necessarily be made in isolation from decisions in other areas. As an illustration, both Schwarz (1981) and Wagner (1974) point out that there are important interactions between the design of a physical distribution system (e.g., the number, nature, and location of warehouses) and the selection of an appropriate inventory management system. (Chapter 11 will deal further with this particular issue.) Other researchers address another type of interaction, namely, between transportation (shipping) decisions and those of inventory management. (See Ernst and Pyke 1993; Henig et al. 1997.) Another interaction, vividly portrayed by the success of Japanese management systems, is between inventory management and quality assurance. Finally, interactions between pricing and inventory decisions have generated much interest in research and in practice. (See Chan et al. 2004; Fleischmann et al. 2004; Chen and Chen 2015). In this book, we do not develop mathematical models that directly deal with these large-scale interactive problems. However, in several places, further descriptive advice and references will be presented.

1.6 Summary

In this chapter, we have examined a number of strategic issues related to inventory management and production planning and scheduling. In much of the remainder of the book, our attention will be focused on specific issues of how best to design and operate an inventory management and production planning and scheduling decision system for a specific organizational context. Nevertheless, it is important not to lose sight of the broader perspective of this chapter.

Problems

- 1.1 Select two appropriate, recent articles from a newspaper and identify how their contents could impinge on decision making in inventory management and production planning.
- 1.2 Why do inflation and supply shortages tend to increase real inventories? What is the effect of a reduction in the rate of inflation on inventory investment?
- 1.3 Investigate and discuss the inventory control procedures at a small (owner-managed) company. When does a company become too small to have its inventories managed scientifically?
- 1.4 Repeat Problem 1.3 for production planning and scheduling procedures.
- 1.5 Repeat Problem 1.3 for a small restaurant.
- 1.6 Look for evidence in the newspaper or business press to support or contradict the research on business cycles cited in Section 1.1. Discuss your findings.
- 1.7 Find an article in the business press that indicates the position of top management toward manufacturing issues, such as inventory management or production planning and scheduling. Discuss your findings.
- 1.8 Consider any other course you have taken. Briefly discuss how some key concepts from it are likely to interact with decisions in the inventory management or production planning and scheduling areas.
- 1.9 For each of three or four specific local industries, what do you think is the approximate breakdown between raw materials, WIP, and finished goods inventories? Discuss your answer.

- 1.10 In Section 1.5, mention was made of an interaction between the design of a physical distribution system and the choice of an inventory management system. Briefly comment on this interaction.
- 1.11 Examine the interaction specified in Section 1.5 again, but now between the choice of a transportation mode and the selection of inventory decision rules. Briefly discuss your results.
- 1.12 Briefly comment on how inventory management might interact with each of the following functional areas of the firm:
 - a. Maintenance
 - b. Quality control
 - c. Distribution (shipping)
- 1.13 A management decision system designed for providing information on how well a business is doing, relative to other businesses in the same industry segment, is appropriate for
 - a. Top management
 - b. Middle management
 - c. Marketing management
 - d. Factory management
 - e. Office staff

Explain the reasoning behind your choice.
- 1.14
 - a. What components of Figure 1.4 would change, and in what direction, if production planning and scheduling systems were improved?
 - b. What components of Figure 1.4 would change, and in what direction, if inventory management systems were improved?
 - c. What components of Figure 1.4 would change, and in what direction, if supply chain management systems were improved?
- 1.15 Discuss the potential interactions between the production planning and scheduling lever and the inventory management lever.
- 1.16 What new topics do you think will be included on the list of management levers in 10 years? What levers will be eliminated?
- 1.17 If it is true that U.S. firms focused on cost, then quality, and then on delivery and flexibility (i.e., time), what will they focus on next?
- 1.18 An important financial performance measure is the return on investment:

$$\text{ROI} = \frac{\text{Profit}}{\text{Total assets}}$$

Assume that inventories are 34% of current assets.

- a. Consider the typical company and assume that it has a ratio

$$\frac{\text{Current asset}}{\text{Total assets}} = 0.40$$

What impact will a 25% decrease in inventory have on the ROI? What percent decrease in inventories is needed to increase the ROI by 10%?

- b. Develop an algebraic expression for the impact on ROI of an $x\%$ decrease in inventories when

$$\frac{\text{Current asset}}{\text{Total assets}} = f$$

- 1.19 Analyze the financial statements for the last 5 years of two companies from different industries. (E.g., a grocery chain and an electronic goods manufacturer.) Calculate their turnover ratios. Try to explain why their turnover ratios are different. Are there any trends apparent that can be explained?
- 1.20 Why is the turnover ratio widely used? What are its benefits and what are its drawbacks?
- 1.21 Blood is collected for medical purposes at various sites (such as mobile units), tested, separated into components, and shipped to a hospital blood bank. Each bank holds the components in inventory and issues them as needed to satisfy transfusion requests. There are eight major types of blood and each type has many components (such as red cells, white cells, and plasma). Each component has a different medical purpose and a different lifetime. For example, the lifetime for white cells is only 6 hours, but the lifetime for red cells is now 42 days. What type of management issues are involved in this particular type of inventory management situation? (Discuss from strategic, tactical, and operational perspectives.) Does perishability require different records than for other situations? How?

References

- Blinder, A. S. and L. J. Maccini. 1991. Taking stock: A critical assessment of recent research on inventories. *Journal of Economic Perspectives* 5, 73–96.
- Bloom, N., M. Floetotto, N. Jaimovich, I. Saporta Eksten, and S. Terry. 2014. Really uncertain business cycles. *US Census Bureau Center for Economic Studies Paper No. CES-WP-14-18*.
- Chan, L. M., Z. M. Shen, D. Simchi-Levi, and J. L. Swann. 2004. Coordination of pricing and inventory decisions: A survey and classification. In D. Simchi-Levi, S. David Wu, and Z. M. Shen (Eds.), *Handbook of Quantitative Supply Chain Analysis*, pp. 335–392. US: Springer.
- Chen, H., M. Z. Frank, and O. Q. Wu. 2005. What actually happened to the inventories of American companies between 1981 and 2000? *Management Science* 51(7), 1015–1031.
- Chen, H., M. Z. Frank, and O. Q. Wu. 2007. US retail and wholesale inventory performance from 1981 to 2004. *Manufacturing and Service Operations Management* 9(4), 430–456.
- Chen, M. and Z.-L. Chen. 2015. Recent developments in dynamic pricing research: Multiple products, competition, and limited demand information. *Production and Operations Management* 24(5), 704–731.
- Droms, W. G. 1979. *Finance and Accounting for Nonfinancial Managers*. Reading, MA: Addison-Wesley.
- Ernst, R. and D. Pyke. 1993. Optimal base stock policies and truck capacity in a two-echelon system. *Naval Research Logistics* 40, 879–903.
- Fine, C. and A. Hax. 1985. Manufacturing strategy: A methodology and an illustration. *Interfaces* 15(6), 28–46.
- Fisher, M. L. 1997. What is the right supply chain for your product? *Harvard Business Review* 75(2)(March/April), 105–116.
- Fleischmann, M., J. M. Hall, and D. F. Pyke. 2004. Smart pricing. *MIT Sloan Management Review* 45(2), 9–13.
- Garvin, D. A. 1987. Competing on the eight dimensions of quality. *Harvard Business Review* (November–December), 101–109.
- Gaur, V., M. L. Fisher, and A. Raman. 2005. An econometric analysis of inventory turnover performance in retail services. *Management Science* 51(2), 181–194.
- Grant, R. M. and J. J. Jordan. 2015. *Foundations of Strategy*. UK: John Wiley & Sons.
- Hendricks, K. B. and V. R. Singhal. 2009. Demand–supply mismatches and stock market reaction: Evidence from excess inventory announcements. *Manufacturing and Service Operations Management* 11(3), 509–524.
- Henig, M., Y. Gerchak, R. Ernst, and D. Pyke. 1997. An inventory model embedded in a supply contract. *Management Science* 43(2), 184–189.

- Jain, N., K. Girotra, and S. Netessine. 2013. Managing global sourcing: Inventory performance. *Management Science* 60(5), 1202–1222.
- Kesavan, S., T. Kushwaha, and V. Gaur. 2016. Do high and low inventory turnover retailers respond differently to demand shocks? *Manufacturing and Service Operations Management* 18(2), 198–215.
- Mack, R. P. 1967. *Information Expectations and Inventory Fluctuations*. New York: Columbia University Press.
- Mintzberg, H. and J. B. Quinn. 1991. *The Strategy Process: Concepts, Contexts, Cases*. Englewood Cliffs, NJ: Prentice-Hall.
- Porter, M. E. 1980. *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. New York: Free Press.
- Rajagopalan, S. and A. Malhotra. 2001. Have US manufacturing inventories really decreased? An empirical study. *Manufacturing and Service Operations Management* 3(1), 14–24.
- Reagan, P. and D. P. Sheehan. 1985. The stylized facts about the behavior of manufacturers' inventories and backorders over the business cycle: 1959–1980. *Journal of Monetary Economics* 15(2), 217–246.
- Rumyantsev, S. and S. Netessine. 2007. What can be learned from classical inventory models? A cross-industry exploratory investigation. *Manufacturing and Service Operations Management* 9(4), 409–429.
- Schmitt-Grohé, S. and M. Uribe. 2012. What's news in business cycles. *Econometrica* 80(6), 2733–2764.
- Schwarz, L. 1981. Physical distribution: The analysis of inventory and location. *AIIE Transactions* 13(2), 138–150.
- Sheu, C. and L. J. Krajewski. 1996. A heuristic for formulating within-plant manufacturing focus. *International Journal of Production Research* 34(11), 3165–3185.
- Wagner, H. M. 1974. The design of production and inventory systems for multifacility and multiwarehouse companies. *Operations Research* 22(2), 278–291.
- Wheelwright, S. 1978. Reflecting corporate strategy in manufacturing decisions. *Business Horizons* (February), 56–62.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Chapter 2

Frameworks for Inventory Management and Production Planning and Scheduling

In this chapter, we first discuss, in Sections 2.1 and 2.2, the complexity of inventory management and production planning and scheduling decisions (even without considering interactions with other functional areas, as outlined in Section 1.3 of Chapter 1). Then, in Sections 2.3 through 2.5, we present various aids for coping with the complexity. Section 2.6 discusses costs and other relevant variables. Sections 2.7 and 2.8 provide some suggestions concerning the use of mathematical models. Sections 2.9 and 2.10 are concerned with two very different approaches for estimating relevant cost factors identified earlier in the chapter. Finally, Section 2.11 provides a detailed look at the process of a major study of inventory management or production planning and scheduling systems.

2.1 The Diversity of Stock-Keeping Units

Some large manufacturing companies and military organizations stock over 500,000 distinct items in inventory. Large distributors and retailers, such as department stores, carry about 100,000 goods for sale. A typical medium-sized manufacturing concern keeps in inventory approximately 10,000 types of raw materials, parts, and finished goods.

Items produced and held in inventory can differ in many ways. They may differ in cost, weight, volume, color, or physical shape. Units may be stored in crates, in barrels, on pallets, in cardboard boxes, or loose on shelves. They may be packaged by the thousands or singly. They may be perishable because of deterioration over time, perishable through theft and pilferage, or subject to obsolescence because of style or technology. Some items are stored in dust-proof, temperature-controlled rooms, while others can lie in mud, exposed to the elements.

Demand for items also can occur in many ways. Items may be withdrawn from inventory by the thousands, by the dozen, or unit by unit. They may be substitutes for each other, so that, if one item is out of stock, the user may be willing to accept another. Items can also be complements; that

is, customers will not accept one item unless another is also available. Units could be picked up by a customer, or they may have to be delivered by company-owned vehicles or shipped by rail, boat, airplane, or truck. Some customers are willing to wait for certain types of products; others expect immediate service on demand. Many customers will order more than one type of product on each purchase order submitted.

Goods also arrive for inventory by a variety of modes and in quantities that can differ from how they will eventually be demanded. Some goods arrive damaged; others differ in number or kind from that which was requisitioned (from a supplier or from an in-house production operation). Some items are unavailable because of strikes or other difficulties in-house or at a supplier's plant. Delivery of an order may take hours, weeks, or even months, and the delivery time may or may not be known in advance.

Decision making in production, inventory, and supply chain management is therefore basically *a problem of coping with large numbers and with a diversity of factors external and internal to the organization*. Given that a specific item is to be stocked at a particular location, three basic issues must be resolved:

1. How often the inventory status should be determined
2. When a replenishment order should be placed
3. How large the replenishment order should be

Moreover, the detailed daily individual item decisions must, in the aggregate, be consistent with the overall objectives of management.

2.2 The Bounded Rationality of a Human Being

An inescapable fact alluded to in the first chapter was the need to view the inventory management and production planning and scheduling problems from a systems standpoint. Before we plunge into a detailed discussion of the possible solutions to the problem as we have posed it, let us consider explicitly and realistically our expectations regarding the nature of the results we seek.

Simon (1957, pp. 196–206) has pointed out that all decision makers approach complex problems with a framework or model that simplifies the real situation. A human being's brain is simply incapable of absorbing and rationalizing all of the many relevant factors in a complex decision situation. A decision maker cannot effectively conceive of the totality of a large system without assistance. All decision makers are forced to ignore some relevant aspects of a complex problem and base their decisions on a smaller number of carefully selected factors. These selected factors always reflect decision makers' personal biases, abilities, and perceptions of the realities they think they face, as well as the decision technology that is available to them at any point in time.

The decisions of inventory management and production planning and scheduling are complex. They extend beyond the intuitive powers of most decision makers because of the many interconnected systems, both physical and conceptual, that have to be coordinated, rationalized, adapted to, or controlled. The decisions must be viewed *simultaneously* from the point of view of the individual item in its relation to other similar items, the total aggregate inventory investment, the master plan of the organization, the production–distribution systems of suppliers and customers, and the economy as a whole.

The challenge is therefore twofold. Since a decision maker's ability to cope with diversity is limited, decision systems and rules must be designed to help expand the bounds put upon that

individual's ability to rationalize. However, because most decision makers probably have already developed personalized approaches to the inventory/production decision, decision systems have to be designed, and their use advocated, in the context of existing resources and managerial capabilities.

Inventory management and production planning and scheduling as practiced today, like the practice of accounting, because of the predominance of numerical data and a relatively long history, are mixtures of economically sound theory, accepted industrial practices, tested personalized approaches, and outright fallacies. A goal of this book is to influence the amount of sound theory that actually gets used, while trying to clarify existing fallacies. However, one must accept the fact that existing theory is, and will be, for some time to come, insufficient to do the whole job. There will always be room for personalized, tested-in-practice, approaches to fill the gaps in theory. Therefore, within this book, we in effect weave our own brand of personalized approaches with theory. This is why, from an intellectual point of view, we find inventory and production decisions both challenging and exciting.

2.3 Decision Aids for Managing Diverse Individual Items

The managerial aids are basically of two kinds, conceptual and physical.

2.3.1 Conceptual Aids

The following is a partial list of conceptual aids for decision making:

1. Decisions in an organization can be considered as a hierarchy—extending from long-range strategic planning, through medium-range tactical planning, to short-range operational control. Typically, different levels of management deal with the three classes of decisions. We have more to say about such a hierarchy, specifically related to production situations, in Chapter 13.
2. A related type of hierarchy can be conceptualized with respect to decision making in inventory management and production planning and scheduling. At the highest level, one chooses a particular type (or types) of control system(s). At the next level, one selects the values of specific parameters within a chosen system (e.g., the desired level of customer service). Finally, one operationalizes the system, including data collection, calculations, reporting of results, and so forth (Brown 1982).
3. The large number of physical units of inventory can be classified into a smaller number of relatively homogeneous organizational categories. Managers then manage inventories in the context of this smaller number of decision classifications (to be discussed in Section 2.4).
4. The complexity of production planning and inventory management decisions can be reduced by identifying, through analysis and empirical research, only the most important variables for explicit consideration (e.g., ordering costs and demand rates). This is an especially important essence of the operational theory we expound in this book.

2.3.2 Physical Aids

Primary physical decision aids include the following:

1. First, and foremost, decision makers can employ spreadsheets and other computer programs. Throughout this book, we comment on how spreadsheets can be used to compute many of the formulas presented; and we ask the reader to develop his or her own spreadsheets in many of the end-of-chapter problems. Occasionally, we note that the algorithms are too difficult for simple spreadsheet implementation and suggest the use of other computer programs. In either case, the computer can handle vast amounts of data faster than the manager could ever dream of doing manually. The American Production and Inventory Control Society (APICS) publishes a magazine (*APICS—The Performance Advantage*) that regularly reviews relevant software.
2. A decision maker's physical span of control can also be expanded by harnessing the *principle of management by exception*. Decision rules are used within the computer system that focus the attention of the decision maker only on unusual or important events.

2.4 Frameworks for Inventory Management

In this section, we present several frameworks designed to help reduce the complexity of managing thousands of items.

2.4.1 Functional Classifications of Inventories

We recommend six broad decision categories for controlling aggregate inventories: cycle stock, congestion stock, safety (or buffer) stock, anticipation inventories, pipeline inventories, and decoupling stock. In our opinion, senior management can and must express an opinion on how much aggregate inventory is required in each of these broad categories. While we cannot expect that such opinions will always be in the form of clear-cut, simple answers, the categorization helps managers view inventories as being controllable rather than as an ill-defined evil to be avoided at all costs.

Cycle inventories result from an attempt to order or produce in batches instead of one unit at a time. The amount of inventory on hand, at any point, that results from these batches is called cycle stock. The reasons for batch replenishments include (1) economies of scale (because of large setup costs), (2) quantity discounts in purchase price or freight cost, and (3) technological restrictions such as the fixed size of a processing tank in a chemical process. The amount of cycle stock on hand at any time depends directly on how frequently orders are placed. As we will see, this can be determined in part by senior management who can specify the desired trade-off between the cost of ordering and the cost of having cycle stock on hand. A major consideration in Chapters 4, 5, 7, and 10 will be the determination of appropriate cycle stocks.

Congestion stocks are inventories due to items competing for limited capacity. When multiple items share the same production equipment, particularly when there are significant setup times, inventories of these items build up as they wait for the equipment to become available. Chapter 16 will discuss congestion stocks in some depth.

Safety stock is the amount of inventory kept on hand, on the average, to allow for the uncertainty of demand and the uncertainty of supply in the short run. Safety stocks are not needed when the future rate of demand and the length of time it takes to get complete delivery of an order are known with certainty. The level of safety stock is controllable in the sense that this investment is directly related to the desired level of customer service (i.e., how often customer demand is met from stock). The determination of safety stock levels in response to prespecified managerial measures of customer service will be discussed in Chapters 6 through 11.

Anticipation inventory consists of stock accumulated in advance of an expected peak in sales. When demand is regularly lower than average during some parts of the year, excess inventory (above cycle and safety stock) can be built up so that, during the period of high anticipated requirements, extra demand can be serviced from stock rather than from, for example, working overtime in the plant. That is, the aggregate production rate can be stabilized through the use of anticipatory stock. Anticipation stock can also occur because of seasonality of supply. For example, tomatoes, which ripen during a rather short period, are processed into ketchup that is sold throughout the year. Likewise, climatic conditions, such as in the Arctic, can cause shipments to be restricted to certain times of the year. Anticipation stocks can also consist of inventories built up to deal with labor strikes, war crises, or any other events that are expected to result in a period of time during which the rate of supply is likely to be lower than the rate of demand. Chapter 14 will deal with the rational determination of anticipation inventories as a part of production planning.

Pipeline (or *WIP*) inventories include goods in transit (e.g., in physical pipelines, on trucks, or in railway cars) between levels of a multiechelon distribution system or between adjacent work stations in a factory. The pipeline inventory of an item between two adjacent locations is proportional to the usage rate of the item and to the transit time between the locations. Land's End, Dell Computer, and other direct-mail firms have negligible pipeline inventory in the distribution system because they deliver directly to the consumer. Pipeline inventories will be considered in Chapters 11, 15, and 16.

Decoupling stock is used in a multiechelon situation to permit the separation of decision making at the different echelons. For example, decoupling inventory allows decentralized decision making at branch warehouses without every decision at a branch having an immediate impact on, say, the central warehouse or factory. In this light, Zipkin (1995) notes that inventory is a "boundary" phenomenon. Large amounts of inventory tend to exist between firms, or between divisions within a firm. As these inventory levels are reduced in an attempt to reduce costs, inventory managers must take into account the larger system of suppliers and customers throughout the supply chain. The topics of multiechelon inventory management and supply chain management (to be discussed in Chapters 11 and 12) attempt to capture some of these system-wide effects and manage them effectively. As we dig deeply into the fundamental principles and formulas, it is essential to keep in mind the upstream and downstream effects of our decisions. Chapter 16 will include discussion related to decoupling stocks.

Note that our six functional categories were defined in order to concentrate attention on the organizational purposes of the inventories, especially with regard to control and manageability, rather than on accounting measures (e.g., raw materials, supplies, finished components, etc.). Most managers are intuitively aware of the functions that inventories in each of the six categories must play. Some of the benefits (and costs) may even be measured by existing cost-accounting systems. However, most accounting systems do not keep track of opportunity costs, such as customer disservice through lost sales or the cost of extra paperwork generated by an order that must be expedited to meet an unexpected emergency.

While it will not always be possible to precisely measure the costs and benefits of each subgroup of inventories, over the years, many experienced managers have garnered a sense of perspective regarding the effect of reducing or increasing the amount of inventory investment allocated to each of these six categories. One of our goals in this book is the development of procedures to harness this experience, along with measurable costs and benefits and analytical techniques, to yield better decisions from an overall systems viewpoint.

It is quite possible that an individual manager may not have had sufficient experience to have developed an intuitive feel for the six functional categories that we have defined above. This could

especially be true in large organizations. In such cases, instead of classifying a company's total aggregate inventory investment, it may be more meaningful for a manager to deal with a smaller inventory investment, such as that of one division or strategic business unit.

Meaningful functional groupings do vary from organization to organization. But, from a theoretical and practical viewpoint, at some point, the six functional categories—cycle stock, congestion, safety, anticipation, pipeline, and decoupling stocks—need to be analyzed from a cost–benefit standpoint to provide an aggregate perspective for the control of individual items.

2.4.2 The A–B–C Classification as a Basis for Designing Individual Item Decision Models

Managerial decisions regarding inventories must ultimately be made at the level of an individual item or product. The specific unit of stock to be controlled will be called a *stock-keeping unit* (or *SKU*), where an *SKU* will be defined as an item of stock that is completely specified as to function, style, size, color, and, often, location. For example, the same-style shoes in two different sizes would constitute two different SKUs. Each combination of size and grade of steel rod in raw stock constitutes a separate SKU. An oil company must regard each segregation of crude as a separate SKU. A tire manufacturer would normally treat the exact same tire at two geographically remote locations as two distinct SKUs. Note that this classification system can result in the demand for two SKUs being highly correlated in practice, because a certain portion of customers will always be willing to substitute, say, a blue jacket for the one that is red.* As we will see, particularly in Chapters 10, 11, and 15, decision rules often must coordinate a number of distinct SKUs.

Close examination of a large number of actual multi-SKU inventory systems has revealed a useful statistical regularity in the usage rates of different items. (We deliberately use the broader-term usage, rather than demand, to encompass supplies, components, spare parts, etc.) Typically, somewhere on the order of 20% of the SKUs account for 80% of the total annual dollar usage. This suggests that all SKUs in a firm's inventory should not be controlled to the same extent. Figure 2.1 illustrates the typical Distribution by Value (DBV) observed in practice.†

A DBV curve can be developed as follows: The value v , in dollars per unit, and the annual usage (or demand) D , in specific units, of each SKU in inventory are identified. Then, the product Dv is calculated for each SKU, and the Dv values for all SKUs are ranked in descending order starting with the largest value, as in Table 2.1. Then, the corresponding values of the cumulative percent of total dollar usage and the cumulative percent of the total number of SKUs in inventory are plotted on a graph such as Figure 2.1. Experience has revealed that inventories of consumer goods will typically show a lesser concentration in the higher-value SKUs than will an inventory of industrial SKUs. Furthermore, empirically, it has been found that quite often the distribution of usage values across a population of items can be adequately represented by a lognormal distribution. Therefore, we can relatively simply estimate the aggregate effects of different inventory control policies. In the Appendix of this chapter, we show (1) the mathematical form of the distribution, (2) its moments (expected values of various powers of the variable), (3) a graphic procedure for testing the adequacy

* In practice, of course, one can define an SKU in other (e.g., less detailed) ways depending on the decision being made and the level of detail at which one wants to control inventories.

† Incidentally, the same DBV concept is applicable to a wide variety of other phenomena such as the distribution of (1) incomes within a population of individuals, (2) donation sizes among a group of donors, and so forth. It is also very similar to the Pareto chart used in quality management. Similar methods are used in market analysis.

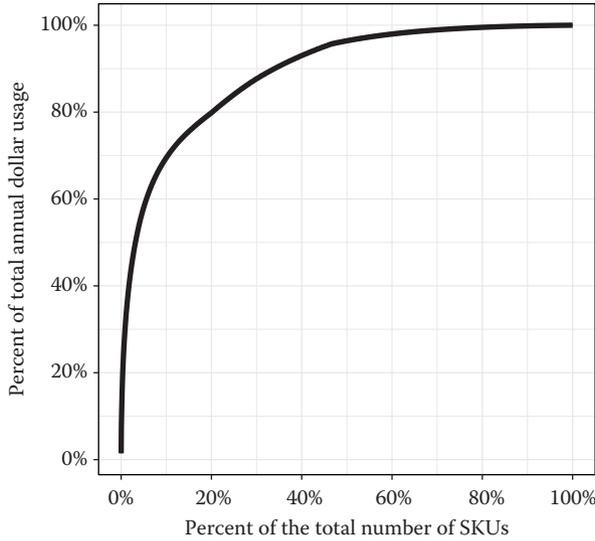


Figure 2.1 DBV of SKUs.

of the fit to a given DBV, and (4) a graphic method of estimating the two parameters of the distribution (m , the average value of Dv , and b which is a measure of the dispersion or spread of the Dv values).*

A table, such as Table 2.1, is one of the most valuable tools for handling the diversity of disaggregate inventories because it helps to identify the SKUs that are the most important. These SKUs will be assigned a higher priority in the allocation of management time and financial resources in any decision system we design. It is common to use three priority ratings: A (most important), B (intermediate in importance), and C (least important). The number of categories appropriate for a particular company depends on its circumstances and the degree to which it wishes to differentiate the amount of effort allocated to various groupings of SKUs. For example, one can always subdivide the DBV into further categories, such as “moderately important,” etc., as long as the resulting categories receive differentiated treatment. A minimum of three categories is almost always used, and we use this number to present the basic concepts involved.

Class A items should receive the most personalized attention from management. The first 5%–10% of the SKUs, as ranked by the DBV analysis, are often designated for this most important class of items although up to 20% of the SKUs can be designated as class A in some settings. Usually these items also account for somewhere in the neighborhood of 50% or more of the total annual dollar movement ($\sum Dv$) of the population of items under consideration. In Chapters 5, 7, 9, 10, and 11, we discuss decision systems that are suitable for use with class A items.

* The parameter b is not the standard deviation of Dv . It is the standard deviation of the underlying normally distributed variable, namely $\ln(Dv)$.

Table 2.1 Sample Listing of SKUs by Descending Dollar Usage

Sequential Number	SKU I.D.	Cumulative Percent of SKU	Annual Usage Value (<i>Dv</i>)	Cumulative Usage	Cumulative Percent of Total Usage
1	–	0.5	\$3,000	\$3,000	13.3
2	–	1.0	2,600	5,600	24.9
3	–	1.5	2,300	7,900	35.1
–	–	–	–	–	–
–	–	–	–	–	–
–	–	–	–	22,498	–
199	–	99.5	2	22,500	100
200	–	100.0	0	22,500	100

Class B items are of secondary importance in relation to class A. These items, because of their *Dv* values or other considerations, rate a moderate but significant amount of attention. The largest number of SKUs fall into this category—usually more than 50% of total SKUs, accounting for most of the remaining 50% of the annual dollar usage. Some books on inventory control tend to recommend a somewhat lower portion of total SKUs for the B category. However, because of the availability of computers, we suggest that as many SKUs as possible be monitored and controlled by a computer-based system, with appropriate management-by-exception rules. In Chapters 4, 5, 6, 9, 10, and 11, we present a large number of decision systems that are suitable for use with class B items. Some of the decision rules are also useful for controlling A items, although in the case of A items, such rules are *more* apt to be overruled by managerial intervention. Furthermore, model parameters, such as costs and the estimates of demand, will be reviewed more often for A items.

Class C items are the relatively numerous remaining SKUs that make up only a minor part of the total dollar investment. For these SKUs, decision systems must be kept as simple as possible. One objective of A–B–C classification is to identify this large third group of SKUs that can potentially consume a large amount of data input and managerial time. Typically, for low-value items, most companies try to keep a relatively large number of units on hand to minimize the amount of inconvenience that could be caused by a stockout of such insignificant parts.

For C items especially, and to a lesser degree for the others, as much grouping as possible of SKUs into control groups based on similar annual usage rates, common suppliers, similar seasonal patterns, same end users, common lead times, and so on, is desirable to reduce the total number of discrete decisions that must be processed. Each control group can be designed to operate using a single order rule and monitoring system. For example, if one SKU in the group requires an order because of low inventories, most of the other items will also be ordered at the same time to save on the cost of decision making. Two bin systems, because they require a minimum of paperwork, are especially popular for controlling class C items. They will be discussed in Chapters 6 and 8. Coordinated control systems will be presented in Chapter 10.

While Figure 2.1 is typical, the precise number of members in each of the A, B, and C categories depends, of course, on how spread out the DBV curve actually is. For example, the greater the spread of the distribution, the more SKUs fall into class C.

An A–B–C classification need not be done on the basis of the DBV curve alone. Managers may shift some SKUs among categories for a number of reasons. For example, some inexpensive SKUs may be classified as “A” simply because they are crucial to the operation of the firm. Some large-volume consumer distribution centers plan the allocation of warehousing space on the basis of usage rate and cubic feet per unit. The SKUs with high usage and large cubic feet are stored closer to the retail sales counter. Similarly a distribution by (profit \times volume) per SKU is sometimes used to identify the best-selling products. Items at the lower end of such an A–B–C curve become candidates for being discontinued. Krupp (1994) argues for a two-digit classification, where the first digit is based on dollar usage as in traditional A–B–C analysis, and the second is based on the number of customer transactions. Flores and Whybark (1987) also recommend using a two-dimensional classification where the first is the traditional A–B–C and the second is based on criticality.

In related research, Cohen and Ernst (1988) use a statistical technique called cluster analysis to group items across many dimensions, including criticality. Reynolds (1994) provides a classification scheme, appropriate for process industries, that helps managers focus attention on important items even if they are rarely used. Note that manufacturers and distributors often have different perspectives on the A–B–C classification: An A item for one may be a C item for the other. See for example, Lenard and Roy (1995).

2.5 A Framework for Production Planning and Scheduling

In this section, we develop an important framework, called the product-process matrix, for classifying production planning and scheduling systems. We begin by discussing the product life cycle.

2.5.1 A Key Marketing Concept: The Product Life Cycle

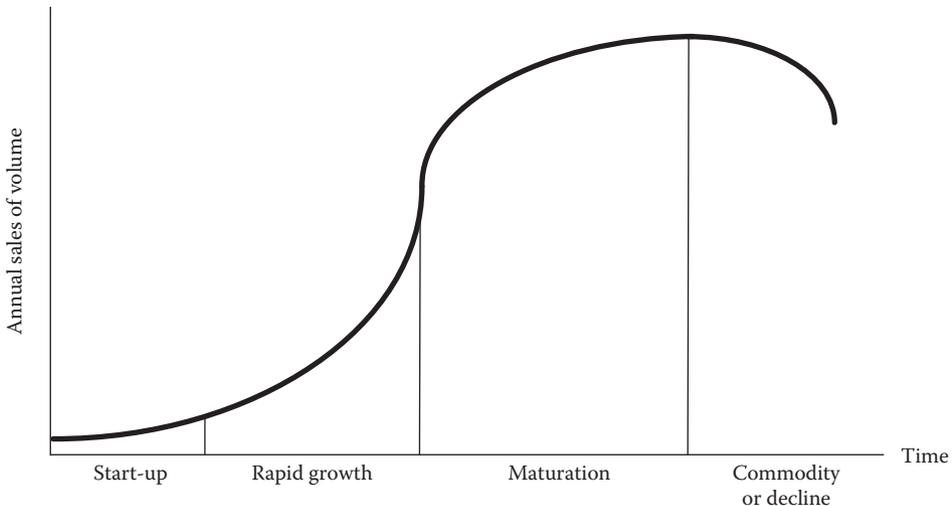
The lifetime sales of many branded products reveal a typical pattern of development, as shown in Figure 2.2. The length of this so-called life cycle appears to be governed by the rate of technological change, the rate of market acceptance, and the ease of competitive entry.

Each year, new clothing styles are introduced in the knowledge that their whole life cycle may last only a season (a few months). Other products, such as commercial aircraft, are expected to be competitive for decades. Some products have been known to begin a new cycle or to revert to an earlier stage as a result of the discovery of new uses, the appearance of new users, or the invention of new features. Kotler (1967) cites the example of television sales, which have exhibited a history of spurts as new sizes of screens were introduced, and the advent of color television, which put sales back to an earlier, rapid growth stage in that industry. As high-definition television becomes popular, and as television is blended with computers for Internet access, it is likely that sales will spurt again. Although the life cycle is an idealized concept, three basic generalizations seem to hold:

1. Products have a limited life. They are introduced to the market, may (or may not) pass through a strong growth phase, and eventually decline or disappear.

2. Product profits tend to follow a predictable course through the life cycle. Profits are absent in the introductory stage, tend to increase substantially in the growth stage, slow down and then stabilize in the maturity and saturation stages, and can disappear in the decline stage.
3. Products require a different marketing, production planning, inventory management, and financial strategy in each stage. The emphasis given to the different functional areas must also change.

When product life cycles are reasonably short, production processes must be designed to be reusable or multipurpose in nature. Alternatively, the amount of investment in single-purpose equipment must be recoverable before the product becomes obsolete. Moreover, the appropriate planning and scheduling systems to use for an individual product are likely to change as the relative importance of the item changes during its life cycle. In particular, special control aspects for the development and decline stages of Figure 2.2 will be discussed in Chapters 3, 5, 7, and 8.



<i>Product variety</i>	Great variety	Increasing standardization	Emergence of a "dominant design"	High standardization, commodity characteristics
<i>Industry structure</i>	Small competitors	Fallout and consolidation	Few large companies	"Survivors"
<i>Form of competition</i>	Product characteristics	Product quality and availability	Price and delivery dependability	Price
<i>Change in growth rate</i>	Little	Increasing rapidly	A little	Decreases rapidly, then slow, or little, if the market is renewed
<i>Change in product design</i>	Very great	Great	Little, unless major renewal of the product	Little
<i>Functional areas most involved</i>	R&D	Engineering, production	Production, marketing, and finance	Marketing, finance

Figure 2.2 The fundamental stages of product/market evolution.

2.5.2 Different Types of Production Processes

As one walks around different factories and talks to factory managers, it becomes clear that there are vast differences among different types of production processes. In this section, we distinguish among four broad classes of processes—job shop, batch flow, assembly line, and continuous process—because the differences among them have important implications for the choice of the production planning and scheduling system. Job shops typically manufacture customized products such as printed circuit boards and certain types of metal parts. A commercial printer, which must tailor each job to the customer's wishes, is also an example of a job shop. Batch flow processes manufacture, for example, apparel, machine tools, and some pharmaceuticals. Assembly products encompass, for example, automobiles, home furnishings, electrical equipment, and computers. Typical illustrations of continuous process (also known as process, continuous flow, or simply flow) products include chemicals (plastics, drugs, soaps, fertilizers, etc.), refined petroleum products, foods and beverages, and paper goods. The differences among these are vast and multidimensional. In what follows, we draw, in particular, on the work of Taylor et al. (1981) and Schmenner (1993).

The contrast in product and market characteristics can be seen in Table 2.2. We highlight just a few points, but we encourage the reader to examine the table in detail. Because of the more standardized nature of products in the process industries, there tends to be more production to stock. Assembly lines can make-to-stock or assemble-to-order, whereas job shops usually make-to-order. Batch flow processes can either make-to-order or make-to-stock depending on the products and customers.

Process industries tend to be the most capital intensive of the four, due to higher levels of automation. Also, process industries tend to have a flow-type layout that is clear but inflexible. That is, materials flow through various processing operations in a fixed routing. However, as we move to the left in this table, the flow increasingly follows numerous, different, and largely unconstrained paths. The production lines in process industries tend to be dedicated to a relatively small number of products with comparatively little flexibility to change either the rate or the nature of the output. In this environment, capacity is quite well defined by the limiting (or pacing/bottleneck) operation. Toward the left of the table, there tends to be more flexibility in both the rate and output. Moreover, both the bottleneck and the associated capacity tend to shift with the nature of the workload (which products are being produced and in what quantities).

Because of the relatively expensive equipment and the relatively low flexibility in output rate, process industries tend to run at full capacity (i.e., 3 shifts per day, 7 days per week) when they are operating. This and the flow nature of the process necessitate highly reliable equipment, which, in turn, normally requires substantial preventive maintenance. Such maintenance is usually carried out when the entire operation is shut down for an extended period of time. Moreover, much longer lead times are typically involved in changing the capacity in a process industry (partly because of environmental concerns, but also because of the nature of the plant and equipment). Job shops and batch flow industries tend to use less-expensive equipment. Maintenance can be done on an “as-needed” basis. Assembly lines are more flexible than continuous processes, but less so than job shops and batch flow lines. Often, the equipment is very inexpensive—simply hand tools; but sometimes automated assembly is used, which is very expensive. Maintenance policies, then, tend to be based on the cost of the equipment.

The number of raw materials used tends to be lower in process situations as compared with assembly; in fact, coordination of raw materials, components, and so on, as well as required labor input, is a major concern in assembly. However, there can be more natural variability in the characteristics of these raw materials in the process context (e.g., ingredients from agricultural or mineral

Table 2.2 Differences in Product and Market Characteristics

<i>Characteristics</i>	<i>Type of Process/Industry</i>			
	<i>Job Shop</i>	<i>Batch Flow</i>	<i>Assembly</i>	<i>Process</i>
Number of customers	Many	Many, but fewer	Less	Few
Number of products	Many	Fewer	Fewer still	Few
Product differentiation	Customized	Less customized	More standardized	Standardized (commodities)
Marketing characteristics	Features of the product	Quality and features	Quality and features or availability/price	Availability/price
Families of items	Little concern	Some concern	Some concern	Primary concern
Aggregation of data	Difficult	Less difficult	Less difficult	Easier
By-products	Few	Few	Few	More
Need for traceability	Little	Intermediate	Little	High
Material requirements	Difficult to predict	More predictable	Predictable	Very predictable
Control over suppliers	Low	Moderate	High	Very high
Vertical Integration	None	Very Little	Some backward, often forward	Backward and forward
Inventories				
Raw materials	Small	Moderate	Varies, frequent deliveries	Large, continuous deliveries
WIP	Large	Moderate	Small	Very small
Finished goods	None	Varies	High	Very high

(Continued)

Table 2.2 (Continued) Differences in Product and Market Characteristics

Characteristics	Type of Process/Industry			
	Job Shop	Batch Flow	Assembly	Process
QC responsibility	Direct labor	Varies	QC specialists	Process control
Production information requirements	High	Varies	Moderate	Low
Scheduling	Uncertain, frequent changes	Frequent expediting	Often established in advance	Inflexible, sequence dictated by technology
Operations challenges	Increasing labor and machine utilization, fast response, and breaking bottlenecks	Balancing stages, designing procedures, responding to diverse needs, and breaking bottlenecks	Rebalancing line, productivity improvement, adjusting staffing levels, and morale	Avoiding downtime, timing expansions, and cost minimization
End-of-period push for output	Much	Frequent	Infrequent	None (can't do anything)
Capital versus labor/material intensive	Labor	Labor and material	Material and labor	Capital
Typical factory size	Usually small	Moderate	Often large	Large
Level of automation	Low	Intermediate	Low or high	High
Number of raw materials	Often low	Low	High	Low
Material requirements	Difficult to predict	More predictable	Predictable	Very predictable
Bottlenecks	Shifting frequently	Shifting occasionally, but predictable	Generally known and stationary	Known and stationary
Speed (units/day)	Slow	Moderate	Fast	Very fast

(Continued)

Table 2.2 (Continued) Differences in Product and Market Characteristics

<i>Characteristics</i>	<i>Type of Process/Industry</i>			
	<i>Job Shop</i>	<i>Batch Flow</i>	<i>Assembly</i>	<i>Process</i>
Process flow	No pattern	A few dominant patterns	Rigid flow pattern	Clear and inflexible
Type of equipment	General purpose	Combination of specialized and general purpose	Specialized, low, or high tech	Specialized, high tech
Flexibility of output	Very	Intermediate	Relatively low (except some assemble to order)	Low
Run length	Very short	Moderate	Long	Very long
Definition of Capacity	Fuzzy, often expressed in dollars	Varies	Clear, in terms of output rates	Clear, expressed in physical terms
Capacity addition	Incremental	Varies	Chunks, requires rebalancing	Mostly in chunks, requires synchronization
Nature of maintenance	As needed	As needed, or preventive when idle	As needed	Shutdown
Energy usage	Low	Low, but can be higher	Low	High
Process changes required by new products	Incremental	Often Incremental	Incremental or radical	Always radical

sources).^{*} Job shops and batch flow operations may have very few inputs as well, but this can vary with the context.

Although there may be relatively few products run on a particular flow line in the process industries, the products do tend to group into families according to a natural sequence. An example would be paints, where the natural sequence of changes would be from lighter to darker colors with a major changeover (once a cycle) from the darkest to the lightest color. As a consequence, in contrast with the other three types of processes, a major concern is ascertaining the appropriate sequence and the time interval between consecutive cycles among the products. The relative similarity of items run on the same line in process industries also makes it easier to aggregate demand data, running hours, and so forth than is the case in the other three.

The flow nature of production in the process industries leads to less WIP inventories than is the case, for example, in job shops. This relative lack of buffering stock, in turn, implies a crucial need for adequate supplies of the relatively few raw materials. However, in this case, the same line of reasoning applies to high volume assembly lines. Job shops tend to have large amounts of WIP, but little finished goods inventory. Without the high levels of WIP, the output of the shop may decrease dramatically. However, because so many job shops manufacture to-order, the products are shipped immediately upon completion, and there is little inventory of finished goods.

Finally, note the operations challenges listed in Table 2.2. Managers in job shops focus on labor and machine utilization and breaking bottlenecks, because the shifting bottlenecks can idle people and equipment. A severe bottleneck can, at certain times, delay shipment of critical customer orders. Batch flow managers focus on balancing stages of production, often because there is a single bottleneck with which other work centers must coordinate. Staffing and line balancing are major concerns in assembly operations. Output tends to be very high, and different operations highly synchronized, so that small changes in the allocation of work to different stations on the line can have a significant impact on profitability. However, assembly line work is tedious, and managers find that they must focus on issues of worker morale. Finally, as is clear from the descriptions of process industries, managers in these industries must keep equipment running with a minimum of downtime.

2.5.3 The Product-Process Matrix

Recall that we are striving to divide production situations into classes that require somewhat different strategies and procedures for planning and scheduling. The product life cycle and the process characteristics are major aids in that direction. However, Hayes and Wheelwright (1979a,b) present a concept that provides us with considerable further insight. They point out that while the product life cycle is a useful framework, it focuses only on marketing. They suggest that many manufacturing processes go through a cycle of evolution, sometimes called the process cycle, beginning with job shops, making few-of-a-kind products, that are labor intensive and generally not very cost efficient, and moving to continuous processes, which are less flexible, automated, and capital intensive. Hayes and Wheelwright suggest that the product life and process life cycle stages cannot be considered separately. One cannot proceed from one level of mechanization to another, for

^{*} Thompson 1988 notes that it is difficult to define a “unit” in process industries, so that on-hand figures in an automated inventory system may be wrong. The product differs by grade of material, potency, and quality specifications. Therefore, adjustments in traditional inventory-recording systems are necessary, including recording shelf life, date of receipt, and expiration date.

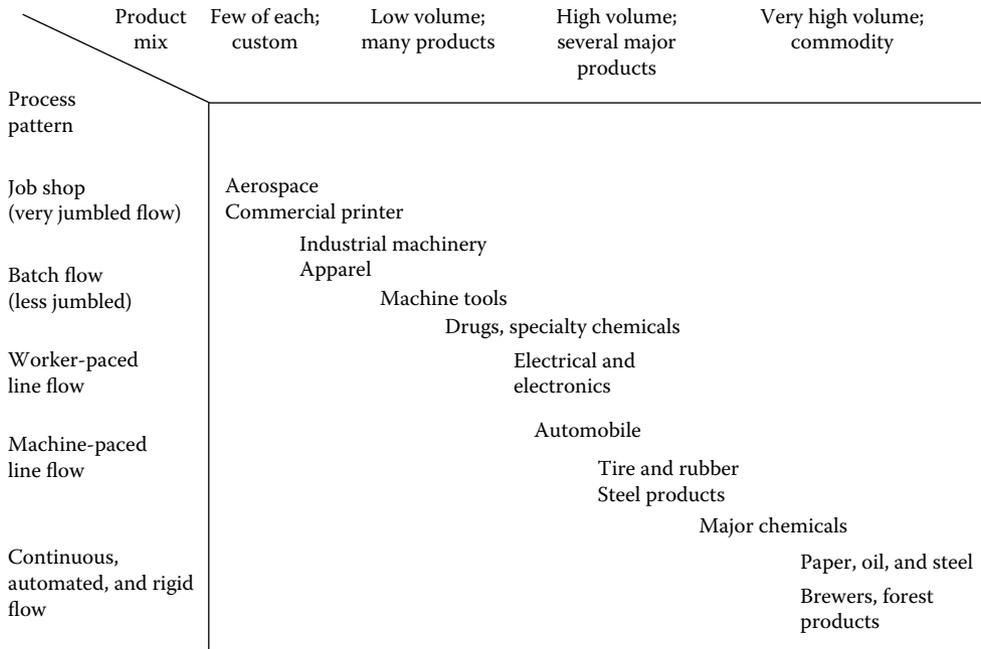


Figure 2.3 The product-process matrix.

example, without making some adjustments to the products and management decision systems involved. Nor can new products be added or others discontinued without considering the effect on the production process.

Hayes and Wheelwright encapsulate the above remarks into a graphical representation known as a product-process matrix. Rather than showing their original form of the matrix, we have portrayed in Figure 2.3 an adapted version (based partly on suggestions in Schmenner (1993), and Taylor et al. (1981)) that is more suitable for our purposes. The columns of the matrix represent the product life cycle phases, going from the great variety associated with startup products on the left-hand side, to standardized commodity products on the right-hand side. The rows represent the major types of production processes. Note that we have now divided assembly into worker-paced and machine-paced flow.*

A number of illustrations are shown in the figure. We see that, by and large, fabrication is in the top-left corner, process industries toward the bottom-right corner, and assembly in the middle. However, there are some exceptions. For example, drugs and specialty chemicals, which are often thought of as process industry products, are centrally located because they are produced in batch flow lines. Likewise, steel products are toward the bottom right, even though they require some fabrication.

* Spencer and Cox (1995) take a slightly different angle on the product-process matrix, defining the vertical axis as the “volume of available capacity devoted to the product line.” At the top of this axis, where we have placed job shops, this value is low, whereas for continuous processes, the value is high. Repetitive manufacturing is then placed between batch and continuous, and is defined as “the production of discrete units in high volume concentration of available capacity using fixed routings.” See also Silver (1996).

Hayes and Wheelwright discuss the strategic implications of nondiagonal positions. Operating in the top-right corner of the matrix can lead to high opportunity costs because the firm is attempting to meet high market demand with a process flow that is designed for low-volume flexibility. The result should be high levels of lost sales. Operating in the lower-left corner can lead to high out-of-pocket costs because the firm has invested in expensive production equipment that is rarely utilized because market demand is low. See also Safizadeh et al. (1996) who survey 144 managers and discover that most factories actually do operate close to the diagonal, and those that do not tend to exhibit poor performance.

To provide additional insight, we draw on the work of Schonberger (1983) who points out that the strategy for production planning and scheduling should depend on how easily one can associate raw material and part requirements with the schedule of end products. As we will now show, there is a direct connection between the position on the product-process matrix and the ease of this association. In the lower right-hand corner of Figure 2.3, the association tends to be quite easy (continuous process systems). This position is, by and large, occupied by capacity-oriented process industries. As we move up to the left and pass through high-volume assembly into lower-volume assembly and batch flow, the association becomes increasingly difficult. In this region, one is dealing primarily with materials and labor-oriented fabrication and assembly industries. Finally, in the top left-hand corner, where fabrication is dominant, the association may again become easier but, because production here is primarily make-to-order, detailed, advanced planning and scheduling are like usually not feasible.

Based on these comments, and others in this section, we can introduce different production planning and scheduling systems that are appropriate for firms located at various positions on the product-process matrix. Because we deal with these systems in depth later in the book, we simply note where they will be discussed. See Figure 2.4 and Table 2.3, where we briefly describe the primary focus of each system.

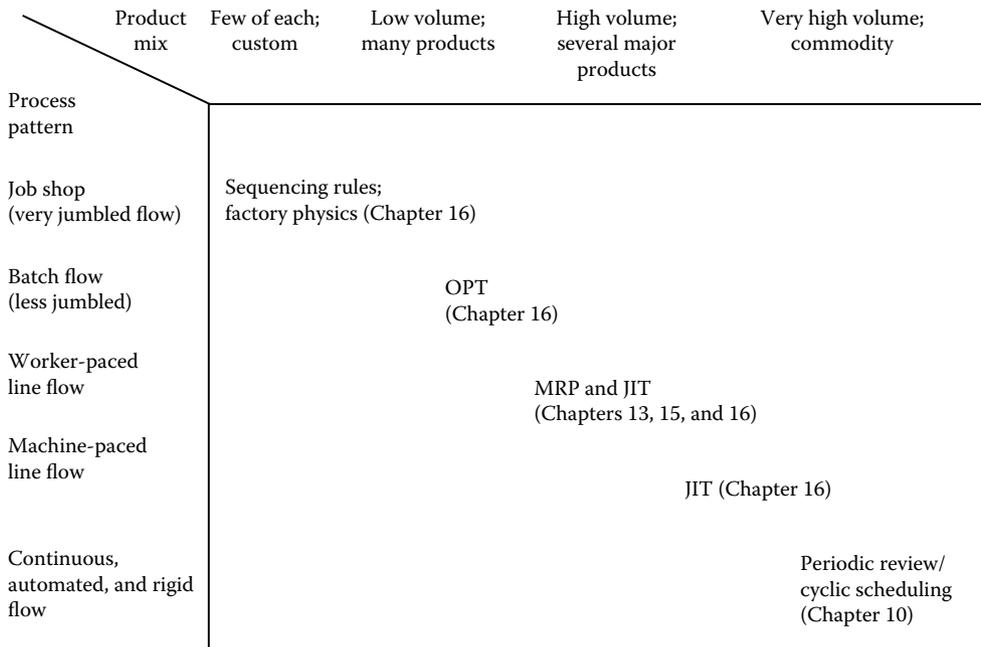


Figure 2.4 Production planning and scheduling and the product-process matrix.

Table 2.3 Production Planning and Scheduling Systems

<i>System</i>	<i>Relevant Chapters of the Book</i>	<i>Nature of Relevant Industries</i>	<i>Primary Focus of the System</i>
Sequencing rules Factory physics	16	Low-volume fabrication	Flexibility to cope with many different orders; Meeting due dates; Increasing throughput; and Predicting lead times
OPT	16	Batch; low-volume assembly	Bottleneck management
Material resources planning (MRP)	13, 15	Medium-volume assembly	Effective coordination of material and labor
JIT	16	High volume, repetitive fabrication, and assembly	Minimizing setup times and inventories; High quality
Periodic review/cyclic scheduling	10	Continuous process	Minimizing sequence-dependent setups; High-capacity utilization

Finally, we note that advanced production technology can allow a firm to move off the diagonal and still be profitable. Flexible manufacturing systems (FMS) achieve flexible, low-volume production with the efficiency of high-volume systems. Automated setups allow rapid changeovers among different products, and computer-controlled machining allows for efficient production. Thus, it is possible to operate profitably to the lower left of the diagonal.

2.6 Costs and Other Important Factors

Through empirical studies and deductive mathematical modeling, a number of factors have been identified that are important for inventory management and production planning and scheduling decisions.

2.6.1 Cost Factors

Here, we briefly describe a number of cost factors. Sections 2.9 and 2.10 will be concerned with estimating their values. The discussion of one other type of cost, the cost of changing the aggregate production rate, or work force size, will be left until Chapter 14 where, for the first time, we deal with medium-range aggregate production-planning problems in which such costs are relevant.

2.6.1.1 The Unit Value or Unit Variable Cost, v

The unit value (denoted by the symbol v) of an item is expressed in dollars per unit. For a merchant* it is simply the price (including freight) paid to the supplier, plus any cost incurred to make it ready for sale. It can depend, via quantity discounts, on the size of the replenishment. For producers, the unit value of an item is usually more difficult to determine. However, one thing is certain; it is seldom the conventional accounting or “book value” assigned in most organizations. The value of an item ideally should measure the actual amount of money (variable cost) that has been spent on the SKU to make it available for usage (either for fulfilling customer demand or for internal usage as a component of some other items). As we will see in Section 2.9, the determination of v is not an easy task. Nevertheless, we shall also see that most models we present are relatively robust to errors in cost estimation. A good starting point is the cost figure given by accounting, adjusted for obvious errors.

As an example, a small retailer buys ski hats for \$20, and then hires a high-school student to unpack them and put them on the shelves. The cost per unit of the student’s time is \$1, so that the value of v should be \$21. A second example is a manufacturer whose cost-accounting system shows that an item costs \$4.32 per unit to manufacture. Upon further examination, however, it is clear that this item is particularly difficult to manufacture, and that a production supervisor devotes up to 30% of her time overseeing the process. The true cost of the item should reflect the value of the supervisor’s time. At 30% of a \$40,000 salary, and 25,000 units produced per year, the cost of the item should be increased by $(\$40,000)(0.30)/(25,000 \text{ units}) = \0.48 per unit.

The unit value is important for two reasons. First, the total acquisition (or production) costs per year clearly depend on its value. Second, the cost of carrying an item in inventory depends on v .

2.6.1.2 The Cost of Carrying Items in Inventory

The cost of carrying items in inventory includes the opportunity cost of the money invested, the expenses incurred in running a warehouse, handling and counting costs, the costs of special storage requirements, deterioration of stock, damage, theft, obsolescence, insurance, and taxes. The most common convention of costing is to use

$$\text{Carrying costs per year} = \bar{I}vr \tag{2.1}$$

where \bar{I} is the average inventory in units (hence $\bar{I}v$ is the average inventory expressed in dollars), and r is the carrying charge, the cost in dollars of carrying one dollar of inventory for 1 year.

By far, the largest portion of the carrying charge is made up of the opportunity costs of the capital tied up that otherwise could be used elsewhere in an organization and the opportunity costs of warehouse space claimed by inventories. Neither of these costs are measured by traditional accounting systems.

The opportunity cost of capital can be defined easily enough. It is, theoretically speaking, the ROI that could be earned on the next most attractive opportunity that cannot be taken advantage of because of a decision to invest the available funds in inventories. Unfortunately, such a marginal cost concept is difficult to implement in practice. For one thing, the next most attractive investment opportunity can change from day to day. Does this mean that the cost of the capital portion of

* We use the term merchant to denote a nonproducer.

carrying should also be changed from day to day? From a theoretical point of view, the answer is yes. In fact, one should theoretically account for the time value of money (a topic we discuss in Chapter 4).^{*} In practice, such factors are difficult to administer; instead, the cost of capital is set at some level by decree and is changed only if major changes have taken place in a company's environment. For example, after due consideration by senior management, a policy is declared that "only investments which expect to earn more than a specified percent can be implemented." The cost of capital used, of course, has to depend on the degree of risk inherent in an investment. (For the same reason, banks charge a higher rate of interest on second mortgages than on the first.) As a result, in practice, the opportunity cost of capital can range from the bank's prime lending rate to 50% and higher. Small companies often have to pay the higher rate because they suffer from severe capital shortages and they lack the collateral to attract additional sources of working capital. A further complication is that inventories are often financed from a mix of sources of funds, each of which may have a different rate. The true cost of carrying the inventory should be based on a weighted average of these rates.

Inventory investment, at least in total, is usually considered to be of relatively low risk because in most cases, it can be converted into cash relatively quickly. However, the degree of risk inherent in inventory investment varies from organization to organization, and the inventory investment itself can change the firm's level of risk (Singhal 1988; Singhal and Raturi 1990; Singhal et al. 1994). Some of the most important impediments to quick conversion to cash can be obsolescence, deterioration, pilferage, and the lack of immediate demand at the normal price. Each of these factors increases the cost of capital over and above the prime rate because of a possible lack of opportunity of quick cash conversion.

The value of r is not only dependent on the relative riskiness of the SKUs, it also depends on the costs of storage that are a function of bulkiness, weight, special handling requirements, insurance, and possibly taxes. Such detailed analysis is seldom applied to all SKUs in inventory. To make the inventory decision more manageable both from a theoretical and a practical point of view, a single value of r is usually assumed to apply for most items. Note from Equation 2.1 that this assumes that more expensive items are apt to be riskier to carry and more expensive to handle or store. Such a convenient relationship doesn't always hold true for all SKUs.[†] Note again that r itself could depend on the total size of the inventory; for example, r would increase if a company had to begin using outside warehousing.

As an example, imagine a company that can borrow money from the bank at 12%. Recent investment opportunities, however, suggest that money available to the firm could earn in the neighborhood of 18%. The value of r should be based on the latter figure, perhaps inflated for the other reasons noted above, because the opportunity cost of the money tied up in inventory is 18%.

2.6.1.3 *The Ordering or Setup Cost, A*

The symbol A denotes the fixed cost (independent of the size of the replenishment) associated with a replenishment. For a merchant, it is called an ordering cost and it includes the cost of order

^{*} The time value of money asserts that a dollar today is worth more than a dollar in the future because it can earn a return in the interim.

[†] In principle, one could instead use $\bar{I}b$ where b is the cost to carry one unit in inventory for a year, or a combination of $I(vr + b)$.

forms, postage, telephone calls, authorization,* typing of orders, receiving, (possibly) inspection, following up on unexpected situations, and handling of vendor invoices. The production setup cost includes many of these components and other costs related to interrupted production. For example, we might include the wages of a skilled mechanic who has to adjust the production equipment for this part. Then, once a setup is completed, there often follows a period of time during which the facility is produced at lower quality or slower speed while the equipment is fine tuned and the operator adjusts to the new part. This is normally called the “learning effect” and the resulting costs are considered to be part of the setup cost because they are the result of a decision to place an order.† Finally, notice that during the setup and learning period, opportunity costs are, in effect, incurred because production time on the equipment is being lost during which some other item could be manufactured. Hence, the latter opportunity cost is only incurred if the production facility in question is being operated at capacity.

Several of these factors can become quite complicated. Consider for a moment the wages of the skilled mechanic who performs the setup. If this person is paid only when setting up a machine, the wages are clearly a part of the setup cost. What happens, however, if the person is on salary, so that whether or not the machine is set up, the wages are still paid? Should these wages be a part of the setup cost? The answer is the subject of much debate in operations and cost accounting. Our answer is that it depends on the use of the mechanic’s time when he is not setting up the machine, and on whether a long-term or short-term perspective is taken. If he would be involved in other activities, including setting up other machines, when he is not setting up the machine for this part, the cost should be included. In other words, there is an opportunity cost for his time. Most factories we have visited are in this situation. People are busy; when they are not working on a particular setup, they are working on other things. There is another side to the story, however. If we decide not to set up the machine for this part, we do not generate actual dollar savings in the short term. The wages are still paid. So, in the short term, the cost of the mechanic’s time is fixed, and therefore, one could argue that it should *not* be a part of the setup cost. Let us assume that the mechanic’s time is used for other activities. The key then is whether a long-term or short-term perspective is taken. A long-term view suggests that the wages should be included because in the long term, this person could be laid off, so that the decision to set up infrequently affects the money the firm pays. (Kaplan (1988) argues that virtually all costs are variable in the long term, because people can be fired, plants closed, and so on.) A short-term view suggests that the wages should not be included. The final issue, then, is whether a long-term or short-term decision is being made. In most cases, the setup cost, A , will be applied to repeated decisions over a long period of time—say a year. We would then argue that the wages of this mechanic should be included in A . In other words, we want to include all costs that are relevant to the decision. To borrow a phrase: “If what you do does not affect the cost, the cost should not affect what you do.” This type of reasoning also applies to many of the other factors listed in the previous paragraph. Further discussion on this topic is in Section 2.9.

Clearly, one could spend months trying to nail down the A value precisely. Often, it is far more beneficial to change the underlying processes that determine A . For example, the widespread adoption of information technology that facilitates computer-to-computer transactions has substantially

* The level of authorization may depend on the dollar size of the replenishment; for example, special authorization from a senior executive may be needed if the order exceeds \$10,000.

† This learning effect often implicitly assumes that the setup process is completely forgotten between production runs.

lowered fixed order setup costs. We shall comment further on the effects of such technologies throughout the book.

2.6.1.4 *The Costs of Insufficient Capacity in the Short Run*

These costs could also be called the costs of avoiding stockouts and the costs incurred when stockouts take place. In the case of a producer, they include the expenses that result from changing over equipment to run emergency orders and the attendant costs of expediting, rescheduling, split lots, and so forth. For a merchant, they include emergency shipments or substitution of a less-profitable item. All these costs can be estimated reasonably well. However, there are also costs, which are much more nebulous, that can result from not servicing customer demand. Will the customer be willing to wait while the item is backordered, or is the sale lost for good? How much goodwill is lost as a result of the poor service? Will the customer ever return? Will the customer's colleagues be told of the disservice? Such questions can, in principle, be answered empirically through an actual study for only a limited number of SKUs. For most items, the risks and costs inherent in disservice have to remain a matter of educated, considered opinion, not unlike the determination of the risks inherent in carrying inventories. In Chapter 6, we examine a number of different methods of modeling the costs of disservice. In every case, however, it is important to explicitly consider the customer's perspective.

2.6.1.5 *System Control Costs*

System control costs are those associated with the operation of the particular decision system selected. These include the costs of data acquisition, data storage and maintenance, and computation. In addition, there are less-tangible costs of human interpretation of results, training, alienation of employees, and so on. Although difficult to quantify, this category of "costs" may be crucial in the choice of one decision system over another.

2.6.2 *Other Key Variables*

Table 2.4[†] provides a rather extensive listing of potentially important variables. We elaborate on just three factors.[‡]

2.6.2.1 *Replenishment Lead Time, L*

A stockout can only occur during periods when the inventory on hand is "low." Our decision about when an order should be placed will always be based on how low the inventory should be allowed to be depleted before the order arrives. The idea is to place an order early enough so that the expected number of units demanded during a replenishment lead time will not result in a stockout very often. We define the *replenishment lead time*, as the time that elapses from the moment at which it

* See Reichheld and Sasser Jr. (1990) who discuss the extremely high cost of losing a customer in service industries. See also Chang and Niland (1967) and Oral et al. (1972).

† Based on A. Saïpe, Partner, Thorne Stevenson & Kellogg, "Managing Distribution Inventories," Executive Development Programme, York University, Toronto, Canada, September 12, 1982.

‡ See Silver (1981), Barancsi and Chikán (1990), Porteus (1990), and Prasad (1994) for other classification schemes, as well as overviews of inventory theory. An interesting aside is that Edgeworth (1888) is the earliest reference we know of a probabilistic inventory theory.

Table 2.4 Inventory Planning Decision Variables

<p><i>Service requirements</i></p> <ul style="list-style-type: none"> • Customer expectations • Competitive practices • Customer promise time required • Order completeness required • Ability to influence and control customers • Special requirements for large customers 	<p><i>Customer-ordering characteristics</i></p> <ul style="list-style-type: none"> • Order timing • Order size • Advanced information for large orders • Extent of open or standing orders • Delay in order processing
<p><i>Demand patterns</i></p> <ul style="list-style-type: none"> • Variability • Seasonality • Extent of deals and promotions • Ability to forecast • Any dependent demand? • Any substitutable products? 	<p><i>Supply situation</i></p> <ul style="list-style-type: none"> • Lead times • Reliability • Flexibility • Ability to expedite • Minimum orders • Discounts (volume, freight) • Availability • Production versus nonproduction
<p><i>Cost factors</i></p> <ul style="list-style-type: none"> • Stockout (pipeline vs. customer) • Carrying costs • Expediting • Write-offs • Space • Spoilage, etc. 	<p><i>Nature of the product</i></p> <ul style="list-style-type: none"> • Consumable • Perishable • Recoverable/repairable
	<p><i>Other issues</i></p> <ul style="list-style-type: none"> • A–B–C pattern • Timing and quality of information • Number of stocking locations • Who bears the cost of inventory?

is decided to place an order, until it is physically on the shelf ready to satisfy customers' demands. The symbol L will be used to denote the replenishment lead time. It is convenient to think of the lead time as being made up of five distinct components:^{*}

1. Administrative time at the stocking point (order preparation time): the time that elapses from the moment at which it is decided to place the order until it is actually transmitted from the stocking point.
2. Transit time to the supplier: this may be negligible if the order is placed electronically or by telephone, but transit time can be several days if a mailing system is used.

^{*} The discussion deals specifically with the case of an external (purchased) replenishment but is also conceptually applicable for internal (production) replenishments.

3. Time at the supplier: if L is variable, this time is responsible for most of the variability. Its duration is materially influenced by the availability of stock at the supplier when the order arrives.
4. Transit time back to the stocking point.
5. Time from order receipt until it is available on the shelf: this time is often wrongly neglected. Certain activities, such as inspection and cataloging, can take a significant amount of time.

We will see that the variability of demand over L can be the result of several causes. If demand varies in a reasonably predictable manner depending on the time of the year, we say it is seasonal. If it is essentially unpredictable, we label the variability as random. Demand over lead time also varies because the time between ordering and receipt of goods is seldom constant. Strikes, inclement weather, or supplier production problems can delay delivery. During periods of low sales, lead times can turn out to be longer than expected because the supplier is accumulating orders to take advantage of the efficiency of longer production runs. When sales are high, longer lead times can result because high demand causes backlogs in the supplier's plant. Finally, in periods of short supply, it may be impossible to acquire desired quantities of items, even after any lead time.

2.6.2.2 *Production versus Nonproduction*

Decisions in a production context are inherently more complicated than those in nonproduction situations. There are capacity constraints at work centers as well as an interdependency of demand among finished products and their components. Gregory et al. (1983) note that the production and inventory planning and control procedures for a firm should depend on (1) whether production is make-to-stock or make-to-order (which, in turn, depends on the relation between customer promise time and production lead time) and (2) whether purchasing is for known production or anticipated production (a function of the purchasing lead time). Chapters 13 through 16 (and to some extent, Chapters 10 and 11) will address production situations.

2.6.2.3 *Demand Pattern*

Besides the factors listed in Table 2.4, we mention the importance of the stage in the product life cycle (Section 2.5.1). Different control procedures are appropriate for new, mature, and declining items. The nature of the item can also influence the demand pattern; for example, the demand for spare parts is likely to be less predictable than the requirements for components of an internally produced item.

2.7 Three Types of Modeling Strategies

Now that we have identified the costs and other key variables, we comment on three types of modeling strategies. First, however, we wish to emphasize that the results of any mathematical analysis must be made consistent with the overall corporate strategy (Section 1.2) and must be tempered by the behavioral and political realities of the organization under study. In particular, a proper problem diagnosis is often more important than the subsequent analysis (see Lockett 1981) and, in some cases, corporate strategy and the behavioral or political circumstances may rule out the meaningful use of any mathematical model.

There are three types of strategies that involve some modelings:

1. Detailed modeling and analytic selection of the values of a limited number of decision variables.
2. Broader-scope modeling, with less attempt at optimization.
3. Minimization of inventories with very little in the way of associated mathematical models.

2.7.1 Detailed Modeling and Analytic Selection of the Values of a Limited Number of Decision Variables

The strategy here is to develop a mathematical model that permits the selection of the values of a limited set of variables so that some reasonable measure of effectiveness can be optimized. A classical example (to be treated in Chapter 4) is the economic order quantity, which, under certain assumptions, minimizes the total of ordering and inventory carrying costs per unit time. In general, a mathematical model may permit a deductive (closed-form) solution, an iterative solution (such as in the Simplex Method of linear programming), or a solution by some form of trial-and-error procedure (such as in the use of a simulation model).

2.7.2 Broader-Scope Modeling with Less Optimization

Here, the strategy is to attempt to develop a more realistic model of the particular situation. However, the added realism often prevents any clearly defined optimization; in fact, there may not even be a mathematically stated objective function. One strives for a feasible solution that one hopes will provide reasonable performance. This is the philosophy underlying Material Requirements Planning (to be discussed in Chapters 13 and 15).

2.7.3 Minimization of Inventories with Little Modeling

Here, the strategy is to attempt to minimize inventories without the help of mathematical models. The Just-in-Time (JIT) and Optimized Production Technology (OPT) philosophies are fall into this category. They strive for elimination of wastes (including inventories) and for continuous improvement. These powerful philosophies, to be discussed in Chapter 16, have since been supplemented with mathematical models that help managers refine and explain their operation. However, the philosophies are not grounded in models, but in an approach to doing business.

2.8 The Art of Modeling

Whichever modeling strategy is chosen, one must attempt to incorporate the important factors, yet keep the model as simple as possible. Incorrect modeling can lead to costly, erroneous decisions. As models become more elaborate, they require more managerial time, take more time to design and maintain, are more subject to problems of personnel turnover, and require more advocacy with top management and everyone else in an organization (i.e., system control costs are increased). There is considerable art in developing an appropriate model; thus, a completely prescriptive approach is not possible. There are some excellent references (see, e.g., Little 1970; Ackoff 1981; Hillier and Lieberman 1990; Powell 1995a,b; Powell and Batt 2011) on the general topic of modeling. In particular, Little recommends that decision models should be understandable

to the decision maker (particularly the underlying assumptions), complete, evolutionary, easy to control, easy to communicate with, robust (i.e., insensitive to errors in input data, etc.), and adaptive. We now list a number of suggestions that analysts may find helpful in modeling complex production/inventory situations:

1. As discussed in Chapter 1, the measures of effectiveness used in a model must be consistent with the objectives of the organization.
2. One is usually not looking for absolute optimization, but instead for significant improvements over current operations. Thus, we advocate the development and use of the so-called heuristic decision rules. These are procedures, based on sound logic, that are designed to yield reasonable, not necessarily mathematically optimal, answers to complex real problems (see Zanakis 1989; Reeves 1993; Powell and Batt 2011).
3. A model should permit results to be presented in a form suitable for management review. Therefore, results should be aggregated appropriately and displayed graphically if possible.
4. One should start with as simple a model as possible, only adding complexities as necessary. The initial, simpler versions of the model, if nothing else, provide useful insights into the form of the solution. We have found it helpful in multi-item problems to first consider the simplest cases of two items, then three items, and so on. In the same vein, it is worthwhile to analyze the special case of a zero lead time before generalizing to a nonzero lead time. We recall the incident of an overzealous mathematician who was employed by a consulting firm. Given a problem to solve involving just two items, he proceeded to wipe out the study budget by first solving the problem for the general case of n items and then substituting $n = 2$ in the very last step. In the same vein, one should introduce probabilistic elements with extreme caution for at least three reasons: (i) the substantially increased data requirements, (ii) the resulting computational complexities, and, perhaps most important, (iii) the severe conceptual difficulties that many practicing managers have with probabilistic reasoning. Throughout the book, we first deal with deterministic situations, and then modify the decision rules for probabilistic environments. Moreover, results are often insensitive to the particular probability distribution used for a specific random variable having a given mean and variance (see Section 6.7.14 of Chapter 6); thus, a mathematically convenient distribution can be used (e.g., the normal distribution described in Appendix II at the end of the book).
5. In most cases in the book, we advocate modeling that leads to analytic decision rules that can be implemented through the use of formulas and spreadsheets. However, particularly when there are dynamic or sequential* effects with uncertainty present (e.g., forecast errors), it may not be possible to analytically derive (through deductive mathematical reasoning) which of two or more alternative courses of action is best to use in a particular decision situation. In such a case, one can turn to the use of simulation. A simulation may be used, which still involves a model of the system. However, now, instead of using deductive mathematical reasoning, one instead, through the model, simulates through time the behavior of the real system under each alternative of interest. Conceptually, the approach is identical with that of using a physical simulation model of a prototype system—for example, the use of a small-scale hydraulic model of a series of reservoirs on a river system. More basically, prior to

* An example of a sequential effect would be the following: if A occurs before B, then C results; but if B occurs before A, then D takes place.

even considering the possible courses of action, simulation can be used to simply obtain a better understanding of the system under study (perhaps via the development of a descriptive analytic model of the system). References on simulation include Fishman (1978) and Kelton and Law (2000).

6. Where it is known *a priori* that the solution to a problem will possess a certain property, this information should be used, if possible, to simplify the modeling or the solution process. For example, in many cases, a very costly event (such as a stockout) will only occur on an infrequent basis. By neglecting these infrequent occurrences, we can simplify the modeling and obtain a straightforward trial solution of the problem. Once this trial solution is found of the above, the assumption is tested. If it turns out to be reasonable, as is often the case, the trial solution becomes the final solution. (See White 1969.)
7. When facing a new problem, one should at least attempt to show an equivalence with a different problem for which a solution method is already known (see Naddor 1978; Nahmias 1981).

2.9 Explicit Measurement of Costs

It often comes as a surprise to technically trained individuals that cost accountants and managers usually cannot determine exactly the costs of some of the variables they specify in their models. That cost measurement is, in practice, a problem that has not been conclusively solved, is evident from the fact that a number of alternative cost-accounting systems are in use. Kaplan (1988) points out that there are three fundamental purposes for cost-accounting systems: valuing the inventory for financial statements, providing feedback to production managers, and measuring the cost of individual SKUs. The choice of a cost system should be consistent with its purpose, and many firms should have more than one system.

One basic problem arises because it is not possible and often is not economical to trace all costs (variable, semivariable, or fixed) to each and every individual SKU in the inventory. An allocation process that distributes, often arbitrarily, fixed or overhead costs across all units is inevitable. (Recall the discussion of the ordering cost A , in Section 2.6.1.) Rosen (1974) categorizes some of the many possible cost systems, and we have summarized these in Table 2.2.

In Table 2.5, column 1, process costing is defined as an accounting method whereby all costs are collected by cost centers such as the paint shop, the warehouse, etc. After a predetermined collection period, each SKU that passed through the process will get allocated a share of the total costs incurred in the cost center. For example, all chairs painted in a particular month could be allocated exactly the same painting cost (which could be different the next month).

Alternatively, *job order costing* could be used. Under this accounting method, a particular order for chairs would be tracked as it progressed through the shop, and all costs incurred by this particular order would be recorded. Thereby, the “cost” of producing the same chair under either of the two costing systems could be very different. The reason for having these two different systems in practice is simple. The job order costing system is more expensive in terms of the amount of bookkeeping required, but it provides information for more detailed cost control than process costing. (Note that the accountant has to select between decision systems for collecting and keeping track of costs just like we must design inventory/production decision systems that have the desired level of sophistication.)

In addition to selecting between bookkeeping systems, the accountant must decide on the basis of valuation. There are four choices listed in column 2 of Table 2.5. Under actual costing, the

Table 2.5 Costing and Control Alternatives

(1)	(2)	(3)
Basic Characteristics of the System	Which Valuation?	Include versus Exclude
	Actual?	Fixed manufacturing
	Predetermined?	Overhead in inventory cost?
	Standard?	
A	A	A
Process-Costing Methods	Actual direct material	Full absorption costing (includes a fixed portion of manufacturing overhead)
	Actual direct labor	
	Actual manufacturing overhead	
B	B	B
Job order costing methods	Actual direct material	Direct variable costing (does not include a fixed portion of manufacturing overhead)
	Actual direct labor	
	Predetermined manufacturing overhead	
	C	
	Actual direct material	
	Predetermined direct labor, overhead	
	D	
	Standard direct material	
	Standard direct labor	
	Standard manufacturing overhead	

product cost v is determined on the basis of accumulated actual cost directly incurred during a given period. Materials, labor, and overhead costs are accumulated, which are and then are spread over all units produced during the period. Note that, as a result, the cost of a given item may vary from one period to another.

Under *standard costing* the materials, labor, and factory overhead are charged against each production unit in accordance with a standard (hourly) rate regardless of how much actual effort (time) it takes. Any deficit (negative variance) between total standard costs charged and actual

costs incurred over a costing period (say a year) is charged as a separate expense item on the income statement and does not end up as a part of the cost of an item, *v*. Alternatively, surpluses (positive variances), that is, where production takes less time than the standard, can artificially inflate *v* unless all items produced are revalued at the end of an accounting year. The possibility of a difference between standard and actual costs results from the fact that standards may have been set for different purposes or they may have been set at ideal, long-run, or perceived “normal” rates that turn out, after the fact, to be unattainable. Also, existing standards may not have been adjusted to account for process improvements and learning.

The term *predetermined* in Table 2.5 refers to the allocation of costs to an SKU based on short-term budgets as opposed to engineered standards. Predetermined cost allocations are usually somewhat arbitrary and may not reflect the amount of expense (as could be determined from a careful engineering study) incurred by an SKU. Predetermined costing has the virtue of ensuring that, from an accounting point of view, all costs incurred over a costing period get allocated somewhere.

Cost-accounting systems are further differentiated in column 3 of Table 2.5 by whether they utilize absorption costing or direct costing. Under *absorption costing* (also called full costing), little distinction is made between fixed and variable manufacturing costs. All overhead costs are charged according to some formula, such as the percent of total direct labor hours used by the product being manufactured. In spite of the fact that direct labor now represents as little as 5% of the total product cost, many firms still allocate overhead on that basis. Activity-based costing (also denoted by ABC) attempts to allocate the actual amount of each overhead resource spent on each SKU. Imagine a firm that produces three products, each of which uses about the same number of direct labor hours. Product X2Z, however, requires a significant amount of effort by supervisors, quality assurance, and expeditors because the production processes are not well suited to it. Allocation of overhead on the basis of direct labor hours will make it appear that the other two products are less profitable than they really are; and product X2Z’s cost will include a smaller portion of overhead than it should. Managers armed with this information may decide to drop a product that is actually very profitable, and to keep a product that actually is losing money. ABC systems have grown in popularity because of these very issues. See also Cooper and Kaplan (1987), Shank and Govindarajan (1988), and Karmarkar and Pitbladdo (1994).

The concept of *direct costing* involves the classification of costs into fixed and variable elements. Costs that are a function of time rather than volume are classified as fixed and are not charged to individual SKUs. For example, all fixed manufacturing expenses, executive and supervisory salaries, and general office expenses are usually considered to be fixed since they are not directly a function of volume. However, note that some costs are in fact semivariable; they vary only with large fluctuations in volume. A doubling of the production rate, for example, could make it necessary to hire extra supervisory personnel.

Semivariable costs (Figure 2.5) in practice are often handled by assuming that volume can be predicted with sufficient accuracy so that, over a relevant range, costs can be viewed as being only of two kinds: variable and fixed. This is not always a simple matter.

It should be clear that the first step in any attempt at explicit cost measurement has to be a determination of what assumptions are made by the existing cost-accounting system within a firm. As Rosen (1974) points out, a small business needing information for external reporting and bookkeeping purposes may choose from Table 2.5 a system (one selection from each of the three columns) that is A–A–A because it may be the least costly from a bookkeeping standpoint. A similar company wishing to postpone tax payments may choose A–A–B, and so forth. Again, in spite of the resistance from financial managers, it may be desirable to use more than one system.

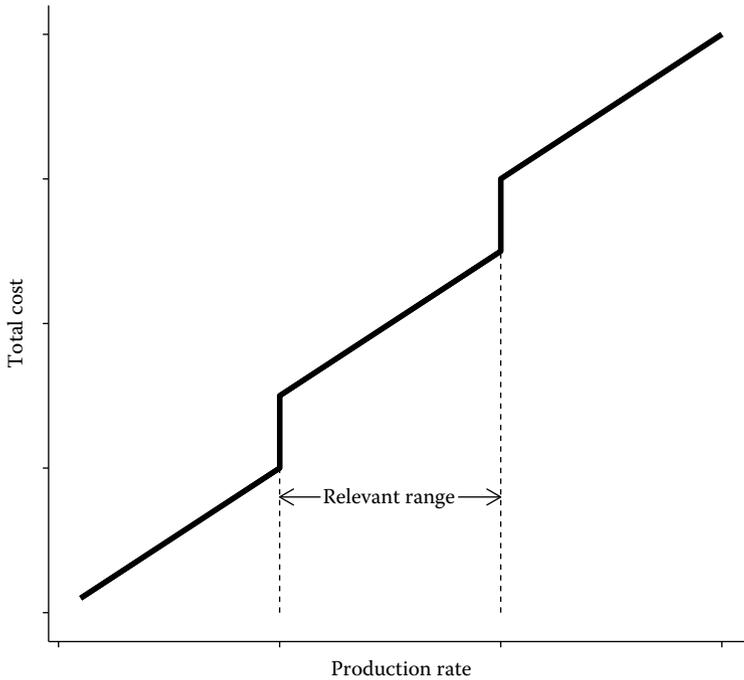


Figure 2.5 Semivariable costs.

Note that while the above description of costing systems provides many options, the most important costs of all, from an inventory-planning standpoint, do not appear at all. Inventory and production-planning decisions require the *opportunity costs*. These costs do not appear in conventional accounting records. Accountants are primarily concerned with the recording of *historical costs*, whereas decision makers must anticipate future costs so that they can be avoided if possible. Existing cost data in standard accounting records, therefore, may not be relevant to decision making or at best have to be recast to be useful.

To a decision maker, *relevant costs* include those that will be incurred in the *future* if a particular action is chosen. They also include costs that could be avoided if a particular action is not taken. Therefore, the overall overhead costs are only relevant if they can be affected by an inventory or production-planning decision. Allocated (predetermined) overhead, based on some ad hoc formula that is useful for financial reporting purposes, generally is not relevant for our purposes. One must therefore carefully examine the costing procedures of an organization from the point of view of their relevancy to the decision at hand.

2.10 Implicit Cost Measurement and Exchange Curves

Suppose we examine carefully the existing accounting systems of a firm and determine the particular species with which we are dealing. Suppose further that we modify the historical cost data to the best of our ability to reflect all relevant costs (including opportunity costs). Then, finally, suppose we use the most advanced techniques of inventory control to determine the appropriate

inventory investment, based on our adjusted costs. What if this “appropriate” total investment in inventories that we propose turns out to be larger than the top management is willing to accept?

Two possible conclusions could be drawn. First, one could argue that we did not correctly determine the relevant costs and that we did not properly capture the opportunity costs. Second, one could argue that the top management is wrong. But what to do? The key lies in realizing that the inventory-planning decision deals with the design of an entire system consisting of an ordering function, a warehousing system, and the servicing of customer demand—all to top management specification. One cannot focus on an individual SKU and ask: What is the marginal cost of stocking this or that individual item without considering its impact on other SKUs or on the system as a whole? Brown (1967, pp. 29–31) argues that there is no “correct” value for r , the cost of carrying inventories, in the explicit measurement sense used by accounting. Instead, the carrying charge r , he says, is a top management policy variable that can be changed from time to time to meet the changing environment. The only “correct” value of r , according to Brown, is the one that results in a total system where aggregate investment, total number of orders per year, and the overall customer-service level are in agreement with the corporate, operations, and marketing strategy. For example, the specification of a low r value would generate a system with relatively large inventory investment, good customer service, and low order replenishment expenses. Alternatively, higher values of r would encourage carrying of less inventories, poorer customer service, and higher ordering costs.

There is of course, no reason, from a theoretical point of view, why all costs (including A , v , and the cost of disservice) could not be considered as policy variables. In practice, this is in effect what is often done, at least partially, by many inventory consultants. An attempt is first made to measure all costs explicitly to provide some baseline data (“ballpark” estimates). Then, the resulting inventory decision system is modified to conform to the aggregate specifications. Only cost estimates that achieve the aggregate specifications are ultimately used during implementation. Such an approach is feasible in practice partly because most of the decision models for inventory management and production planning and scheduling are relatively insensitive to errors in cost measurement. In Chapters 4 and 6, we describe in detail a methodology called exchange curves, for designing inventory decision systems using cost information and policy variables specified by the top management. See also the discussion in Lenard and Roy (1995). For now, we just give a brief overview illustration for the decision of choosing order quantities for a population of items.

For the economic order quantity decision rule (to be discussed in Chapter 4), it turns out that the order quantity of an individual item depends on the ratio A/r . As A/r increases, the order quantity increases; hence, the average inventory level of the item increases and the number of replenishments per year decreases. These same general effects apply for all items in the population under consideration. Therefore, as A/r increases, the total average inventory (in dollars) increases and the total number of replenishments per year decreases. Thus, an aggregate curve is traced as A/r is varied (as shown in Figure 2.6). If the management selects a desired aggregate operating point on the curve, this implies a value of A/r . The use of this A/r value in the economic order quantity formula for each item will give the desired aggregate operating characteristics.

2.11 The Phases of a Major Study of an Inventory Management or Production Planning and Scheduling System

We have now outlined some of the important costs and other factors for inventory and production decisions. The final part of this chapter describes the process of a major (internal or external)

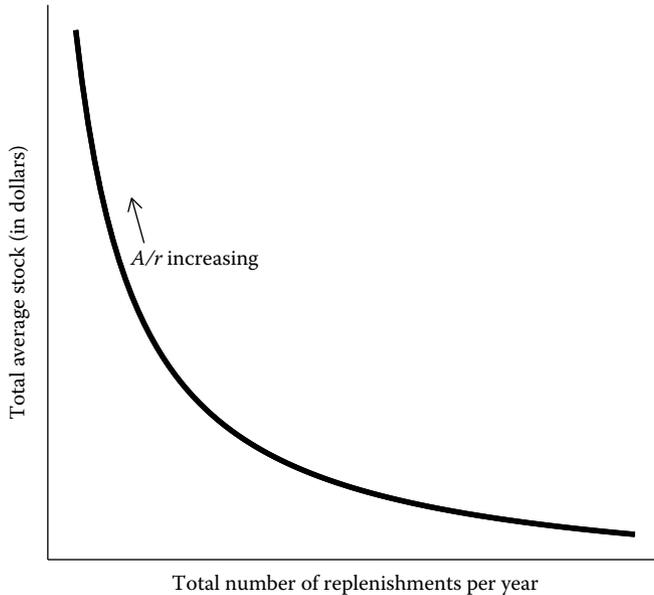


Figure 2.6 Example of an exchange curve.

consulting study. In this process, we strongly recommend building decision systems as a process of organizational intervention, whereby a model builder attempts to improve on the quality of decisions being made. By intervening, the consultant accepts a great responsibility—that of ensuring that the daily routines in the client’s organization are carried on, at least as well as before, during the period of transition to the newly designed decision procedures and thereafter. Acceptance of such a responsibility implies that implementation considerations have to dominate all phases of the design of an operational decision system. In particular, successfully implemented innovations require careful planning, and once in place, they must be maintained, evaluated, and adapted, ad infinitum, through an effective control and evaluation system.

A major study can be an overwhelming task. Therefore, we recommend dividing it into six phases, based on the classic work of Morgan (1963): consideration, analysis, synthesis, choosing among alternatives, control, and evaluation. See Figure 2.7. We briefly present a number of questions and issues that should be addressed in each phase.

2.11.1 Consideration

This first phase focuses on conceptualizing the problem and covers a number of strategic and organizational issues, as well as some detailed modeling concerns. The questions and issues include:

- What are the important operations objectives for this firm? (Recall the four objectives—cost, quality, delivery, and flexibility—discussed in Chapter 1.) In other words, one should perform at least a brief audit of the operations strategy of the firm.
- Who has the overall responsibility for inventory management and production planning and scheduling? Identify the capabilities and limitations of this person as they pertain to the potential goals of the study.

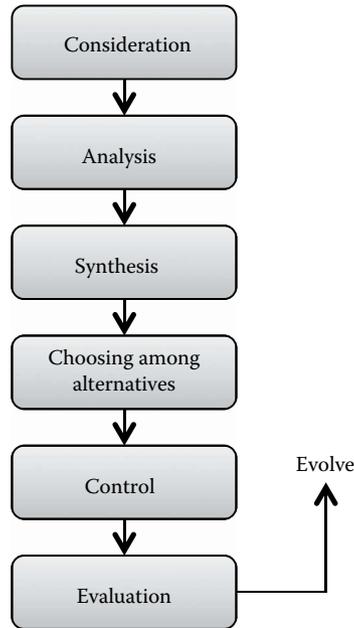


Figure 2.7 Phases of a consulting study.

- What is the organizational structure? In particular, what is the relationship between operations and marketing?
- What are annual sales of the organization?
- What is the current average aggregate inventory level in dollars? This and the previous question help to identify the magnitude of savings possible.
- What are the current inventory management and production planning and scheduling procedures? How are decisions made, and by whom? Are any models used?
- What are the computer resources and skills available?
- What is the SKU? Is it a completed item, or a subassembly that will be finished to order? Are there other alternatives to holding the finished product inventory?
- Is the inventory necessary? Perhaps, it is possible to eliminate some inventories that exist only for historical reasons.
- What modeling strategy (from Section 2.7) seems to be appropriate? One should keep in mind the comments in Section 2.8.

2.11.2 Analysis

The second phase is one of the two modeling stages, and focuses on data collection and a detailed understanding of the uncontrollable and controllable variables. Uncontrollable variables are those, such as the number of items and the level of seasonality, which the consultant or the firm cannot easily change. Controllable variables are those quantities that can be manipulated to achieve the goals. We provide quite a long list that can be used to trigger information gathering. (A reexamination of Table 2.4 may also be helpful.) Questions and issues for uncontrollable variables include:

- How many items are to be studied? If there are thousands of items, it is wise to select, say, every second or fifth item in the SKU list and analyze them first. A study of a small, representative sample can indicate the magnitude of potential savings. It is important, however, that both A and B items are chosen. In addition, it is desirable, if practical, to develop the entire DBV curve. This is easy if current data are available in electronic form. If not, it may not be worth the effort.
- Are the items independent of one another? Dependence may be due to multiple items on a customer order, complementary or substitute items, and shared production equipment.
- What does the supply chain look like? Specifically, can we consider this stocking point in isolation of upstream and downstream locations, or is a multiechelon model necessary due to significant interrelationships with other production or distribution facilities?
- Are customer transaction sizes in single units or batches? If transactions are not necessarily unit sized, are they still discrete or can they take on a continuum of values (e.g., the number of gallons or liters of gasoline demanded by a vehicle in a fill-up)? As we shall see in Chapters 6 and 7, unit-sized transactions are the easiest to model.
- Is demand deterministic or variable? If it is variable, is there a known probability distribution that adequately describes it?
- Do customers arrive at a constant or variable rate? Again, if it is variable, can it be described by a known probability distribution? These latter three bullet points often are considered together by looking at the total demand over some time period, such as a day, week, or month.
- Is the average demand seasonal, or is it somewhat constant over the year?
- What historical data are available? Forecasting models are the most accurate when there is sufficient historical data to calibrate the model.
- How many customers are served? If the number is small, demands may not be independent across time. For example, a large customer order today would reduce the probability of a customer order tomorrow.
- How many, and how powerful, are the firm's competitors?
- What is the typical customer promise time? If customers do not demand immediate delivery, it may be possible to reduce inventories by holding stocks of components, and then assembling to order.
- How is customer service measured? What are the service-level targets?
- How many suppliers serve the firm? Do they offer quantity discounts? Is supply seasonal? Are all the units received of adequate quality to be used?
- What is the replenishment lead time (and its variability) for each item studied?
- Is there a concern about obsolescence or spoilage?
- What is the number, location, and capacity of each storage and manufacturing facility?
- What are the costs that we identified in Section 2.6.1 (A , v , r , and so on)?

For production operations, the following questions are also important:

- How many components go into each end item?
- Through how many processing stages do the products go?
- What is the capacity of the production equipment? How many shifts, hours per week, and weeks per year? Are there any complete shutdowns during the year?
- What are the setup or changeover times for each part? Is there a sequence dependency in these times? What are the processing rates for each part?

- How reliable is the production equipment? Are breakdowns a problem? Does the firm regularly schedule preventive maintenance?
- What is the layout of the factory? In other words, it should be positioned on the product-process matrix.
- What are the yields of the equipment? (Yield is the proportion of a batch that is of acceptable quality.)

Now we examine controllable variables:

- If there is more than one stocking or production location, where should each item be stored or produced?
- How frequently should inventory levels be reviewed?
- When should each item be reordered or produced?
- How much should be ordered or produced? In what sequence, or in what groupings?
- How much productive capacity should be established? Should we recommend hiring, firing, or overtime?
- Should the product price be adjusted?
- Should the service targets be changed?
- What transportation modes, and other factors should be changed?

2.11.3 Synthesis

The second modeling stage, which is the third phase of the process, attempts to bring together the vast amount of information gathered in the previous phase. The consultant now should establish relationships among objectives, uncontrollable variables, and the means of control. For example, how is the expected cost per unit time related to the demand rate, ordering cost, order quantity, and so on? A mathematical objective function stated in terms of controllable variables is the most common result of this phase, although it is not always so. This synthesis phase is a primary focus of the remainder of the book, and therefore we will not devote more space to it here.

2.11.4 Choosing among Alternatives

This phase is the second primary focus of this book. Once we have an objective function, or established relationships among objectives, uncontrollable, and controllable variables, it is necessary to determine reasonable settings of the controllable variables by manipulating the model. This can be done with calculus, search methods, and so on. If the decision is repetitive, it is desirable to develop mathematical decision rules that provide the settings of the controllable variables in terms of the uncontrollable variables. For example, the economic order quantity, to be discussed in Chapter 4, determines the order quantity as a function of A , v , r , and the annual demand rate. These decision rules should also allow for other factors that are not explicitly considered in the model. (See Section 1.5 of Chapter 1.) Moreover, the model should accommodate sensitivity analysis so that managers can get a feel for the result of changes in the input data. If the decision is not repetitive, there is no need to develop a framework for repeated use. A “one-off” decision is adequate.

2.11.5 Control

The control phase is primarily focused on implementation of the decision rules, or more generally, the choices among alternatives. Here, the consultant must ensure that the system functions properly on a tactical or short-range basis. One aspect will be discussed in more detail in Section 2.11.8. Other questions and issues include:

- Training of the staff who will use the new system.
- How often should the values of the control parameters (e.g., the order quantity) be recomputed?
- How will the uncontrollable variables be monitored?
- How will the firm keep track of the inventory levels? In Section 2.11.9, we discuss this issue in more depth.
- How will exceptions be handled? For example, when and how should managers override the decision rules?
- What provision will be made for monitoring sales against forecasts? There are two reasons for overseeing the forecasting system—(a) to take action to prevent serious shortages or overages of stock; and (b) possibly to change the forecast model.
- What actions should be taken when stockouts occur or are imminent? Does the firm have provision for expediting, alerting sales people, finding alternative sources of supply, and so on?
- In the production environment, how will actual production be monitored against scheduled production?

2.11.6 Evaluation

The final phase is evaluation, or measuring how well the control system functions. If the consultant gathered data on actual costs and customer service prior to implementation, which is extremely wise, he or she can now compare them with the actual values under the new system. Likewise, to test the validity of any models, it is useful to compare predicted with actual costs and service. Breaking down the components of the total cost can yield an insight into the sources of any discrepancies.

2.11.7 General Comments

Even a carefully executed study can run into major difficulties that are not the responsibility of the consultant. It is possible, for example, that an external forecasting system provides forecasts that are worse than anticipated. These may reflect badly on the new inventory or production system. The consultant should attempt to identify the reasons for unusually bad, or unusually good, performance. If the model or data are wrong, adjustments can be made. On the other hand, there may be resistance from people at the firm that led to poor performance.

It should be clear that implementation is a long, drawn-out process whose importance and duration are both often underestimated. *Gradual* implementation, accompanied by extensive education, is essential. Where possible, the so-called *pilot approach* should be first utilized. Specifically, the new system should be first implemented on a trial basis on a limited class of items (e.g., on only one of the several product lines manufactured by the firm). As noted above, provision should be made for manual overrides of the output of computerized decision rules. In this regard, Bishop (1974) describes a successful implementation where the apparent key factor that led to the adoption

was a design option within the new decision system that allowed manual overrides of computer decisions by the manager whenever he disagreed with them. “Early in the operation of the system, approximately 80% of the items were overridden by the use of this feature . . . a year later this proportion had declined to less than 10%.” According to Bishop, the inventory control manager, his staff, and his boss used the new decision system simply because they felt that they were in charge, rather than at the mercy, of the innovation.

2.11.8 *Transient Effects*

Each time the parameters of a decision model are adjusted, a series of transient effects result that must be carefully monitored and controlled. Consider the recommendation of an analyst that affects the reallocation of inventory tied up in safety stocks in such a manner that *total inventory investment is not changed in the long run*. The reorder points for some SKUs are raised, for others the reorder points are lowered, and for the balance there is no change. If the management implemented all the new reorder points at the *same* time, some of the newly calculated reorder points, raised from the previous levels, would immediately trigger orders. If there were many changes, a sudden wave of orders could overload a company’s manufacturing plant or order department and also result, after a lead time, in an increase in on-hand dollar inventory investment.

The obvious solution is, of course, not to make all the changes at the same time. Some form of phased implementation strategy that lengthens the period of transition and thereby dampens the disruptions in the system must be chosen. To determine an acceptable transition strategy, the following need to be balanced:

1. The cost of carrying extra inventories during the transition period
2. The expense of a higher than normal ordering rate
3. The opportunity cost of delaying benefits from a new production/inventory decision system that is not implemented all at once
4. Any benefits resulting from the better than normal service level resulting from extra inventory on hand during the transition period

In practice, the changeover to new decision rules is often carried out at the time when an SKU is reordered. This works well if different SKUs are ordered at “random” times. Alternatively, revised decision rules are implemented in phases, starting with A items.

In either case, it is best to try to plan a large transition carefully; in particular, one should closely monitor the resulting stockout rate, the net increase in orders placed, and the rise in total dollar inventory. It is unlikely that the costs of transition can ever be determined and balanced analytically. Therefore, a simulation of transient effects that result from alternative implementation strategies should be undertaken whenever the scope of the problem warrants such an expense.

2.11.9 *Physical Stock Counts*

Physical stock counts, taken annually in many organizations, are intended to satisfy the auditors that account records reasonably represent the value of inventory investment. But the accuracy of stock records is also an important aspect of production/inventory decisions. Most of the decision systems described in this book would be seriously crippled by inaccurate records. See Swann (1989), DeHoratius (2006), and DeHoratius and Raman (2008).

As a check on records, some physical counting of SKUs actually in stock has always been deemed necessary. But physical counts have proven to be expensive and time consuming.* The authors recall the words of a general manager during a tour of a steel-manufacturing operation: “I am deliberately letting the inside inventory drop down to facilitate the annual stock counting.” Performing a physical stock count has in the past in many companies involved could involve shutting down all operations, and thereby losing valuable productive time. Because of this large opportunity cost, many companies have adopted different procedures that achieve the same purpose. Given a limited budget to be spent on corroborating inventory records with the physical inventory, a system that more effectively rations the available clerical resources is known as *cycle counting*.

In cycle counting, as the name might suggest, a physical inventory of each particular item is taken once during each of its replenishment cycles. There are several versions of cycle counting, but probably the most effective one involves counting the remaining physical stock of an SKU just as a replenishment arrives. This has two key advantages. First, the stock is counted at its lowest level. Second, a clerk is already at the stocking position delivering the replenishment, so that a special trip is not needed for counting purposes. Clearly, one drawback is that, if the records are faulty on the high side, we may already have gone into an unexpected stockout position. Note that cycle counting automatically ensures that low-usage SKUs are counted less often than high-usage items because the frequency of reordering increases with sales volume (at least under a decision system based on economic order quantities). Most auditors agree with physically counting low-value items less often than the more expensive ones.

Paperwork, relating to an SKU being counted, can be outstanding at any point in time. When outstanding paperwork exists, a disagreement between counted and recorded inventory may not necessarily mean that the written records are in error. Cycle counts are for this reason in some companies confined to the first thing in the morning, or the last thing in the day, or on the off shift, when outstanding paperwork should be at a minimum. In any case, some form of cutoff rule such as the following is required:

1. If a discrepancy is less than some low percent of the reorder point level, adjust your records to the lower of the two figures and report any loss to accounting.
2. If the discrepancy is greater than some low percent, wait for a few days and request a recount. One hopes by this time that outstanding paperwork will have been processed. If the discrepancy is the same on both counts, we have probably found an error and should adjust the records accordingly. If the second count results in a different discrepancy, the problem requires investigation.

There are some practical disadvantages and problems with cycle counting. Cutoff rules are not easy to follow while normal everyday activity is going on. Checking up on the paperwork in the system, so that stock levels may be properly corroborated, requires considerable perseverance and diplomacy. The responsibility for cycle counting is often assigned to stockroom employees. As a result, too often, cycle counting gets relegated to a lower priority, and may even be discontinued; it may even be discontinued when business activity picks up and stockroom personnel find it difficult to carry out their regular activities as well as stock counting. One solution in such cases is to assign a permanent team to carry out cycle counting all year round. In an interesting study,

* Counting need not be done in a literal sense. An effective, accurate way of estimating the number of units of certain types of products is the use of weigh scales.

Millet (1994) found that inaccuracies in inventory records were often due to inaccurate physical counts, and could be eliminated by providing incentives for those responsible for counting. See also Hovencamp (1995) and Ziegler (1996).

Cycle counting rations effort in proportion to annual dollar usage. It is important to recognize that not all causes of stock discrepancies are a function of dollar usage. The authors once encountered a situation illustrating this point. Items that could be resold on the black market tended to be out of stock more often. Although it could not be proven, it was possible that some of the stolen parts were being resold back to the company through supposedly legitimate wholesalers. Therefore, the accessibility, physical size, and resalability of items, as well as their dollar usage, had to be considered in establishing procedures for stock counts.

Stratified sampling procedures can also be useful for reducing the required sample sizes, particularly for audit purposes. Buck and Sadowski (1983) suggest this approach.

2.12 Summary

In this chapter, we have provided a number of frameworks for the manager and consultant that are helpful in reducing the complexity of the task of managing inventories and production. In the next and last chapter in this part of the book, we turn to the topic of demand forecasting in that an estimate of future demand is clearly needed for all decisions in the production/inventory area.

Problems

- 2.1 Briefly describe the inventory management problem(s) faced by a typical small neighborhood restaurant. Recall Problem 1.5. Point out the differences in the inventory problem(s) faced by a particular store in a more specialized chain such as McDonald's or KFC.
- 2.2 Comment on the following statement: "Optimization models, because they make unrealistic assumptions and are difficult for the user to comprehend, are useless in inventory management or production planning and scheduling."
- 2.3 Discuss the fundamental differences in the approach required to analyze the following two types of inventory control problems:
 - a. A one-opportunity type of purchase that promises large savings
 - b. A situation where the options remain the same from one buying opportunity to the next
- 2.4 It has been argued that the fixed cost of a setup A is not really fixed but can vary throughout the year when the production workload is seasonal. Discuss, and include the implications in terms of the impact on replenishment lot sizes.
- 2.5 Assuming that you have no data on the total annual dollar movement ($\sum Dv$), which of the following SKUs (and under what circumstances) should be classified as A, B, or C items?
 - a. Spare parts for a discontinued piece of manufacturing equipment
 - b. Bolts and washers
 - c. Subassemblies
 - d. Imported SKUs for resale
 - e. Motor oil
 - f. Perishable food stuffs
 - g. A Right Handed PS-37R-01
 - h. Widget invented by the company owner's nephew
 - i. Platinum bushings

If, in addition, you were supplied with figures on the annual dollar movement (Dv) for each SKU, how could such figures affect your classification above? What other information would you like to have? Finally, describe and classify one example of A, B, and C from your experience.

- 2.6 An inventory, mostly made up of machined parts, consisted of 6000 SKUs valued (at full cost) by the accounting department at \$420,000. The company had recently built a new warehouse at a cost of \$185,000 that was financed through a 12% mortgage. The building was to be depreciated on a 25-year basis. The company's credit rating was sound, and a bank loan of \$50,000 was under negotiation with the bank. The main operating costs per year in the new warehouse were estimated to be as follows:

Municipal and other taxes	\$5,732.67
Insurance on building and contents	3,200.00
Heating/air conditioning	12,500.00
Electricity and water	3,200.00
Labor (supervisor, three clerks, and a half-time janitor) ^a	69,000.00
Pilferage	5,000.00
Obsolescence	5,000.00
Total	\$103,632.67

^a Includes 60% manufacturing overhead.

It was estimated that on average, each SKU involved one dollar of labor cost per dollar of material cost.

- Recommend a value for the carrying cost r in \$/\$/year.
 - Suppose that your recommended value for r is accepted by the management and the accounting department. You proceed to calculate the economic order quantities for all 6000 items and discover that the new warehouse is too small to physically accommodate all the inventory that is indicated by your calculations. What action would you take now?
 - Alternatively suppose that your recommended value for r is accepted by the management and the accounting department, but that the total dollar investment (based on economic order quantities) that you estimate will be needed is approximately \$800,000. The top management is unwilling to have a total investment greater than \$500,000. What action would you take under such a circumstance?
- 2.7 Use Figures 2.8 through 2.10 for this problem. Assume that the raw materials and purchased parts each represent 30% of the cost of goods sold and that value added (labor plus factory overhead) constitutes the remaining 40%. Also assume that each production cycle involves the production of 1000 units of finished goods, each valued (at cost) at \$10/unit. In each case depicted in the figures, ascertain the appropriate *overall* value of $\bar{I}v$ (i.e., the average inventory level, expressed in dollars). Also describe, in terms of the production and demand processes, how each of the patterns could result.

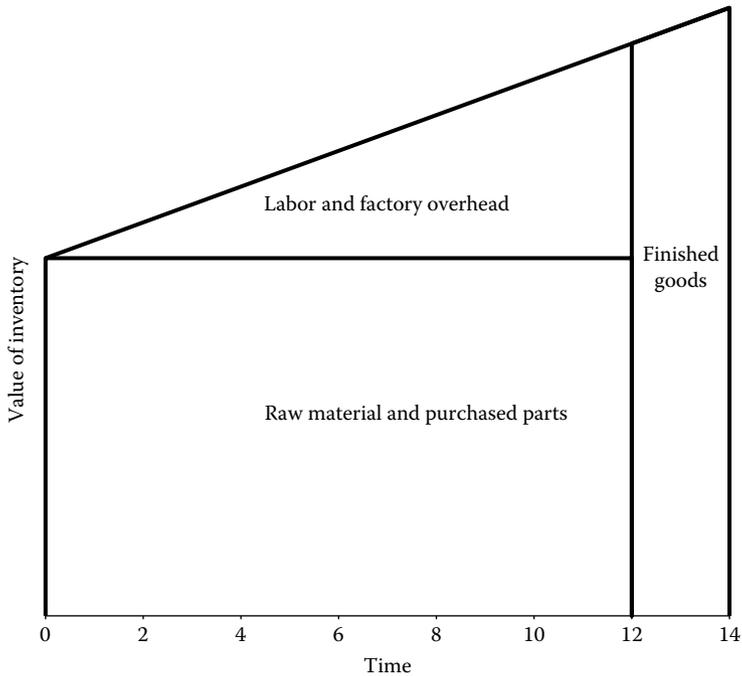


Figure 2.8 Case a for inventory valuation Problem 2.7.

- 2.8 What parameters that affect the different functional categories of inventories could be changed? For example, cycle stocks are determined in part by the setup cost, and it can be changed. Give other examples, and note how the parameters could change.
- 2.9 Have a class representative (or representatives) select a random sample of 30 items stocked in your university bookstore. For each item, obtain an estimate of Dv (from records or from the manager of the store). Prepare a DBV table and a graph. See Table 2.1 and Figure 2.1.
- 2.10 The type of operation characterized by a large volume of identical or similar products flowing through identical stages of production is typical in which of the following industries?
 - a. Aerospace
 - b. Automobiles
 - c. Machine tools

For such a situation, which type of production planning and scheduling system is appropriate? Explain your reasoning.
- 2.11 An important problem in the operation of a retail store is the allocation of limited shelf space among a wide variety of products.
 - a. Discuss how demand is influenced by the allocation. Can individual items be treated separately in this regard?
 - b. How are inventory-related costs of an item affected by the space allocated?
- 2.12 Consignment stock is stock held by a retailer but owned by the supplier until it is sold by the retailer. For the same SKU, discuss how you would model the associated inventory carrying costs of the retailer (i) without consignment and (ii) with consignment.

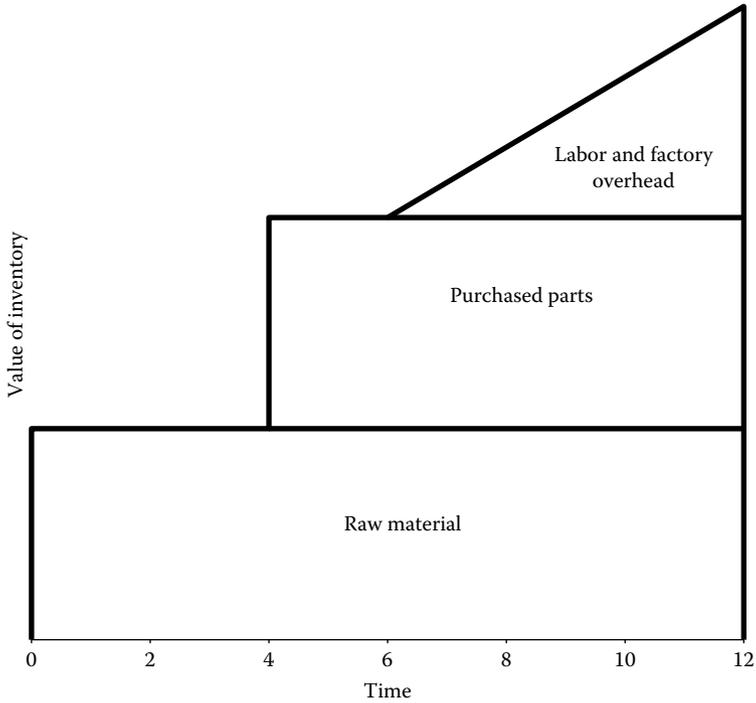


Figure 2.9 Case b for inventory valuation Problem 2.7.

- 2.13 a. Construct a spreadsheet that permits the development of a DBV table and a graph. Input data should be for each of n items (where, for any particular application, n would be specified by the user):
- i. Item identification
 - ii. D_i , the annual usage of item i , in units/year
 - iii. v_i , the unit variable cost of item i , in \$/unit
- b. Illustrate the use of your program for the following example involving 20 items:

Item, i	ID	D_i	v_i
1	A	80	422.33
2	B	514	54.07
3	C	19	0.65
4	D	2,442	16.11
5	E	6,289	4.61
6	F	128	0.63
7	G	1,541	2.96
8	H	4	22.05

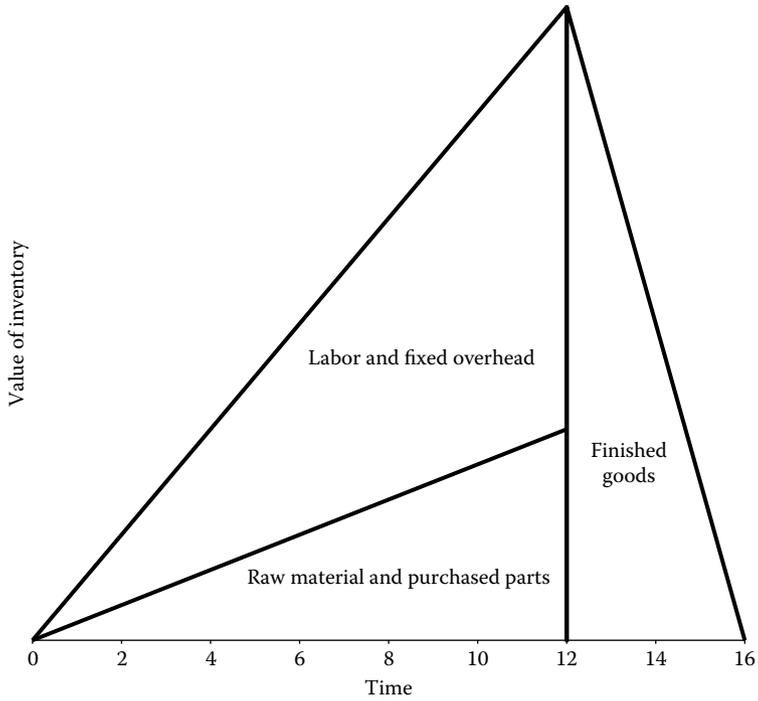


Figure 2.10 Case c for inventory valuation Problem 2.7.

9	I	25	5.01
10	J	2,232	2.48
11	K	2	4.78
12	L	1	38.03
13	M	6	9.01
14	N	12	25.89
15	O	101	59.50
16	P	715	20.78
17	Q	1	2.93
18	R	35	19.52
19	S	1	28.88
20	T	4	29.86

What percentage of the total dollar usage is contributed by the top 20% of the items and by the bottom 50% of the items and

2.14 Repeat Problem 2.13 for the following data:

<i>SKU ID</i>	<i>Annual Usage in Units</i>	<i>Unit Variable Cost (\$/Unit)</i>
A	172	27.73
B	8	110.40
C	3	60.60
D	4	19.80
E	2	134.34
F	15	160.50
G	4	49.48
H	3	8.46
I	4	40.82
J	1	34.40
K	212	23.76
L	4	53.02
M	12	71.20
N	2	67.40
O	2	37.70
P	1	28.80
Q	4	78.40
R	12	33.20
S	3	72.00
T	18	45.00
U	50	20.87
V	19	24.40
W	1	48.30
X	10	33.84
Y	27	210.00
Z	33	73.44
1A	48	55.00
1B	5	30.00

SKU ID	Annual Usage in Units	Unit Variable Cost (\$/Unit)
1C	210	5.12
1D	27	7.07
1E	10	37.05
1F	12	86.50
1G	48	14.66
1H	117	49.92
1I	12	47.50
1J	94	31.24
1K	4	84.03
1L	7	65.00
1M	2	51.68
1N	4	56.00
1O	12	49.50
1P	2	59.60
1Q	100	28.20
1R	2	29.89
1S	5	86.50
1T	60	57.98
1U	8	58.45

- 2.15 Select 30 items at random from an inventory population in a local organization. These could include the campus bookstore or supplies in a department office. For each item, ascertain the value of Dv and develop a DBV table and graph. What percent of the total dollar usage is contributed by the top 20% of the items (ranked by dollar usage)?
- 2.16 One of the major cost categories in this chapter is called “carrying costs.” See Section 2.6.1. Until recently people did not appreciate that some quality-related costs could be appropriately included in this category. Indicate why annual quality-related costs might change with the production lot size.
- 2.17 For a local organization, identify the major cost factors included in Section 2.6.1, including the fixed cost of ordering or setup, A , the carrying charge, r , and the unit variable cost, v , for just one item. Describe the components of each. For A and r , indicate briefly why you decided to include each component. What components did you ignore in your analysis?

Appendix 2A: The Lognormal Distribution

2A.1 Probability Density Function and Cumulative Distribution Function

The density function of a lognormal variable x is given by*

$$f_x(x_0) = \frac{1}{bx_0\sqrt{2\pi}} \exp\left[-\frac{(\ln x_0 - a)^2}{2b^2}\right] \quad (2A.1)$$

where a and b are the two parameters of the distribution.

We can see that the range of the lognormal is from 0 to ∞ . It has a single peak in this range. A sketch of a typical lognormal distribution is shown in Figure 2A.1. The name is derived from the fact that the logarithm of x has a normal distribution (with mean a and standard deviation b).

The cumulative distribution function

$$p_{x \leq}(x_0) = \int_{-\infty}^{x_0} f_x(z) dz$$

can be shown to be given by

$$p_{x \leq}(x_0) = 1 - p_{u \geq}\left(\frac{\ln x_0 - a}{b}\right) \quad (2A.2)$$

where $p_{u \geq}$ is the probability that a unit normal variable takes on a value of u_0 or larger, a function tabulated in Appendix II.

2A.2 Moments

One can show (See Aitchison and Brown 1957) that the expected value of the j th power (for any real j) of x is given by†

$$E(x^j) = m^j \exp[b^2 j(j-1)/2] \quad (2A.3)$$

where

$$m = \exp[a + b^2/2] \quad (2A.4)$$

is the mean or expected value of x . The variance of x defined by

$$\sigma_x^2 = E(x^2) - [E(x)]^2$$

can be shown, through the straightforward use of Equations 2A.3 and 2A.4, to be

$$\sigma_x^2 = [E(x)]^2 [\exp(b^2) - 1]$$

* $f_x(x_0)dx_0$ represents the probability that x takes on a value between x_0 and $x_0 + dx_0$.

† In general, $E(x^j) = \int_{-\infty}^{\infty} x_0^j f_x(x_0) dx_0$.

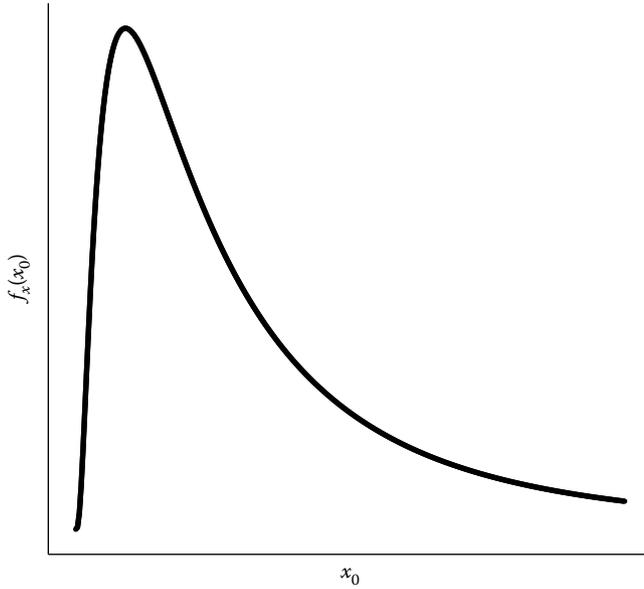


Figure 2A.1 The lognormal probability density function.

Hence, the standard deviation σ_x is given by

$$\sigma_x = E(x) \sqrt{\exp(b^2) - 1}$$

and the coefficient of variation by

$$\sigma_x/E(x) = \sqrt{\exp(b^2) - 1}$$

the latter depending only on the value of b .

If, as in a DBV of a population of n items, x is actually a discrete variable that takes on the values x_i ($i = 1, 2, \dots, n$) and the x_i 's can be represented by a lognormal distribution with parameters m and b , then

$$E(x^j) = \frac{\sum_{i=1}^n x_i^j}{n} \tag{2A.5}$$

and from Equation 2A.3 it follows that

$$\sum_{i=1}^n x_i^j = nm^j \exp[b^2 j(j-1)/2] \tag{2A.6}$$

Note that the parameter m can be estimated using the sample mean, that is,

$$m = \frac{\sum_{i=1}^n x_i}{n} \tag{2A.7}$$

2A.3 Testing the Fit of a Lognormal

One can easily test the fit of a lognormal distribution to a set of data by taking natural logs and then using a normal probability plot routine. Such routines are available in most statistical software packages. See Section 3.5.5 for further discussion. Herron (1976) also provides an alternative approach.

References

- Ackoff, R. 1981. The art and science of mess management. *Interfaces* 11(1), 20–26.
- Aitchison, J. and J. A. C. Brown. 1957. *The Lognormal Distribution*. Cambridge, England: Cambridge University Press.
- Barancsi, É. and A. Chikán. 1990. *Inventory Models*, Volume 16. Dordrecht, the Netherlands: Kluwer Academic Publishers.
- Bishop, J. L. 1974. Experience with a successful system for forecasting and inventory control. *Operations Research* 22(6), 1224–1231.
- Brown, R. G. 1967. *Decision Rules for Inventory Management*. New York: Holt, Rinehart and Winston.
- Brown, R. G. 1982. *Advanced Service Parts Inventory Control*. Norwich, VT: Materials Management Systems, Inc.
- Buck, J. R. and R. P. Sadowski. 1983. Optimum stratified cycle counting. *IIE Transactions* 15(2), 119–126.
- Chang, Y. S. and P. Niland. 1967. A model for measuring stock depletion costs. *Operations Research* 15(3), 427–447.
- Cohen, M. A. and R. Ernst. 1988. Multi-item classification and generic inventory stock control policies. *Production and Inventory Management Journal* 29(3), 6–8.
- Cooper, R. and R. S. Kaplan. 1987. How cost accounting systematically distorts product costs. In W. J. J. Bruns and R. S. Kaplan (Eds.), *Accounting and Management: Field Study Perspectives*, pp. 204–228. Cambridge, MA: Harvard Business School Press.
- DeHoratius, N. 2006. Inventory record inaccuracy in retail supply chains. In J. J. Cochran, L. A. Cox, P. Keskinocak, J. P. Kharoufeh, and J. C. Smith (Eds.), *Wiley Encyclopedia of Operations Research and Management Science*, pp. 1–14. Hoboken, NJ: John Wiley & Sons, Inc.
- DeHoratius, N. and A. Raman. 2008. Inventory record inaccuracy: An empirical analysis. *Management Science* 54(4), 627–641.
- Edgeworth, F. Y. 1888. The mathematical theory of banking. *Journal of the Royal Statistical Society* 51, 113–127.
- Fishman, G. S. 1978. *Principles of Discrete Event Simulation*. New York: Wiley-Interscience.
- Flores, B. E. and D. C. Whybark. 1987. Implementing multiple criteria ABC analysis. *Journal of Operations Management* 7(1–2), 79–85.
- Gregory, G., S. A. Klesniks, and J. A. Piper. 1983. Batch production decisions and the small firm. *Journal of the Operational Research Society* 34(6), 469–477.
- Hayes, R. and S. Wheelwright. 1979a. Link manufacturing process and product life cycles. *Harvard Business Review* 57(1), 133–140.
- Hayes, R. and S. Wheelwright. 1979b. The dynamics of process-product life cycles. *Harvard Business Review* 57(2), 127–136.
- Herron, D. P. 1976. Industrial engineering applications of ABC curves. *AIIE Transactions* 8(2), 210–218.
- Hillier, F. S. and G. J. Lieberman. 1990. *Introduction to Mathematical Programming*. New York: McGraw-Hill.
- Hovencamp, D. P. 1995. Where is it? *APICS* (July), 50–54.
- Kaplan, A. J. 1988. Bayesian approach to inventory control of new parts. *IIE Transactions* 20(2), 151–156.
- Karmarkar, U. and R. Pitbladdo. 1994. Product-line selection, production decisions and allocation of common fixed costs. *International Journal of Production Economics* 34, 17–33.
- Kelton, W. D. and A. M. Law. 2000. *Simulation Modeling and Analysis*. New York: McGraw Hill Boston.
- Kotler, P. 1967. *Marketing Management: Analysis, Planning, and Control*. Englewood Cliffs, NJ: Prentice-Hall.

- Krupp, J. A. G. 1994. Are ABC codes an obsolete technology? *APICS—The Performance Advantage* (April), 34–35.
- Lenard, J. D. and B. Roy. 1995. Multi-item inventory control: A multicriteria view. *European Journal of Operational Research* 87, 685–692.
- Little, J. 1970. Models and managers: The concept of a decision calculus. *Management Science* 16(8), 466–485.
- Lockett, G. 1981. The management of stocks—Some case histories. *OMEGA* 9(6), 595–604.
- Millet, I. 1994. A novena to Saint Anthony, or how to find inventory by not looking. *Interfaces* 24(2), 69–75.
- Morgan, J. 1963. Questions for solving the inventory problem. *Harvard Business Review* 41(1), 95–110.
- Naddor, E. 1978. Sensitivity to distributions in inventory systems. *Management Science* 24(16), 1769–1772.
- Nahmias, S. 1981. Approximation techniques for several stochastic inventory models. *Computers and Operations Research* 8(3), 141–158.
- Oral, M., M. Salvador, A. Reisman, and B. Dean. 1972. On the evaluation of shortage costs for inventory control of finished goods. *Management Science* 18(6), B344–B351.
- Porteus, E. L. 1990. The impact of inspection delay on process and inspection lot sizing. *Management Science* 36(8), 999–1007.
- Powell, S. G. 1995a. Six key modeling heuristics. *Interfaces* 25(4), 114–125.
- Powell, S. G. 1995b. Teaching the art of modeling to MBA students. *Interfaces* 25(3), 88–94.
- Powell, S. G. and R. J. Batt. 2011. *Modeling for Insight: A Master Class for Business Analysts*. Hoboken, NJ: John Wiley & Sons.
- Prasad, S. 1994. Classification of inventory models and systems. *International Journal of Production Economics* 34, 209–222.
- Reeves, C. R. 1993. Improving the efficiency of tabu search for machine sequencing problems. *Journal of the Operational Research Society* 44, 375–382.
- Reichheld, F. R. and W. E. Sasser Jr. 1990. Zero defections: Quality comes to services. *Harvard Business Review* 68(5), 105–111.
- Reynolds, M. P. 1994. Spare parts inventory management. *APICS—The Performance Advantage* 4(4), 42–46.
- Rosen, L. S. 1974. *Topics in Managerial Accounting* (2nd ed.). Ryerson, Toronto: McGraw-Hill.
- Safizadeh, M. H., L. P. Ritzman, D. Sharma, and C. Wood. 1996. An empirical analysis of the product-process matrix. *Management Science* 42(11), 1576–1591.
- Schmenner, R. W. 1993. *Production/Operations Management: From the Inside Out* (5th ed.). New York: Macmillan.
- Schonberger, R. J. 1983. Applications of single-card and dual-card kanban. *Interfaces* 13(4), 56–67.
- Shank, J. K. and V. Govindarajan. 1988. The perils of cost allocation based on production volumes. *Accounting Horizons* 2, 71–79.
- Silver, E. A. 1981. Operations research in inventory management: A review and critique. *Operations Research* 29(4), 628–645.
- Silver, E. A. 1996. A concern regarding the revised product-process matrix. *International Journal of Production Research* 34(11), 3285.
- Simon, H. A. 1957. *Models of Man, Social and Rational*. New York: John Wiley & Sons.
- Singhal, V. R. 1988. Inventories, risk, and the value of the firm. *Journal of Manufacturing and Operations Management* 1, 4–43.
- Singhal, V. R. and A. S. Raturi. 1990. The effect of inventory decisions and parameters on the opportunity cost of capital. *Journal of Operations Management* 9(3), 406–420.
- Singhal, V. R., A. S. Raturi, and J. Bryant. 1994. On incorporating business risk into continuous review inventory models. *European Journal of Operational Research* 75, 136–150.
- Spencer, M. S. and J. F. Cox. 1995. An analysis of the product-process matrix and repetitive manufacturing. *International Journal of Production Research* 33(5), 1275–1294.
- Swann, D. 1989. Silver bullets and basics. *Production and Inventory Management Review* (January), 34–35.
- Taylor, S., S. Seward, and S. Bolander. 1981. Why the process industries are different. *Production and Inventory Management Journal* 22(4), 9–24.

- White, D. 1969. Problems involving infrequent but significant contingencies. *Operational Research Quarterly* 20(1), 45–57.
- Zanakis, S., J. Evans, and A. Vazacopoulos. 1989. Heuristic methods and applications: A categorized survey. *European Journal of Operational Research* 43, 88–110.
- Ziegler, D. 1996. Selecting software for cycle counting. *APICS November*, 44–47.
- Zipkin, P. 1995. Processing networks with planned inventories: Tandem queues with feedback. *European Journal of Operational Research* 80, 344–349.

Chapter 3

Forecasting Models and Techniques

Inventory and production-planning decisions nearly always involve the allocation of resources in the presence of demand uncertainty. For inventory decisions, financial resources must be deployed to procure goods in the anticipation of a future sale of those goods. Similarly, for production planning, capacity must be committed to certain products in anticipation of their future sale. Effective decision making in inventory management and production planning requires accurate descriptions of possible demands in future periods. As noted earlier in Chapter 1, variability can be a key driver of inventory expense. A study at a large high-tech firm revealed that 40% of the inventories in their system were cycle and pipeline stocks, whereas 60% were due to variability. Of the inventories driven by variability, 2% were held due to supplier performance variability, 2% were held due to manufacturing variability, and 96% were held due to demand uncertainty. A similar study at a large chemical producer revealed that 45% of finished goods inventory were held due to variability, again with the majority of that inventory held due to demand variability. These two examples highlight that while the relative impact of demand uncertainty on the inventory will vary across industrial settings, demand uncertainty will almost always be a significant factor for inventory management.

Forecasts can be based on a combination of an extrapolation of what has been observed in the past (what we call statistical forecasting) and informed judgments about future events. Informed judgments can include the knowledge of firm orders from external customers, preplanned shipments between locations within a single organization, or preplanned usage of service parts in preventive maintenance. They also include marketing judgments such as the effects of promotions, competitor reactions, general economic conditions, and so on. For items or groups of related items with accurately maintained historical demand data, we can apply techniques known as *time-series* forecasting models. With a time-series forecasting model, we use the demand history for an item (or group of items) to construct a forecast for the future demand of that item (or group). Such time-series methods are commonly used in practice and, when the models are appropriately selected and calibrated, they can be quite effective in generating high-quality forecasts. The majority of this chapter is dedicated to such time-series methods. In particular, we focus on generating forecasts for a single item using time-series methods.

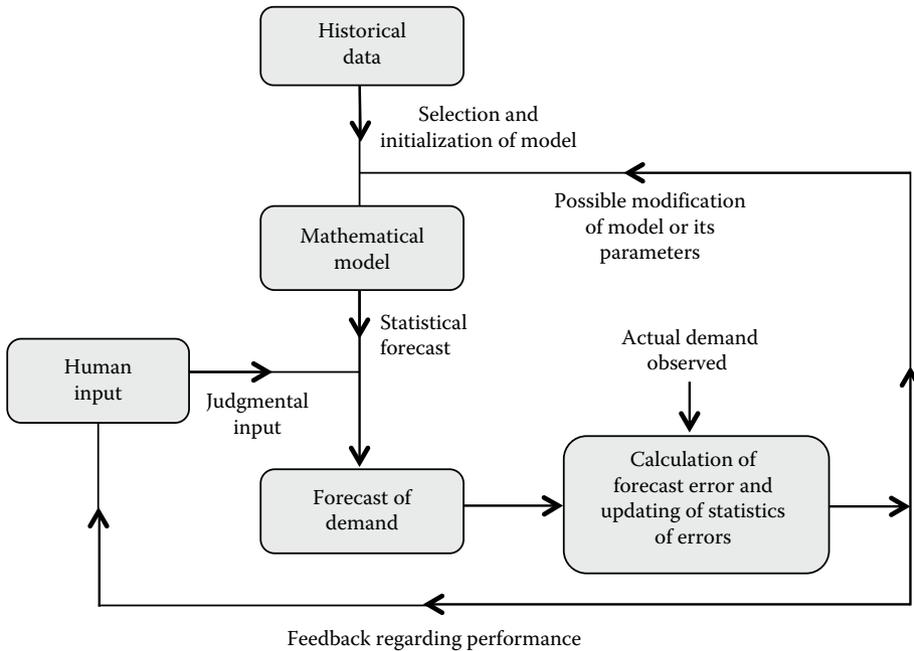


Figure 3.1 A suggested forecasting framework.

The overall framework of a suggested forecasting system is shown in Figure 3.1. The human judgment we have mentioned is a crucial ingredient. Note also that, as the actual demand in a period is observed, it is compared with the earlier forecast so that we can measure the associated error in the forecast. It is important to monitor these errors for at least three reasons. First, the quantity of safety (or buffer) stock needed to provide adequate customer service will depend (as we will see in Chapter 6) on the sizes of the forecast errors. Second, the statistical forecast is based on an assumed underlying mathematical model with specific values for its parameters. The sizes and directions (plus or minus) of the forecast errors should suggest possible changes in the values of the parameters of the model or even in the form of the model itself (e.g., introducing a seasonal component, when it was not already present). Finally, the errors can provide a monitor and feedback on the performance of the subjective input component of the forecasts.

Forecasting is of course a broad topic, and forecasts are used in many contexts from weather predictions, predictions of asset prices, financial or budgeting forecasts, etc. Given the scope of this book, we focus on forecasts of product demand used to support production and inventory resource allocation decisions although we note that many of the forecasting techniques discussed can be applied to other settings. It is often said that the one thing we know about forecasts is that they will be in error. This tends to be true if one views the forecast as a single number—often called a point estimate—but to support an inventory or production decision, we need more than just a single number. We need a richer description of the set of possible demand scenarios. Specifically, we define a forecast as a *description* of future demand scenarios used to *support a resource allocation decision*. Throughout this chapter, we first focus on measuring the performance of the point-estimate forecast. Later, we will discuss how we can use measures of this point-estimate forecasting process to help complete our description.

No matter how forecasts are generated, we are interested in evaluating the quality of the forecasts for two reasons. First, the historical performance of a forecasting process helps us create the description of future demand to support resource allocation decisions. Second, by tracking the performance of a forecasting process over time, we may be able to identify ways to improve that process. This feedback loop is noted in Figure 3.1.

We present some definitions and notation for use throughout this and the following chapters:

- Demand observed in period t is represented by x_t .
- \hat{x} represents a point forecast for a future period. This value will be indexed by two relevant time periods: the period in which the forecast is made and the period the forecast is for. A forecast made at the end of time period t for a future period $t + \tau$ will then be written as $\hat{x}_{t,t+\tau}$. One would typically refer to this as the “lag- τ ” forecast since it is a forecast for a time period τ periods in the future. Quite often, we will focus on the “lag-1” forecast—the forecast for one period ahead: $\hat{x}_{t-1,t}$, or $\hat{x}_{t,t+1}$.
- Let $e_{t-\tau,t} = x_t - \hat{x}_{t-\tau,t}$ represent the forecast error associated with the forecast made in period $t - \tau$ for period t . Defined in this way, a negative value indicates that the point forecast was *higher* than the observed demand. As mentioned in the previous point, we will often focus on the “lag-1” forecast, and for notational convenience, we may write the lag-1 error $e_{t-1,t}$ as simply e_t .

This chapter is organized as follows. Section 3.1 discusses the plausible components of mathematical models of a demand time series. Then, in Section 3.2, we present the three steps involved in the use of a mathematical model—namely (1) the selection of a general form of the model, (2) the choice of values for the parameters within the model, and, finally, (3) the use of the model for forecasting purposes. Although the emphasis in this chapter, with regard to statistical forecasting, is on individual-item, short-range forecasting (needed for inventory management and production scheduling), we do briefly present material appropriate to medium-range aggregate forecasting (useful for production planning) in Section 3.3. This is followed, in Section 3.4, by a detailed discussion of several different procedures for short-range forecasting, with particular emphasis on what is called as exponential smoothing. Section 3.5 is concerned with various measures of forecast errors, building on our basic forecast error definitions given above. These measures are needed for establishing safety stocks and also for monitoring the forecasting performance. Section 3.6 addresses how to deal with unusual kinds of demand. Then, in Section 3.7, we have considerable discussion concerning how to encourage and use the human judgment input to forecasting. Issues concerning special classes of items are raised in Section 3.8. Finally, in Section 3.9, we return to tactical and strategic issues, specifically to the choice among different forecasting procedures.

3.1 The Components of Time-Series Analysis

Any time series can be thought of as being composed of five components: level (*a*), trend (*b*), seasonal variations (*F*), cyclical movements (*C*), and irregular random fluctuations (ϵ). *Level* captures the scale of a time series. See Figure 3.2. If only a level was present, the series would be constant with time. *Trend* identifies the rate of growth or decline of a series over time. See Figure 3.3 that illustrates a positive linear trend. *Seasonal variations* can be of two kinds: (1) those resulting from natural forces, and (2) those arising from human decisions or customs. For example, in the northern United States and in Canada, the intensity of many types of economic activity depends on

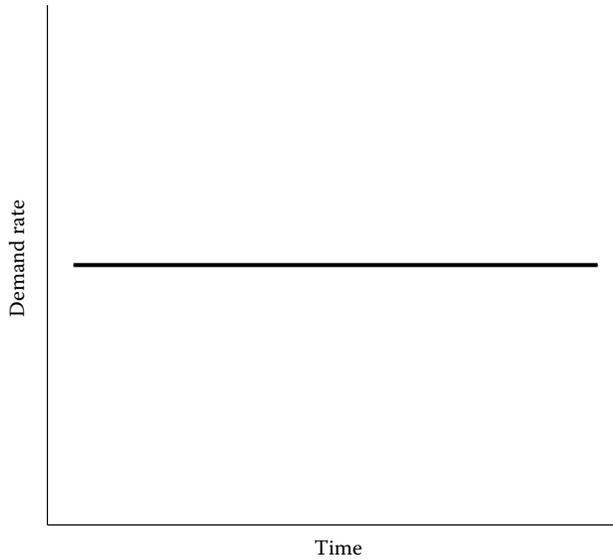


Figure 3.2 Level demand.

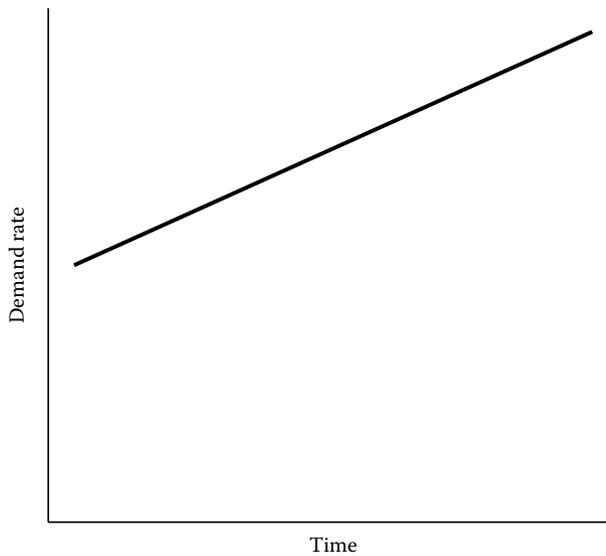


Figure 3.3 Linear trend.

prevailing weather conditions. On the other hand, department store sales increase before the start of the school year, the timing of which is a human decision. Figure 3.4 illustrates an annual seasonal pattern. *Cyclical variations* or alternations between expansion and contraction of economic activity are the result of business cycles. *Irregular fluctuations* in time-series analysis are the residue that remain after the effects of the other four components are identified and removed from the time series. Such fluctuations are the result of unpredictable events such as unusual weather conditions

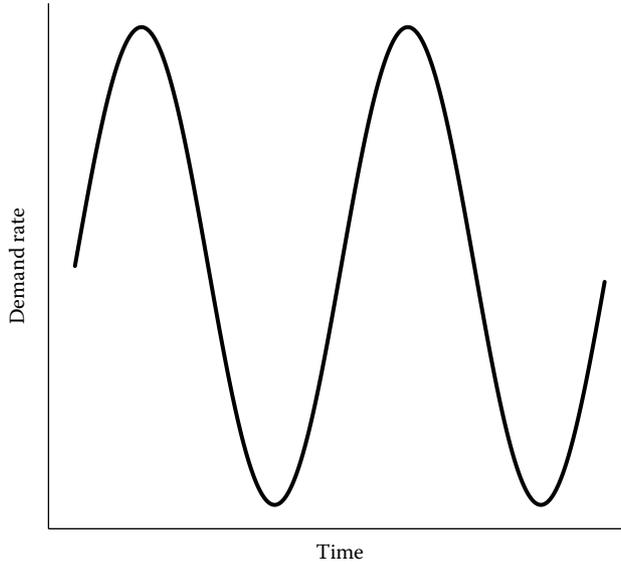


Figure 3.4 Illustrative seasonal pattern.

and unexpected labor strife. They represent our inability to effectively model what else may be affecting the time series we are studying.

Using these concepts, we can formulate a multiplicative model of a time series:

$$\text{Demand} = (\text{Trend})(\text{Seasonal})(\text{Cyclic})(\text{Irregular}) \quad (3.1)$$

$$\text{Demand} = (b)(F)(C)(\epsilon)$$

or an *additive* model:

$$\text{Demand in period } t = (\text{Level}) + (\text{Trend}) + (\text{Seasonal}) + (\text{Cyclic}) + (\text{Irregular}) \quad (3.2)$$

$$x_t = a + bt + F_t + C_t + \epsilon_t$$

or a mixed model (partly additive, partly multiplicative). There are statistical procedures for isolating each component. We comment further on this topic later in the chapter, but our discussion will be restricted to models that, in general, do not incorporate cyclical effects because we concentrate on short- to medium-term forecasts.

3.2 The Three Steps Involved in Statistically Forecasting a Time Series

It is conceptually useful to think in terms of three distinct steps involved in statistically forecasting a time series of demands for an individual SKU (or the aggregated demand for a group of related items). In this section, we provide an overview of these steps, and then in the later sections we

discuss each in greater detail. The steps are

Step 1 Select an appropriate underlying model of the demand pattern through time.

Step 2 Select the values for the parameters inherent in the model.

Step 3 Use the model (selected in Step 1) and the parameter values (chosen in Step 2) to forecast the future demands.

The choice of the general type of underlying model to use for a particular item (or aggregated group of items) depends very much on cost considerations. The more sophisticated and costly models are really only appropriate for (1) medium- or long-term forecasting of an aggregated time series, and (2) perhaps for short-term forecasts of A items. Trend and particularly seasonal components should be introduced with care. It is often the case with B and C items that there is insufficient reliable historical data to adequately estimate the seasonal components. An analysis of historical data (often woefully unavailable in companies embarking on the introduction of formalized forecasting procedures) should suggest the general types of models (e.g., seasonal or not, additive or multiplicative, etc.) that would appear to be appropriate for a given time series. In addition, managerial knowledge of a particular market is invaluable in suggesting the presence of trends or seasonal patterns.

Once a general form of the model is tentatively selected, it is necessary to estimate the values of any parameters of that model. Estimation can be based on some form of statistical fit to historical data—for example, a least-squares criterion will be discussed in Section 3.3.1. Alternatively, particularly when there is limited or no historical data available, the estimates of parameters are made on essentially a subjective basis—for example, saying that a new item in a family of comparable items will have a seasonal pattern similar to that of well-established members of the family.

In the event that historical data exist, the time series generated by the tentatively selected model (and its estimated parameter values) can be compared with the historical pattern. If the fit is judged to be inadequate, then, it is necessary to iterate back through Steps 1 and 2.

Once we are satisfied with the assumed underlying model, then, in Step 3, it is used, with the estimated values of its parameters, to forecast future demands. This procedure, of course, implicitly assumes that the model and its parameter values will not change appreciably over the period being forecasted. If such changes can be predicted, then, appropriate subjective adjustments in parameter values and in the nature of the model itself are called for.

As we described earlier (see Figure 3.1), as forecasts are made and errors are observed, managers can adjust the estimates of the parameter values and judge the suitability of the model.

3.3 Some Aggregate Medium-Range Forecasting Methods

In this section, we provide a *brief* overview of some of the common models and procedures useful for medium- and longer-term forecasting of an aggregate time series. Such forecasts are an important input to the aggregate production-planning procedures to be discussed in Chapter 14. Our intention is simply to discuss the general nature of the available methodology, and to provide references where further details can be found. Moreover, there are several longer-range forecasting techniques that we will not even mention. An excellent starting point for more information on either medium- or long-term forecasting is the article by Makridakis (2008).

In Section 3.1, we showed both a multiplicative and an additive model (Equations 3.1 and 3.2), each giving demand as a function of level, trend, seasonal, and cyclical components. In actual fact, it

may be necessary to first apply a transformation to the demand data in order to achieve a reasonable fit with one of these models. The added complexity of such a transformation is warranted in medium-range forecasting where we are dealing with relatively few (aggregated) time series and forecasts. A class of transformations that has proved to be useful in practice (see Granger and Newbold 2014) is

$$T(x_t) = \ln x_t \quad (3.3)$$

So, $T(x_t)$, as opposed to x_t , would be modeled and forecasted. Then, for any given forecast of $T(x_t)$ the corresponding x_t would be obtained by inverting the transformation, that is, solving Equation 3.3 for x , in terms of T , or,

$$x_t = e^T \quad (3.4)$$

As an example, if $x_t = 20$, $T(x_t)$ would be $\ln(20) = 2.996$. We would use 2.996 in forecasting activities. Then, if a forecast of $T(x_{t+1}) = 3.5$, the actual forecasted demand, x_{t+1} , would be $e^{3.5} = 33.115$.

3.3.1 Regression Procedures

For illustrative purposes, we show only the simple, but common, case of an underlying model that assumes that demand is a linear function of time:

$$x_t = a + bt \quad (3.5)$$

A least-squares criterion is used to estimate values, denoted by \hat{a} and \hat{b} , of the parameters a and b . The resulting modeled value of x_t , denoted \hat{x}_t , is then given by

$$\hat{x}_t = \hat{a} + \hat{b}t + \epsilon_t \quad (3.6)$$

If there are n actual historical observations x_t (for $t = 1, 2, \dots, n$), then, the least-squares criterion involves the selection of \hat{a} and \hat{b} to minimize the sum of the squares (S) of the n differences between x_t and \hat{x}_t . That is,

$$S = \sum_{t=1}^n (x_t - \hat{a} - \hat{b}t)^2 \quad (3.7)$$

Through the use of calculus we can show that

$$\hat{b} = \frac{\sum_{t=1}^n tx_t - \left(\frac{n+1}{2}\right) \sum_{t=1}^n x_t}{n(n^2 - 1)/12} \quad (3.8)$$

$$\hat{a} = \sum_{t=1}^n x_t/n - \hat{b}(n+1)/2 \quad (3.9)$$

Fitting such a model to the aggregate demand data for the item shown in Table 3.1, we obtain

$$\hat{x}_t = 2961 + 48.68t \quad (3.10)$$

Once the parameters a and b are estimated, the model of Equation 3.6 can be used to forecast (extrapolate) the demand in a future period. For example, the use of Equation 3.10 gives an estimate of demand for the item in October 2015 (period 22) as being 4,032 packages. There are also

Table 3.1 Demand Data for a Sample Item

Year	Month	Period, t	x_t (Packages)
2013	January	1	3,025
	February	2	3,047
	March	3	3,079
	April	4	3,136
	May	5	3,268
	June	6	3,242
	July	7	3,285
	August	8	3,334
	September	9	3,407
	October	10	3,410
	November	11	3,499
	December	12	3,598
2014	January	13	3,596
	February	14	3,721
	March	15	3,745
	April	16	3,650
	May	17	3,746
	June	18	3,775
	July	19	3,906
	August	20	3,973

statistical procedures for estimating the adequacy of the straight-line fit and for developing a confidence interval around the forecast (extrapolated) value for any future period. Details are available in several sources such as Berger (2013).

Our discussion to this point has assumed that the independent variable used in the regression is the time period t . More generally, multiple regression includes two or more exogenous variables. Moreover, in autoregression, the demand in period t is postulated as a function of the demand itself in earlier periods. As an illustration of the latter, we might have

$$x_t = a_1 + a_2x_{t-1} + \epsilon_t$$

which postulates that the demand in period t is a linear function of the demand in period $t - 1$ plus an error term. A generalization to a system of autoregressive equations in several variables to be forecast, as well as some exogenous variables, leads one into the field of econometric forecasting.

This includes the use of the so-called leading indicators. A variable y is said to be a leading indicator of another variable x if changes in x are anticipated by changes in y . Multiple regression is sometimes referred to as forecasting by association because the variable of interest is forecasted indirectly by its association with other variables. See, for example, Maddala and Lahiri (1992).

3.3.1.1 The Box–Jenkins Approach

Box and Jenkins (see Box et al. 2011) have had a profound effect on the field of statistical forecasting. They suggest a broad class of underlying statistical models of demand patterns, as well as a procedure for selecting an appropriate member of the class based on the historical data available. Conceptually the models are more complex than either the simple regression models (outlined in Section 3.3.1) or the exponential smoothing procedures (to be covered in Section 3.4). However, this complexity (and its associated additional costs) affords an opportunity for more accurate forecasting, which is certainly of interest at the medium-term, aggregate level. In this section, we briefly introduce some aspects of the Box–Jenkins approach.

The so-called autoregressive-moving average (or ARMA) class of models represents the demand in the current period, x_t , by a weighted sum of past demands and unpredictable random components. In mathematical terms:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} \quad (3.11)$$

where the ϵ 's are the so-called white noise, namely, independent, normally distributed variables with mean 0 and a constant variance, and the ϕ 's, θ 's, p , and q are constants.

Typically, Equation 3.11 can be simplified. Anderson (1976, p. 45) reports that many stationary time series are adequately represented by ARMA models with $p+q = 2$, that is, at most three terms on the right side of Equation 3.11. As an illustration, the following shows the case of $p = 1$, $q = 1$:

$$x_t = \phi_1 x_{t-1} + \epsilon_t + \theta_1 \epsilon_{t-1} \quad (3.12)$$

Equation 3.11 in fact implies that the x_t series is stationary with time (i.e., its expected value and standard deviation do not change with time). At least two circumstances can invalidate this assumption. First, there may be a trend (linear or higher order) in the underlying process. Second, there may be seasonality. The Box–Jenkins approach can handle both of these circumstances by introducing additional terms into Equation 3.11.

In Section 3.2, we discussed three steps involved in statistically forecasting a time series: selection of a model, estimation of the parameters of the model, and use of the model for forecasting purposes. The Box–Jenkins approach involves considerable elaboration on the first two of these steps, because there is such a rich class of models from which to choose. See Box et al. (2011) for further details.

3.4 Individual-Item, Short-Term Forecasting: Models and Procedures

In this major section of the chapter, we concentrate on procedures for estimating and using the model parameters. These procedures have been found to be reasonably accurate in terms of short-term forecasts for individual SKUs. In addition, they are relatively simple and inexpensive to use,

in contrast with the methods of the previous section, and thus are practical to use for the high volume of individual-item, short-term forecasts.

We restrict attention to procedures that are appropriate for at least one of the following assumed underlying models of x_t , the demand in period t :

$$x_t = a + \epsilon_t \text{ (Level model)} \quad (3.13)$$

$$x_t = a + bt + \epsilon_t \text{ (Trend model)} \quad (3.14)$$

$$x_t = (a + bt)F_t + \epsilon_t \text{ (Multiplicative trend-seasonal model)} \quad (3.15)$$

where, as earlier,

a = a level

b = a (linear) trend

F_t = a seasonal coefficient (index) appropriate for period t and the ϵ_t 's are independent random variables with mean 0 and constant variance σ^2

Of course, we never know the exact values of the parameters in Equations 3.13 through 3.15. To make matters worse, the true values are likely to change with the passage of time. For example, the arrival of a new competitor in the market is likely to reduce the level or trend by reducing the current market share. Consequently, the procedures we discuss will be concerned with estimating the current values of the parameters of the specific model under consideration. There are two distinct aspects to estimation. First, there is the question of obtaining initial estimates prior to using the model for forecasting purposes. Second, there is the updating of the most recent estimates as we observe the demand (and hence the forecast error) in the latest period. For each procedure, we discuss these two different dimensions of estimation.

3.4.1 *The Simple Moving Average*

3.4.1.1 *Underlying Demand Model*

The simple moving average method is appropriate when demand is modeled as in Equation 3.13:

$$x_t = a + \epsilon_t \quad (3.16)$$

that is, a level with random noise. As mentioned above, the parameter a is not really known and is subject to random changes from time to time.

3.4.1.2 *Updating Procedure*

The simple N -period moving average, as of the end of period t , is given by

$$\bar{x}_{t,N} = (x_t + x_{t-1} + x_{t-2} + \cdots + x_{t-N+1})/N \quad (3.17)$$

where the x 's are actual, observed demands in the corresponding periods. The estimate of a as of the end of period t is then

$$\hat{a}_t = \bar{x}_{t,N} \quad (3.18)$$

We can show that this estimate of a results from minimizing the sum of squares of errors over the preceding N periods (see Problem 3.3). A slightly simpler version (for updating purposes) of the

N -period moving average is easily developed from Equation 3.17 as

$$\bar{x}_{t,N} = \bar{x}_{t-1,N} + (x_t - x_{t-N})/N \quad (3.19)$$

Note from Equation 3.17 that the moving average is simply the mean of the N most recent observations. Provided that the model of Equation 3.16 holds, we know from basic probability that $\hat{x}_{t,N}$ or \hat{a}_t is distributed with mean a and standard deviation σ_ϵ/\sqrt{N} . The larger the N is, the more precise will be the estimate, provided Equation 3.16 continues to apply over the entire set of periods. However, if there is a change in the parameter a , a small value of N is preferable because it will give more weight to recent data, and thus will pick up the change more quickly. Typical N values that are used range from 3 to as high as 12.

3.4.1.3 Initialization

In terms of initialization, an N -period moving average should not really be used until there are N periods of history available. In actual fact, if early forecasts are essential, one can use a smaller number of periods in the initial stages.

3.4.1.4 Forecasting

Because of the (level) nature of the model of Equation 3.16, it is appropriate to use our estimate \hat{a}_t as a forecast at the end of period t for any future period:

$$\hat{x}_{t,t+\tau} = \hat{a}_t \quad (3.20)$$

where $\hat{x}_{t,t+\tau}$ is the forecast, made at the end of period t , of demand in period $t + \tau$ (for $\tau = 1, 2, 3, \dots$). This procedure can be developed on a spreadsheet quite easily, as is illustrated next. See also Problem 3.14.

3.4.1.5 Illustration

For a product, PSF-008, columns 1–3 of Table 3.2 show the demand data during 2013. The 5-month moving averages, beginning as of the end of May 2013, are shown in column 4 and the associated 1-month ahead forecasts in column 6. We comment further on the results of Table 3.2 in the next section when we discuss the topic of simple exponential smoothing.

3.4.1.6 General Comments

The moving average approach requires the storage of N periods of historical data that must be stored for each SKU. In addition, the weight given to each of these N data points is equal while all older data points are disregarded. Intuitively, we may want to weight more recent observations more heavily, but it seems inappropriate to completely discount the earlier data. It is possible to modify the moving average approach with different weights. The exponential smoothing approach in the next section does precisely this in a compact way that is easy to implement.

Table 3.2 Moving Average and Simple Exponential Smoothing Forecasts for Product PSF-008, Using $N = 5$ and $\alpha = 1/3$ (All Data in 100 Square Meters)

Year (1)	Month t (2)	Demand x_t (3)	5-Month Moving Average $\bar{x}_{t,5}$ (4)	Exponentially Smoothed Average (5)	Forecasts	
					$\hat{x}_{t,t+1} = \bar{x}_{t,5}$ (6)	$\hat{x}_{t,t+1} = \hat{a}_t$ (7)
2013	1	52				
	2	48				
	3	36		(50.0)		
	4	49		(49.7)		
	5	65	50.0	(54.8)	50.0	54.8
	6	54	50.4	54.5	50.4	54.5
	7	60	52.8	56.3	52.8	56.3
	8	48	55.2	53.6	55.2	53.6
	9	51	55.6	52.7	55.6	52.7
	10	62	55.0	55.8	55.0	55.8
	11	66	57.4	59.2	57.4	59.2
	12	62	57.8	60.1	57.8	60.1
2014	13					

3.4.2 Simple Exponential Smoothing

Simple exponential smoothing is probably the most widely used statistical method for short-term forecasting. As we shall see, it is intuitively appealing and very simple to employ. An excellent reference on this topic is Gardner (2006).

3.4.2.1 Underlying Demand Model

As in the case of moving averages, the basic underlying demand pattern assumed is that of Equation 3.13, which, for convenience, we repeat here:

$$x_t = a + \epsilon_t \quad (3.21)$$

However, later, we will mention that the procedure is actually appropriate for a somewhat wider range of demand models.

3.4.2.2 Updating Procedure

We wish to estimate the parameter a in Equation 3.21. Recall that in Section 3.3.1, we used a least-squares regression on historical data to estimate parameters of the underlying models. However, we

pointed out that regular least squares gives equal weight to all historical data considered. Simple exponential smoothing can be derived by selecting \hat{a}_t , the estimate of a at the end of period t , in order to minimize the following sum of discounted squares of residuals:

$$S' = \sum_{j=0}^{\infty} d^{j+1} (x_{t-j} - \hat{a}_t)^2 \quad (3.22)$$

where

d = a discount factor ($0 < d < 1$)

x_{t-j} = the actual demand in period $t - j$

Note that, because $d < 1$, the weighting given to historical data decreases geometrically as we go back in time. Exponential smoothing would really be more appropriately named geometric smoothing (an exponential curve is a continuous approximation to geometrically decaying points). The minimization of Equation 3.22 is carried out in the Appendix to this chapter. The resulting estimate of a satisfies the following updating formula:

$$\hat{a}_t = \alpha x_t + (1 - \alpha)\hat{a}_{t-1} \quad (3.23)$$

where $\alpha = 1 - d$ is known as the smoothing constant.

In addition, it can be shown (see, e.g., Johnson and Montgomery 1974) that \hat{a}_t is an unbiased estimate of a ; that is, $E\hat{a}_t = a$.

Equation 3.23 can be manipulated as follows:

$$\hat{a}_t = \hat{a}_{t-1} + \alpha(x_t - \hat{a}_{t-1})$$

or

$$\hat{a}_t = \hat{a}_{t-1} + \alpha e_t \quad (3.24)$$

where

e_t = the error in period t between the actual demand x_t and the forecast \hat{a}_{t-1} made at the end of the previous period $t - 1$

There is an intuitive appeal to Equation 3.24. We adjust the forecast for period t in the direction of the previous error, with α throttling the strength of this adjustment. If the forecast \hat{a}_{t-1} from the previous period ($t - 1$) was too low (the error is positive), it makes sense that we would want to increase the next forecast, \hat{a}_t , which we do by adding αe_t to the previous forecast.

Note again that these equations can be developed quite easily on a spreadsheet. (See Problem 3.15.) Equation 3.24 states that the new estimate of the level is the old estimate plus a fraction of the most recent error. Practically speaking, we don't ever really know whether any particular error results from a random fluctuation or whether a significant shift in a has occurred. As a result, the simple exponential smoothing model hedges by assuming that only a fraction of the forecast error should be used to revise the estimate of a . Note from Equation 3.23 that as α tends toward 0, our estimate is unchanged by the latest piece of information, while as α moves toward 1, more and more weight is placed on the demand observed in the last period. In Section 3.4.5, we comment on the selection of an appropriate value of the smoothing constant.

The updating procedure of Equation 3.23 can also be shown to minimize the mean-squared error (MSE) of the one-period-ahead forecast for the following models that are both somewhat more general than that of Equation 3.21:

Case 1 (Details can be found in Harrison 1967)

$$x_t = a_t + \epsilon_t$$

with

$$a_t = a_{t-1} + \delta_t$$

where the δ_t 's are random changes with mean 0 and independent of the ϵ 's. That is, the level is subject to random changes through time (which is what we have implicitly assumed earlier).

Case 2 The x_t 's are generated by a modification of the autoregressive process discussed in Section 3.3.1. Specifically,

$$x_t = x_{t-1} + \epsilon_t = \theta_1 \epsilon_{t-1}$$

See Goodman (1974) for details.

The geometric weighting of historical data in exponential smoothing can also be seen by substituting a corresponding relation for \hat{a}_{t-1} in Equation 3.23, and then repeating for \hat{a}_{t-2} , etc. This gives

$$\begin{aligned} \hat{a}_t &= \alpha x_t + (1 - \alpha)[\alpha x_{t-1} + (1 - \alpha)\hat{a}_{t-2}] \\ &= \alpha x_t + \alpha(1 - \alpha)x_{t-1} + (1 - \alpha)^2[\alpha x_{t-2} + (1 - \alpha)\hat{a}_{t-3}] \\ &= \alpha x_t + \alpha(1 - \alpha)x_{t-1} + \alpha(1 - \alpha)^2 x_{t-2} + \dots \\ \hat{a}_t &= \sum_{j=0}^{\infty} \alpha(1 - \alpha)^j x_{t-j} \end{aligned} \quad (3.25)$$

We see the contrast with moving averages where the last N observations were given equal weight. In addition, the recursion relationship of Equation 3.23 shows that only one value of a and the most recent x_t need to be retained. From the weighting of the terms in Equation 3.25, we can ascertain (see Problem 3.4) that the average age of the data encompassed in \hat{a}_t is $1/\alpha$ periods. In an N -month moving average, the average age of the data used is $(N + 1)/2$. Equalizing the average ages gives the following relationship between α and N :

$$\alpha = 2/(N + 1) \quad (3.26)$$

3.4.2.3 Initialization

Where significant historical data exist, one simply uses the average demand in the first several periods as the initial estimate of a . The numerical illustration will demonstrate this point. The case where inadequate history exists will be addressed in Section 3.8.1.

3.4.2.4 Forecasting

As mentioned earlier, \hat{a}_t is an unbiased estimate, as of the end of period t , of the parameter a , where the latter, according to Equation 3.21, is the expected demand in any future period. Thus, our forecast, made at the end of period t , for any future period $t + \tau$ is

$$\hat{x}_{t,t+\tau} = \hat{a}_t \quad (3.27)$$

For deciding when to place a replenishment order and how much to order, we will be interested in the total forecast over several unit time periods (e.g., the replenishment lead time may be 4 months, while the forecast update period is 1 month). If we let $\hat{x}_{t,t,u}$ be the forecast, made at the end of period t , of total demand from time t (end of period t) to time u , then from Equation 3.27 we have

$$\hat{x}_{t,t,u} = (u - t)\hat{a}_t \quad (3.28)$$

3.4.2.5 Illustration

We use the same example of product PSF-008 as we used earlier to illustrate a 5-period moving average. In addition, for comparison, we use Equation 3.26 to select the smoothing constant that gives the same average age of the data; that is,

$$\alpha = 2/(5 + 1) = 1/3$$

(In actual fact, as will be discussed later, one would normally not use such a high value of α .)

In Table 3.2, the average of the first 5 months' sales (50.0) is taken as the initial value for the exponentially smoothed average and plotted opposite $t = 3$, the midpoint of the 5-month range of the average. Then, the exponentially smoothed average in column 5 for month 4 is calculated by using Equation 3.23:

$$\begin{aligned} \hat{a}_t &= \frac{1}{3}x_4 + \frac{2}{3}\hat{a}_3 \\ &= \frac{1}{3}49 + \frac{2}{3}50 = 49.7 \end{aligned}$$

The same procedure was repeated for all subsequent months. For months 4 and 5, no exponentially smoothed forecasts were issued to allow the effect of setting initial conditions at $\hat{a}_3 = 50.0$ to be smoothed out. This is commonly referred to as the “run-in period.” Rather than using $\hat{a}_3 = 50.0$, we could instead set $\hat{a}_0 = 50.0$, and thus make the run-in time equal to 5 periods.

Using Equation 3.27, “exponential smoothing” forecasts were calculated for each period from period 6 onward. These are shown in column 7 of Table 3.2. Note from columns 6 and 7 that the exponential smoothing forecast outperforms the 5-month moving average forecast for 5 of the 7 months of 2013. This appears to be a result of the fact that by using $\alpha = 1/3$, the simple exponential smoothing forecast is made more responsive to recent sales data variation than the 5-month moving average. Since the time series x_t appears to be undergoing changes, this degree of responsiveness appears to be warranted for the particular example shown.

3.4.2.6 General Comments

In summary, simple exponential smoothing is an easy to understand, effective procedure to use when the underlying demand model is composed of level and random components. Even when the underlying demand process is more complicated, exponential smoothing can be used as part of an updating procedure. For example, Steece and Wood (1979) use a Box–Jenkins model for the aggregate demand time series of an entire class of items, and then use simple exponential smoothing to update the estimate of a particular item's demand as a fraction of the aggregate.

3.4.3 Exponential Smoothing for a Trend Model

The simple smoothing procedure presented in the previous section is based on a model without a trend and therefore is inappropriate when the underlying demand pattern involves a significant trend. A somewhat more complicated smoothing procedure is needed under such circumstances.

3.4.3.1 Underlying Demand Model

The basic underlying model of demand is the pattern of Equation 3.14 that we repeat here.

$$x_t = a + bt + \epsilon_t \quad (3.29)$$

Again, we will make a reference later to somewhat more general models for which the suggested smoothing procedure is also appropriate.

3.4.3.2 Updating Procedure

Holt (2004) suggests a procedure that is a natural extension of simple exponential smoothing

$$\hat{a}_t = \alpha_{HW}x_t + (1 - \alpha_{HW})(\hat{a}_{t-1} + \hat{b}_{t-1}) \quad (3.30)$$

$$\hat{b}_t = \beta_{HW}(\hat{a}_t - \hat{a}_{t-1}) + (1 - \beta_{HW})\hat{b}_{t-1} \quad (3.31)$$

where α_{HW} and β_{HW} are smoothing constants and, for the demand model of Equation 3.29, the difference $\hat{a}_t - \hat{a}_{t-1}$ is an estimate of the actual trend in period t . (We use the subscripts HW to represent Holt–Winters since these smoothing constants will be part of a procedure of Winters, to be discussed in Section 3.4.4.)

Harrison (1967), among others, shows that the Holt procedure minimizes the expected one-period-ahead, mean-square forecast error for the trend model of Equation 3.29—in fact, for a more general model where independent random changes in both a and b are possible for each period, in addition to the random noise (the ϵ 's).

We use Holt's updating as part of a popular procedure, to be described in Section 3.4.4, for dealing with an underlying model involving both trend and seasonal components. However, in this section, where there is no seasonality, we recommend instead a single-parameter procedure suggested by Brown (1963), which turns out to be a special case of Holt's method. Besides involving only a single smoothing parameter, Brown's procedure again has the intuitively appealing property of being derived from minimizing the sum of geometrically weighted forecast errors (similar to

the basis for the simple exponential smoothing procedure of Equation 3.23). Brown's updating equations are (see Problem 3.6)

$$\hat{a}_t = [1 - (1 - \alpha)^2]x_t + (1 - \alpha)^2(\hat{a}_{t-1} + \hat{b}_{t-1}) \quad (3.32)$$

$$\hat{b}_t = \left[\frac{\alpha^2}{1 - (1 - \alpha)^2} \right] (\hat{a}_t - \hat{a}_{t-1}) + \left[1 - \frac{\alpha^2}{1 - (1 - \alpha)^2} \right] \hat{b}_{t-1} \quad (3.33)$$

where α is the single smoothing constant ($\alpha = 1 - d$ again, where d is the discount factor used in the geometrically weighted sum of squares).

Note that Equations 3.32 and 3.33 are the special case of Holt's updating equations (Equations 3.30 and 3.31), where

$$\alpha_{HW} = [1 - (1 - \alpha)^2] \quad (3.34)$$

$$\beta_{HW} = \frac{\alpha^2}{1 - (1 - \alpha)^2} \quad (3.35)$$

In Section 3.4.5, we address the issue of the selection of the α value.

3.4.3.3 Initialization

Typically, the initial values of the level (a) and the trend (b) are ascertained by doing a regular (unweighted) least-squares regression on the historical data available. Suppose that there are n periods of data available. Smoothing will begin with the next period after this preliminary data. Thus, we would like to have the level a computed relative to an origin at the end of the last period of the fitted data. That is, we should number the historical periods $0, -1, -2, -3, \dots, -(n-1)$ going back in time. In other words, we wish to select a_0 and b_0 to minimize

$$S = \sum_{t=-n+1}^0 [x_t - (\hat{a}_0 + \hat{b}_0 t)]^2$$

Through the use of calculus, we obtain

$$\hat{a}_0 = \frac{6}{n(n+1)} \sum_t tx_t + \frac{2(2n-1)}{n(n+1)} \sum_t x_t \quad (3.36)$$

and

$$\hat{b}_0 = \frac{12}{n(n^2-1)} \sum_t tx_t + \frac{6}{n(n+1)} \sum_t x_t \quad (3.37)$$

where all the summations range over the integers $-(n-1), -(n-2), \dots, -2, -1, 0$.

Where this initial regression is a relatively straightforward matter, the question arises as to why we don't simply redo a regression after each additional period of information instead of resorting to the smoothing procedure of Equations 3.32 and 3.33. There are two reasons. First, the regression gives equal weight to all historical information, whereas the smoothing reduces the weight geometrically as we go back in time. Second, and less important, smoothing is less computationally intensive, and therefore is more appropriate on a repetitive basis for short-term forecasting. As we

shall see, smoothing can be done by adding a single line to a spreadsheet for each new demand data point.

The case where there is insufficient history for regression initialization will be treated in Section 3.8.1.

3.4.3.4 Forecasting

Because of the underlying model of Equation 3.29, we know that

$$\hat{x}_{t,t+\tau} = \hat{a}_t + \hat{b}_t\tau \tag{3.38}$$

where

$\hat{x}_{t,t+\tau}$ = the forecast, made at the end of period t , of the demand in period $t + \tau$

Using Equation 3.38, we can rewrite Equations 3.30 and 3.31 in a manner similar to Equation 3.24 (for simple exponential smoothing) to highlight the intuition behind the adjustments to the level and trend estimates \hat{a}_t, \hat{b}_t .

$$\begin{aligned} \hat{a}_t &= (\hat{a}_{t-1} + \hat{b}_{t-1}) + \alpha_{HW}[x_t - (\hat{a}_{t-1} + \hat{b}_{t-1})] \\ &= \hat{x}_{t-1,t} + \alpha_{HW}e_t \end{aligned}$$

That is, the new estimate for the level \hat{a}_t is the old estimate for the level *at time* t plus an adjustment based on the error, moderated by the smoothing constant. The trend-updating equation can be rewritten for a similar intuition.

$$\hat{b}_t = \hat{b}_{t-1} + \beta_{HW}[(\hat{a}_t - \hat{a}_{t-1}) - \hat{b}_{t-1}]$$

Here, the “error” used to adjust the new trend estimate is not the forecast error but the difference between the old trend estimate and the recently estimated value for the trend obtained by taking the difference between level estimates in periods $t - 1$ and t .

To make a forecast for *cumulative* demand over a horizon, say from time t (end of period t) to time u (end of period u), is

$$\hat{x}_{t,t,u} = \sum_{\tau=1}^{u-t} (\hat{a}_t + \hat{b}_t\tau) \tag{3.39}$$

When u is not exactly an integer, the last term in the summation should cover an appropriate fraction of the last period.

3.4.3.5 Illustration

Consider another product PSF-016 where it is reasonable to assume that demand is growing, approximately linearly, with time. Table 3.3 shows demand data for 15 consecutive months. Suppose that the first 6 months of data (January 2013 through June 2013) are used to estimate \hat{a}_0 and \hat{b}_0 . The use of Equations 3.36 and 3.37 gives

$$\hat{a}_0 = 27.19 \quad \text{and} \quad \hat{b}_0 = 1.34$$

Suppose we do not use a run-in period and directly use these to forecast demand in July 2013 ($t = 1$), employing Equation 3.38. This gives us $\hat{x}_{0,1} = 28.53$. The resulting error in July 2013

Table 3.3 Illustration of Forecasting with a Trend Model for PSF-016 (All Data in 100 Square Meters)

Year	Month	t	x_t	$e_t = x_t - \hat{x}_{t-1,t}$	\hat{a}_t	\hat{b}_t	$\hat{x}_{t,t+1} = \hat{a}_t + \hat{b}_t$
2013	January	-5	20				
	February	-4	25				
	March	-3	21				
	April	-2	22				
	May	-1	27				
	June	0	28		27.19	1.34	28.53
	July	1	27	-1.53	27.68	1.17	28.85
	August	2	30	1.15	29.49	1.30	30.79
	September	3	34	3.21	32.57	1.66	34.23
	October	4	25	-9.23	29.10	0.63	29.73
	November	5	25	-4.73	27.10	0.10	27.21
	December	6	26	-1.21	26.54	-0.03	26.51
2014	January	7	36	9.49	31.78	1.03	32.81
	February	8	41	8.19	37.36	1.94	39.29
	March	9	39	-0.29	39.13	1.90	41.03

is $27 - 28.53$ or -1.53 . The parameters a and b are updated through the use of Equations 3.32 and 3.33 and a forecast $\hat{x}_{1,2}$ is made at the end of July 2013, and so on. Continuing in this fashion, we obtain the results in the last four columns of Table 3.3. (Note that in this, and all other tables in this chapter, we have used a spreadsheet to compute the values shown. For presentation purposes, we report numbers rounded to the nearest hundredth or so. Because the spreadsheet retains many more digits than those shown, the results computed from the rounded numbers may differ from the results reported in the tables.)

3.4.3.6 A Model with Damped Trend

Sometimes, the data are so noisy, or the trend is so erratic, that a linear trend is not very accurate, especially when forecasting several periods ahead. Gardner and McKenzie (1985) introduce a *damped trend* procedure that works particularly well in these situations. The method follows Equations 3.30 and 3.31 closely, and is quite easy to use. The updating and forecasting formulas are

$$\hat{a}_t = \alpha_{HW}x_t + (1 - \alpha_{HW})(\hat{a}_{t-1} + \phi\hat{b}_{t-1}) \quad (3.40)$$

$$\hat{b}_t = \beta_{HW}(\hat{a}_t - \hat{a}_{t-1}) + (1 - \beta_{HW})\phi\hat{b}_{t-1} \quad (3.41)$$

$$\hat{x}_{t,t+\tau} = \hat{a}_t + \sum_{i=1}^{\tau} \phi^i \hat{b}_t \quad (3.42)$$

where ϕ is a dampening parameter. Note the similarity of these equations with Equations 3.30, 3.31, and 3.38. In addition to fitting a and b , we must also search for the best ϕ using historical data. If $0 < \phi < 1$, the trend is damped and as τ gets large, the forecast approaches an asymptote that is the horizontal line

$$\hat{a}_t + \hat{b}_t\phi/(1 - \phi)$$

If $\phi = 1$, the trend is linear, and therefore is identical with the model presented in this section. If $\phi > 1$, the trend is exponential, which Gardner and McKenzie (1985) note is dangerous in an automated forecasting system. The authors show ways to reduce the number of parameters, analogous to Brown's procedure presented in Equations 3.32 and 3.33. They also test the model and find that it performs better than the simple linear trend models in most cases, particularly as the forecast horizon grows. The related references include Lewandowski (1982), Gardner and McKenzie (1985, 1989), and Gardner (1988, 1990).

3.4.3.7 Numerical Illustration

Using the data from Table 3.3, we can compute the forecasts using the damped trend procedure. Using Equation 3.34, $\alpha_{HW} = 0.556$, because $\alpha = 1/3$, and using Equation 3.35, $\beta_{HW} = 0.2$. Let us assume that a simple search for the best ϕ reveals that the smallest sum of squared errors is obtained when $\phi = 0.4$. Also, to be consistent with the previous example, let $\hat{a}_0 = 27.19$ and $\hat{b}_0 = 1.34$. Then, from Equation 3.42, the forecast for July is $\hat{x}_{0,1} = 27.19 + 0.4(1.34) = 27.73$. The resulting error in July 2013 is $27 - 27.73$ or -0.73 . The updated parameters a and b are found from Equations 3.40 and 3.41, and are 27.32 and 0.46, respectively. Then, a forecast $\hat{x}_{1,2}$ is made at the end of July 2013, and so on. Continuing in this fashion, we obtain the results in the last four columns of Table 3.4. The reader can verify that the sum of squared errors for this example is somewhat smaller using the damped trend procedure than the linear trend approach.

Now consider a longer-term forecast. Specifically, let us forecast the demand for October 2013 (period 4) at the end of June 2013 (period 0). Using the linear trend model, and Equation 3.38, we find the forecast to be

$$\hat{x}_{t,t+\tau} = \hat{a}_t + \hat{b}_t\tau = 27.19 + 1.34(4) = 32.56$$

The damped trend model, using Equation 3.42 gives

$$\hat{x}_{t,t+\tau} = 27.19 + \sum_{i=1}^4 (0.4)^i(1.34) = 28.06$$

Because the actual demand for October is 25, the damped trend model is clearly more accurate for this example. Note, however, that the linear trend model is more accurate when predicting September's demand (in June). As will be discussed in Section 3.4.5, we recommend some form of search to choose an appropriate value of the dampening parameter. If it turns out to be 1, we have simply replicated the linear trend model.

3.4.4 Winters Exponential Smoothing Procedure for a Seasonal Model

In many organizations, there are a number of individual SKUs that exhibit demand patterns with significant seasonality. Examples include lawnmowers, skiing equipment, fertilizer, heating fuel,

Table 3.4 Illustration of Forecasting with a Trend Model Using Damped Trend

Year	Month	t	x_t	$e_t = x_t - \hat{x}_{t-1,t}$	\hat{a}_t	\hat{b}_t	$\hat{x}_{t,t+1} = \hat{a}_t + \hat{b}_t$
2013	January	-5	20				
	February	-4	25				
	March	-3	21				
	April	-2	22				
	May	-1	27				
	June	0	28		27.19	1.34	27.73
	July	1	27	-0.73	27.32	0.46	27.51
	August	2	30	2.49	28.89	0.46	29.08
	September	3	34	4.92	31.81	0.73	32.10
	October	4	25	-7.10	28.16	-0.50	27.96
	November	5	25	-2.96	26.31	-0.53	26.10
	December	6	26	-0.10	26.05	-0.22	25.96
2014	January	7	36	10.04	31.54	1.03	31.95
	February	8	41	9.05	36.98	1.42	37.54
	March	9	39	1.46	38.35	0.73	38.64

and ice cream. For individual items, the most common forecasting update period when seasonality is present is 1 month. However, for illustration, we use a quarterly (trimonthly) update period. Incidentally, the procedure to be discussed is potentially useful in aggregate, medium-range forecasting, in which case quarterly periods are more meaningful.

3.4.4.1 Underlying Model

The underlying model assumed is that of Equation 3.15

$$x_t = (a + bt)F_t + \epsilon_t \quad (3.43)$$

where

a = the level

b = the linear trend

F_t = a seasonal index (coefficient) appropriate for period t

ϵ_t = independent random variables with mean 0 and constant variance σ^2

The season is assumed to be of length P periods and the seasonal indices are normalized so that, at any point, the sum of the indices over a full season is exactly equal to P . With an annual season, $P = 12$ for a 1-month forecast interval, and $P = 4$ for a quarterly interval. Seasonal models are also applicable for a weekly “season” (e.g., demand for medical services at the emergency facilities

of a hospital), or even a daily “season” (e.g., customer demand for electrical energy as a function of the time of the day). The special case when all the F 's equal unity represents a nonseasonal trend model.

3.4.4.2 Updating Procedure

To use the model of Equation 3.43 repetitively for forecasting purposes, we must have estimates of a , b , and the F 's. Winters (1960) first suggested the following procedure, which is not optimal in the sense, for example, of minimizing mean-square forecast errors. However, it is intuitively appealing and is a natural extension of the Holt procedure for a trend model (see Equations 3.30 and 3.31). Specifically, the parameters are updated according to the following three equations:

$$\hat{a}_t = \alpha_{HW}(x_t/\hat{F}_{t-P}) + (1 - \alpha_{HW})(\hat{a}_{t-1} + \hat{b}_{t-1}) \quad (3.44)$$

$$\hat{b}_t = \beta_{HW}(\hat{a}_t - \hat{a}_{t-1}) + (1 - \beta_{HW})\hat{b}_{t-1} \quad (3.45)$$

$$\hat{F}_t = \gamma_{HW}(x_t/\hat{a}_t) + (1 - \gamma_{HW})\hat{F}_{t-P} \quad (3.46)$$

where α_{HW} , β_{HW} , and γ_{HW} are the three smoothing constants that all lie between 0 and 1. The selection of their values will be covered in Section 3.4.5.

In Equation 3.44, the term x_t/\hat{F}_{t-P} reflects an estimate of the *deseasonalized* actual demand in period t in that \hat{F}_{t-P} was the estimate of the seasonal index for the most recent (P periods earlier) equivalent period in the seasonal cycle. The rest of the terms in Equations 3.44 and 3.45 are equivalent to those shown earlier in the nonseasonal Holt procedure (see Equations 3.30 and 3.31). In Equation 3.46, the term provides an estimate of the seasonal factor based on the latest demand observation. Thus, Equation 3.46, consistent with all the earlier exponential smoothing equations, reflects a linear combination of the historical estimate and the estimate based on the latest piece of data. At all times, we wish to have the sum of the indices through an entire season added to P . Thus, when a specific index is updated, we renormalize all indices to achieve this equality. A numerical illustration will be shown after a discussion of the initialization and forecasting procedures.

As with simple exponential smoothing and exponential smoothing with a trend, we can rewrite the updating equations to show how the old estimate is updated.

$$\begin{aligned} \hat{a}_t &= (\hat{a}_{t-1} + \hat{b}_{t-1}) + \alpha e_t/F_{t-P} \\ \hat{b}_t &= \hat{b}_{t-1} + \beta_{HW}[(\hat{a}_t - \hat{a}_{t-1}) - \hat{b}_{t-1}] \\ \hat{F}_t &= F_{t-P} + \gamma_{HW}(x_t/\hat{a}_t - F_{t-P}) \end{aligned}$$

The level \hat{a} is updated in a similar way to exponential smoothing with a trend where the only difference is that the error used is adjusted by the seasonal index F . Since the estimates a and b are already seasonally adjusted, the updating equation for the trend estimate \hat{b} is the same as in exponential smoothing with a trend. The new seasonal index is calculated by updating the index from one season ago (period $t - P$) and adjusting based on the difference between the most recent observation of the seasonal effect x_t/\hat{a}_t and the previous seasonal index.

3.4.4.3 Initialization

With both trend and seasonal factors present, the initialization becomes considerably more complicated. One has to properly separate out the trend and seasonal effects in the historical data. Several different initialization procedures have been suggested, but we restrict our attention to the most commonly used method, the so-called *ratio to moving average procedure* (which differs from Winters' original suggestion). This approach can effectively handle changes in the underlying trend during the historical period. In addition, it tends to eliminate cyclical effects (which we are not explicitly considering). Further details are available in Hamburg and Young (1994).

Ideally, from a statistical standpoint, it is desirable to have several seasons worth of data because each specific seasonal period (e.g., the month of June) occurs only once per season (year). This is especially true for slow-moving items, or for items with highly erratic demand, because the noise in the data can obscure the underlying seasonality. However, when using too much history, one runs the risk of the seasonal pattern having changed during the history, so that the early portion is no longer representative of the current, let alone, future conditions. As a compromise, we suggest a minimum of 4 complete seasons, using 5 or 6 (with care) if that much is available. The case where insufficient data exist will be treated in Section 3.8.1.

There are several steps in the procedure. We use the example, shown in Table 3.5 and Figure 3.5, of quarterly sales (in 2011–2014) for product EDM-617.

Step 1 Initial Estimation of Level (Including Trend) at Each Historical Period. The demand in any historical period is assumed to be composed of trend (incorporating the level), seasonal, and random components. To estimate the seasonal indices, we must first remove the trend effect. The trend point (i.e., the level) for any particular period t is estimated by a moving average of a full season (i.e., P periods) centered at period t . A full season is used in order to have the moving average free of seasonal effects. Because the most common seasons have an even number of periods ($P = 4$ or 12), the standard P -period moving average ends up being centered between two periods, not right at the middle of a period as desired. To overcome this problem, we take the average of two consecutive moving averages, as illustrated in Table 3.5. In columns 5 and 6, we show 4-period moving averages that are computed from consecutive time periods. For example, the first entry of column 5, 54.25, is the average of the first four data points (43, 57, 71, and 46), and therefore is centered between quarters 2 and 3 of 2011. The first entry of column 6, 56.00, is the average of the four periods beginning at quarter 2 of 2011 (or, 57, 71, 46, and 50), and therefore is centered between quarters 3 and 4 of 2011. Column 7 is the average of these two averages, and, therefore, is the centered 4-period moving average for quarter 3 of 2011. These centered moving averages are also portrayed in Figure 3.5.

Step 2 Estimates of Seasonal Factors. The estimate of the seasonal factor (column 8 of Table 3.5) for any particular historical period t is obtained by dividing the demand x_t (column 4) by the centered moving average (column 7). Note that, because of the moving average procedure, we cannot obtain estimates for the first half-season (first two quarters of 2011) nor for the last half-season (last two quarters of 2014). Estimates still include the random components. In an effort to dampen the random effect, we average the seasonal factors for similar periods in different years. For example, in Table 3.5, the average of the 3rd quarter factors is $(1.288 + 1.365 + 1.286)/3$ or 1.313. The averages obtained in this fashion are shown in column 2 of Table 3.6. The total of the averages need not add up to exactly P (4 in our example). Thus,

Table 3.5 Initialization of Seasonal Model for Item EDM-617

Year (1)	Quarter (2)	Period t (3)	Demand x_t (4)	Moving Average 1 (5)	Moving Average 2 (6)	Centered 4-Period Moving Average (7)	Estimate of F_t (8) = (4) - (7)
2011	1	-15	43				
	2	-14	57				
	3	-13	71	54.25	56.00	55.13	1.288
	4	-12	46	56.00	57.00	56.50	0.814
2012	1	-11	50	57.00	60.50	58.75	0.851
	2	-10	61	60.50	60.75	60.63	1.006
	3	-9	85	60.75	63.75	62.25	1.365
	4	-8	47	63.75	67.00	65.38	0.719
2013	1	-7	62	67.00	67.25	67.13	0.924
	2	-6	74	67.25	67.50	67.38	1.098
	3	-5	86	67.50	66.25	66.88	1.286
	4	-4	48	66.25	67.25	66.75	0.719
2014	1	-3	57	67.25	71.25	69.25	0.823
	2	-2	78	71.25	74.50	72.88	1.070
	3	-1	102				
	4	0	61				

we normalize to obtain the estimates of seasonal indices that total to P . The results are shown in column 3 of Table 3.6.

Step 3 Estimating \hat{a}_0 and \hat{b}_0 . We still need estimates of the level and trend terms in Equation 3.43.

Using the seasonal indices found in Step 2, we deseasonalize the data as shown in Table 3.7. (Once again, we have rounded the data in the table, but the spreadsheet that computes column 6 retains many more digits.) The results are also shown graphically in Figure 4.8. They are reasonably close to the centered moving averages found in Table 3.5, because, in this example, the seasonal coefficients found in Table 3.5 do not fluctuate very much from year to year. In any event, a regression line is now fit to the data of column 6 in Table 3.7 using the procedure of Equations 3.36 and 3.37 of Section 3.4.3. This gives us estimates \hat{a}_0 and \hat{b}_0 as of the end of 2014 (the end of period 0). The resulting line is shown in Figure 3.6; the parameter values are

$$\hat{a}_0 = 76.866 \quad \text{and} \quad \hat{b}_0 = 1.683$$

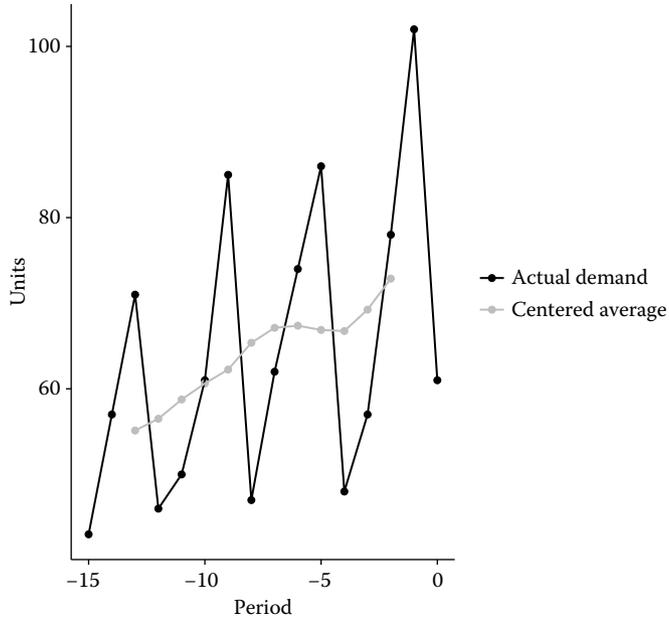


Figure 3.5 Historical data for item EDM-617.

Again, one could use an earlier origin (with \hat{a}_0 appropriately adjusted) and an associated run-in period through the rest of the historical data. Also, in the event that the plotted data indicate a shift in trend part way through the history, the regression line should only be fit to the latest data (beyond the shift). (See Williams (1987) for an alternate way of initializing the parameters.)

Table 3.6 Initial Seasonal Indices for Item EDM-617

Quarter <i>i</i> (1)	Average Estimate (2)	Normalized Index F_i (3)
1	0.866	0.869
2	1.058	1.061
3	1.313	1.317
4	0.751	0.753
Total	3.988	4.000

Table 3.7 Calculations to Estimate \hat{a}_0 and \hat{b}_0 Item EDM-617

Year (1)	Quarter (2)	Period t (3)	Demand x_t (4)	F_t (5)	Estimate of Level (6) = (4)/(5)
2011	1	-15	43	0.869	49.509
	2	-14	57	1.061	53.701
	3	-13	71	1.317	53.908
	4	-12	46	0.753	61.091
2012	1	-11	50	0.869	57.569
	2	-10	61	1.061	57.469
	3	-9	85	1.317	64.538
	4	-8	47	0.753	62.419
2013	1	-7	62	0.869	71.385
	2	-6	74	1.061	69.717
	3	-5	86	1.317	65.297
	4	-4	48	0.753	63.747
2014	1	-3	57	0.869	65.628
	2	-2	78	1.061	73.485
	3	-1	102	1.317	77.445
	4	0	61	0.753	81.012

3.4.4.4 Forecasting

The underlying model includes level, trend, and seasonal factors. As of the end of period t we have

$$\hat{x}_{t,t+\tau} = (\hat{a}_t + \hat{b}_t\tau)\hat{F}_{t+\tau-P} \tag{3.47}$$

where $\hat{x}_{t,t+\tau}$ is the forecast, made at the end of period t , of demand in period $t + \tau$ and $\hat{F}_{t+\tau-P}$ is the most recent estimate of the seasonal index for period $t + \tau$. (This assumes $\tau \leq P$; the last *direct* update of $\hat{F}_{t+\tau-P}$ was made in period $t + \tau - P$. However, when any other seasonal index is updated, renormalization could cause $\hat{F}_{t+\tau-P}$ to change somewhat.)

As earlier, the forecast of cumulative demand from the beginning of period $t + 1$ through to the end of period u is given by

$$\hat{x}_{t,t,u} = \sum_{\tau=1}^{u-t} \hat{x}_{t,t+\tau} \tag{3.48}$$

where again, if u is not an integer, the last term in the summation reflects an appropriate proportion of the last period.

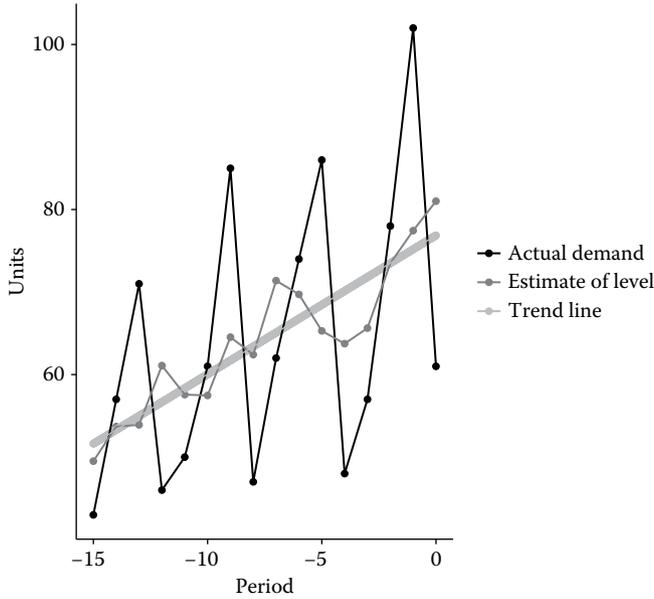


Figure 3.6 Determination of trend line for item EDM-617.

3.4.4.5 Illustration

We again use the item EDM-617. First, consider the forecast $\hat{x}_{0,1}$ of the demand in the first quarter of 2015 made at the end of 2014. The previous first quarter was $t = -3$. From Table 3.6, we thus have $\hat{F}_{-3} = 0.869$. Also $\hat{a}_0 = 76.866$ and $\hat{b}_0 = 1.683$. Thus, from Equation 3.47 we have

$$\hat{x}_{0,1} = (76.866 + 1.683)0.869 = 68.222 \text{ units}$$

Similarly, the forecasts at the end of 2014 for the other three quarters of 2015 are

$$\hat{x}_{0,2} = (76.866 + 1.683(2))1.061 = 85.161 \text{ units}$$

$$\hat{x}_{0,3} = (76.866 + 1.683(3))1.317 = 107.886 \text{ units}$$

$$\hat{x}_{0,4} = (76.866 + 1.683(4))0.753 = 62.946 \text{ units}$$

Suppose that the actual demand in the first quarter of 2015 was 75 units. The level, trend, and seasonality factors can now be updated. Assume that the smoothing constants α_{HW} , β_{HW} , and γ_{HW} have values of 0.2, 0.1, and 0.3, respectively. From Equation 3.44, the updated value of the level is given by

$$\begin{aligned} \hat{a}_1 &= \alpha_{HW}(x_1/\hat{F}_{-3}) + (1 - \alpha_{HW})(\hat{a}_0 + \hat{b}_0) \\ &= 0.2(75/0.866) + 0.8(76.866 + 1.683) = 80.110 \end{aligned}$$

Next, from Equation 3.45, the revised estimate of the trend is given by

$$\begin{aligned}\hat{b}_1 &= \beta_{HW}(\hat{a}_1 - \hat{a}_0) + (1 - \beta_{HW})\hat{b}_0 \\ &= 0.1(80.110 - 76.866) + 0.9(1.683) = 1.839\end{aligned}$$

Then, by using Equation 3.46, we obtain the updated estimate of the seasonality factor for the first quarter as

$$\begin{aligned}\hat{F}_1 &= \gamma_{HW}(x_1/\hat{a}_1) + (1 - \gamma_{HW})\hat{F}_{-3} \\ &= 0.3(75/80.110) + 0.7(0.869) = 0.889\end{aligned}$$

Because of the change in this seasonality factor, we renormalize the four seasonality factors so that they again add to 4:

$$1.061 + 1.317 + 0.753 + 0.889 = 4.020$$

Therefore,

$$\begin{aligned}\hat{F}_{-2} &= 1.061 \left(\frac{4.00}{4.02} \right) = 1.056 \\ \hat{F}_{-1} &= 1.317 \left(\frac{4.00}{4.02} \right) = 1.310 \\ \hat{F}_{-0} &= 0.753 \left(\frac{4.00}{4.02} \right) = 0.749 \\ \hat{F}_1 &= 0.889 \left(\frac{4.00}{4.02} \right) = \underline{0.884} \\ &\text{Total} = 4.000\end{aligned}$$

Note that demand for period 1 was forecast (at the end of period 0) as 68.222 units. Actual demand turned out to be 75 units; that is, demand was underforecast by 6.778 units. Because of this forecast error, the level was adjusted upward from 76.866 to 80.110—an increase of 3.244 units, considerably more than was predicted by the original trend value of 1.683 units. Moreover, the estimate of the trend was increased from 1.683 to 1.839, and the estimate of the seasonality factor was raised from 0.869 to 0.884.

The new estimates of the parameters can now be used in Equation 3.47 to provide a revised set of forecasts as of the end of period 1 (the first quarter of 2015). To illustrate, the new forecast of demand in the second quarter of 2015 is

$$\begin{aligned}\hat{x}_{1,2} &= (\hat{a}_1 + \hat{b}_1)\hat{F}_{-2} \\ &= (80.110 + 1.839)(1.056) = 86.544\end{aligned}$$

The forecasts for the remaining two quarters of 2015 are also shown in Table 3.8.

Table 3.8 Effect of Forecast Error in Period 1 on Forecasts of Demand for Later Periods

<i>Calendar Time</i>	<i>Period t</i>	<i>Forecast Calculated at the End of Period 0</i>	<i>Forecast Calculated at the End of Period 1</i>
1st quarter 2015	1	68.222	75.0 (actual)
2nd quarter 2015	2	85.161	86.544
3rd quarter 2015	3	107.886	109.796
4th quarter 2015	4	62.946	64.149

Once the figure for actual sales in quarter 2 of 2015 becomes available, the above procedures must be repeated—the parameters (level, trend, and seasonal factors) revised and a new set of revised forecasts issued.

3.4.4.6 General Comments

Clearly, updating, forecasting, and particularly initialization are considerably more complicated with a seasonal model than was the case with the earlier nonseasonal models. Implementing the equations on a spreadsheet also requires a bit more creativity. (See Problem 3.17.) Thus, the decision to use a seasonal model should be taken with care. In the initialization stage, if the initial F 's are all close to unity or fluctuate widely from season to season, then, a seasonal model is probably inappropriate. (See Sections 3.6 and 3.7; Chatfield 1978, 2013; McLeavey et al. 1981; Rhodes 1996.)

There are other, quite different approaches to dealing with seasonality. In particular Brown (1981, Chapter 9), among others, argues for the use of a transcendental model that is a mix of sine waves. This approach has the merit of being more parsimonious (fewer parameters than seasonal indices) and hence tends to be more stable, particularly when the demand level is quite low. However, we feel that this advantage is usually outweighed by the much more intuitively understandable nature of the seasonal indices model.

Finally, there are certain “seasonal” effects whose changing nature can be predicted. One of the best examples is the Easter holiday that for some years occurs in March and some in April. Although the monthly data can be distorted by this effect, the distortion can be anticipated, so that it is necessary for human input to modify the forecasts. Regular promotions are another example; they may occur each summer, but at slightly different times. (Section 9.7.1 addresses some of these issues for style goods.) Another anomalous situation arises where it is crucial (perhaps for production-scheduling purposes) to use monthly data to forecast the demand rate per day. Under these circumstances, one should take into account of the varying numbers of days from month to month. In other circumstances, it is simply adequate to consider the differing numbers of days as being part of the contribution to differing seasonal factors.

3.4.5 Selection of Smoothing Constants

We now address the choice of the smoothing constants for the various smoothing procedures discussed earlier. For the moment, we restrict our attention to a static choice, that is, values that will

essentially remain unchanged over time for any particular SKU. In Section 3.5.4, the case where the smoothing constants can vary with time, so-called *adaptive smoothing*, will be treated.

There is a basic trade-off in the choice of the values of the smoothing constants. Small values, which give little weight to the most recent data, hence considerable weight to the history, are appropriate under stable conditions—they effectively smooth out most of the random noise. On the other hand, if the parameters of the underlying demand model are changing appreciably with time, then higher values of the smoothing constants are in order, because they give more weight to recent data and thus more quickly bring the parameter estimates into line with the true changed values. Unfortunately, we are never really sure whether forecast errors are a result of random (transient) noise or a real change in one or more of the parameter values in the underlying demand model. Therefore, a compromise must be made in choosing the values of the smoothing constants. Harrison (1967) quantifies this compromise under special assumptions about the inherent variability of the model parameters. (See also Adshead and Price 1987; Johnston and Boylan 1994.) An additional point related to the same issue is that, all other factors being equal, the smoothing constants should increase with the length of the forecast update period, because the likelihood of changes in the model parameters increases with the passage of time.

In general, we suggest the use of a search experiment to establish reasonable values of the smoothing constants. For each of a number of representative items, the following experiment should be conducted. The available demand history is divided into two sections. The first section is used to initialize the model parameters in the updating procedure selected. Then, with these initial values, the smoothing is carried through the second portion of the data, using a specific combination of values for the smoothing constants. Some appropriate measure of effectiveness (e.g., the mean of the squared forecast errors to be discussed in Section 3.5) is evaluated. This is repeated, on a grid search basis, to find the combination of values of smoothing constants that minimizes the measure of effectiveness across a class of similar items. In actual fact, exact minimization is not required because the measures of effectiveness always tend to be quite insensitive to deviations of the smoothing constants from their best values. Furthermore, we wish to use only a few sets of smoothing constants, each set being employed for a broad class of items.

Moreover, any case where minimization occurs when unusually large values are used for the smoothing constants should be viewed with caution. As an illustration, a large value (>0.3) of α in the simple smoothing procedure should raise the question of the validity of the assumed underlying level model. The high smoothing constant, in this case, suggests that perhaps a trend model is more appropriate. Further discussion on the aforementioned search can be found in Winters (1960), and Gardner and Dannenbring (1980).

Keeping these suggestions in mind, we now present, for each of the smoothing procedures discussed earlier, ranges of reasonable values of the parameters found through our experience and that of several other individuals, as reported in the literature.

3.4.5.1 *Simple Exponential Smoothing*

Recall from Equation 3.23 that the updating procedure is

$$\hat{a}_t = \alpha x_t + (1 - \alpha)\hat{a}_{t-1}$$

The likely range of α is 0.01–0.30 with a compromise value of 0.10 often being quite reasonable.

3.4.5.2 Smoothing for a Trend Model

Earlier, we suggested the use of Brown’s single parameter updating method. For reference purposes, we repeat Equations 3.32 and 3.33.

$$\hat{a}_t = [1 - (1 - \alpha)^2]x_t + (1 - \alpha)^2(\hat{a}_{t-1} + \hat{b}_{t-1}) \tag{3.49}$$

$$\hat{b}_t = \left[\frac{\alpha^2}{1 - (1 - \alpha)^2} \right] (\hat{a}_t - \hat{a}_{t-1}) + \left[1 - \frac{\alpha^2}{1 - (1 - \alpha)^2} \right] \hat{b}_{t-1} \tag{3.50}$$

The range of α values prescribed under simple smoothing is also applicable here.

3.4.5.3 Winters Seasonal Smoothing

Repeating Equations 3.44 through 3.46, the updating procedure is

$$\hat{a}_t = \alpha_{HW}(x_t/\hat{F}_{t-p}) + (1 - \alpha_{HW})(\hat{a}_{t-1} + \hat{b}_{t-1}) \tag{3.51}$$

$$\hat{b}_t = \beta_{HW}(\hat{a}_t - \hat{a}_{t-1}) + (1 - \beta_{HW})\hat{b}_{t-1} \tag{3.52}$$

$$\hat{F}_t = \gamma_{HW}(x_t/\hat{a}_t) + (1 - \gamma_{HW})\hat{F}_{t-p} \tag{3.53}$$

As pointed out earlier (in Equations 3.34 and 3.35), Equations 3.49 and 3.50 are a special case of the Holt procedure where

$$\alpha_{HW} = [1 - (1 - \alpha)^2] \tag{3.54}$$

and

$$\beta_{HW} = \frac{\alpha^2}{1 - (1 - \alpha)^2} \tag{3.55}$$

This suggests that reasonable guidelines for the choices of α_{HW} and β_{HW} in Equations 3.51 and 3.52 are provided by substituting the above suggested α values into Equations 3.54 and 3.55. The results are shown in Table 3.9 along with the suggested values of γ_{HW} , the latter found through experimentation. Again, we emphasize that, where possible, a search experiment should be conducted with demand data for the particular population of items involved.*

Table 3.9 Reasonable Values of Smoothing Constants in the Winters Procedure

	Underlying α Value	α_{HW}	β_{HW}	γ_{HW}
Upper end of range	0.30	0.51	0.176	0.50
Reasonable single value	0.10	0.19	0.053	0.10
Lower end of range	0.01	0.02	0.005	0.05

* For stability purposes (see McClain and Thomas 1973), the value of γ_{HW} should be kept well below that of α_{HW} .

3.5 Measuring the Performance of a Forecasting Process

No matter how forecasts are generated, we are interested in evaluating the quality of the forecasts for two reasons. First, the historical performance of a forecasting process helps us create the description of future demand to support resource allocation decisions. Second, by tracking the performance of a forecasting process over time, we may be able to identify ways to improve that process. When we speak of the *accuracy* of a forecast, we generally mean some measure of how far away our point-estimate forecast is from the actual value. A challenge is trying to separate the *dispersion* of forecast errors from the *bias*. Measures of bias tell us how far off the forecasts are from the target *on average*, but they do not tell us how widely or narrowly scattered the forecast errors are.

Significant bias may signal that the parameters of the underlying demand model have been inaccurately estimated or that the model itself is incorrect. Alternatively, it may be human adjustments to forecasts that are causing the bias. We revisit the impact of human judgment in Section 3.7. Where forecasts are used for production planning and inventory management decision making, either of these situations is undesirable; removing bias from the forecast will improve the performance.

A common measure of bias is the average error over some time horizon, say n periods, where periods could be days, weeks, months, etc. As noted above, we represent the forecast error for lag- τ forecasts by $e_{t-\tau,t}$.

$$\bar{e} = \text{Average Error} = \sum_{t=1}^n e_{t-\tau,t} / n \quad (3.56)$$

A closely related measure sometimes used in practice is the *sum* (rather than the average) of forecast errors. This is sometimes referred to as the running sum of forecast errors.*

$$C_n = \text{Sum of Errors} = \sum_{t=1}^n e_{t-\tau,t} \quad (3.57)$$

Both \bar{e} and C_n allow positive and negative errors to cancel each other out over time. So, with either of these measures, if their value is close to zero, we say that the forecasting process is *unbiased*; forecasts (point estimates) are not *consistently* too high or *consistently* too low. “Close to zero” is not a terribly precise term obviously; so, we need to be concerned with scaling or interpreting \bar{e} and C_n to try to decide whether our forecast is biased. The challenge is that due to the random component in a demand pattern, it is not easy to detect an underlying bias; the noise tends to camouflage it. We return to this issue in Section 3.9.2 after introducing measures of dispersion and accuracy.

The standard deviation of forecast errors for a forecasting process (forecasting for a unit period), σ_1 , gives us a measure of *dispersion* of forecast errors

$$\sigma_1 = \sqrt{\frac{\sum_{t=1}^n (e_{t-\tau,t} - \bar{e})^2}{(n-1)}}$$

* Another measure of bias commonly used in practice is the average of actual values divided by forecasts. If over time this value is close to 100%, we conclude that our forecasting process is unbiased. This kind of measure is more frequently used as a summary performance measure for many forecasts aggregated.

where \bar{e} is the average forecast error defined in Equation 3.56. σ_1 tells how variable forecast errors are but does not speak to bias.

It is critical to note that none of these measures by themselves gives a complete description that could support a resource allocation decision. To see why, suppose we were evaluating a forecasting process that generated forecasts that were always *exactly* 1,000 units too high (e.g., $x_t - \hat{x}_{t,t} = -1,000$). The average error $\bar{e} = -1,000$ and the standard deviation of forecast errors $\sigma_\tau = 0$. This forecasting process has no dispersion or variability but is clearly quite biased. Alternatively, suppose the forecast errors alternated between 1,000 and $-1,000$. Such a forecasting process is unbiased (average error $\bar{e} = 0$) but has high dispersion. As we will discuss further below, when forecast bias is detected, identifying and eliminating the cause of the bias must be a priority.

3.5.1 Measures of Forecast Accuracy

In this section, we introduce three widely used measures of forecast accuracy, MSE, mean absolute deviation (MAD), and mean absolute percent error (MAPE). As will be demonstrated in Chapter 6, for purposes of establishing the safety stock of an individual item (to provide an appropriate level of customer service), we will need an estimate of the standard deviation (σ_L) of the errors of forecasts of total demand over a period of duration L (the replenishment lead time). In this section, we concentrate on estimating σ_1 , the standard deviation of errors of forecasts of demand made for a unit period. (A unit period is a forecast update interval.) In Section 3.5.2, we discuss the conversion of the estimate of σ_1 to an estimate of σ_L . The procedures to be discussed are appropriate for A and B items. C items will be handled by simpler methods to be presented later.

Suppose that we have two types of information for each of n unit time periods, specifically (1) the actual observed demands, x_1, x_2, \dots, x_n and (2) the one-period-ahead forecasts (by some particular forecasting procedure) $\hat{x}_{0,1}, \hat{x}_{1,2}, \dots, \hat{x}_{n-1,n}$ where, as earlier, $\hat{x}_{t,t+1}$ is the forecast, made at the end of period t , of demand in period $t + 1$. The measure of variability that is often used (minimized) in fitting of squared errors of a straight line to historical data is the MSE. When forecasts are unbiased, the MSE is directly related to σ_1 , as we shall see shortly. Because of these properties, and because it can be computed easily on a spreadsheet, we recommend its use. The MSE for one-period-ahead forecasts is given by

$$\text{MSE} = \frac{1}{n} \sum_{t=1}^n (x_t - \hat{x}_{t-1,t})^2 = \sum_{t=1}^n (e_{t-1,t})^2 \quad (3.58)$$

A second measure of variability, the MAD, was originally recommended for its computational simplicity. With the widespread availability of computers, however, the MAD is of less practical importance. Nevertheless, it is intuitive, and still frequently used in practice. For the n periods of data, the estimate of the MAD for one-period-ahead forecasts would be

$$\text{MAD} = \frac{1}{n} \sum_{t=1}^n |x_t - \hat{x}_{t-1,t}| = \frac{1}{n} \sum_{t=1}^n |e_{t-1,t}| \quad (3.59)$$

The term $x_t - \hat{x}_{t-1,t}$ is the error ($e_{t-1,t}$) or the deviation in the forecast for period t . The vertical lines indicate that we take the absolute value of the difference. Finally, the mean of the absolute deviations results by the summation and division by n .

The MAPE is another intuitive measure of variability. It generally is not affected by the magnitude of the demand values, because it is expressed as a percentage. However, it is not appropriate if demand values are very low. For example, a forecast of 1 unit of demand matched with an actual value of 2 units shows an error of 100%. The manager should decide whether a reported error of 100%, or of 1 unit, is more desirable. The MAPE is given by

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{x_t - \hat{x}_{t-1,t}}{x_t} \right| = \frac{1}{n} \sum_{t=1}^n \left| \frac{e_t}{x_t} \right| \tag{3.60}$$

3.5.1.1 Numerical Illustration

Consider an item with the demands and forecasts as shown for six periods in Table 3.10. The MSE is computed using Equation 3.58.

$$\text{MSE} = \frac{1}{6} \sum_{t=1}^6 (e_t^2) = 206/6 = 34.33$$

Using Equation 3.59, the estimate of MAD is

$$\text{MAD} = \frac{1}{6} \sum_{t=1}^6 |e_t| = 32/6 = 5.33$$

Again, as seen in Figure 3.7, the MAD has the appealing graphical interpretation of being the average distance (ignoring direction) between the actual demand in a period and the forecast of that demand.

Table 3.10 Illustration of Computation of MAD, MAPE, and MSE

Period <i>t</i>	1	2	3	4	5	6	Total	Average
Actual demand x_t	100	87	89	87	95	85		
One-period-ahead forecast made at the end of the previous period $\hat{x}_{t-1,t}$	90	92	91	91	90	91		
Error or deviation $e_t = x_t - \hat{x}_{t-1,t}$	10	-5	-2	-4	5	-6		
Squared deviation e_t^2	100	25	4	16	25	36	206	34.33
Absolute deviation $ e_t $	10	5	2	4	5	6	32	5.33
Percent deviation $ e_t/x_t $	10.0%	5.7%	2.2%	4.6%	5.3%	7.1%	34.9%	5.82%

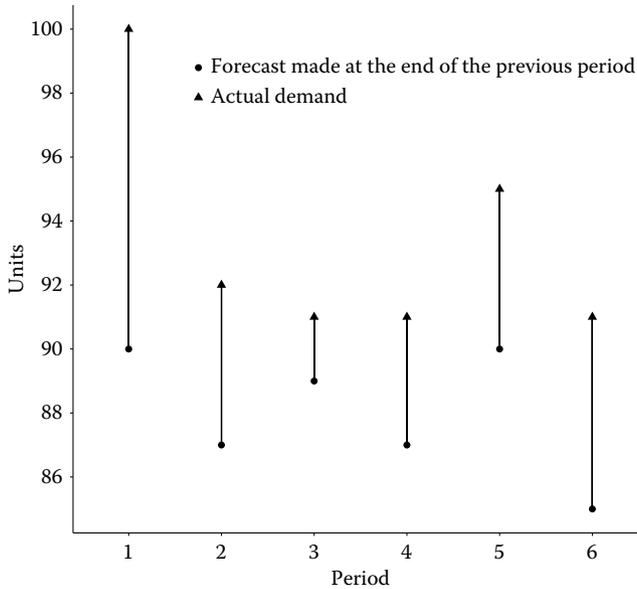


Figure 3.7 Illustration of the MAD.

The MAPE is computed using Equation 3.60

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{e_t}{x_t} \right| = 5.82\%$$

3.5.1.2 Conversion to σ_1

To find the standard deviation of forecast errors, for the purpose of setting safety stock levels, we can use a simple relationship between the value of σ_1 and the MSE:

$$\sigma_1 = \sqrt{\text{MSE}} \quad (3.61)$$

This relationship assumes that forecasts are unbiased. As noted above, removing bias from a forecasting process should be a priority. If bias remains, the approach in Equation 3.61 will tend to overestimate σ_1 . So, with unbiased forecasts, a reasonable estimate of σ_1 , denoted as $\hat{\sigma}_1$, is the square root of the MSE calculated from the sample data. We illustrate how to initialize the MSE below.

The relationship of σ_1 to the MAD is not as simple. However, for the normal distribution and unbiased forecasts, we can show that the following relationship holds between the true MAD (the average absolute deviation from the mean) and the true standard deviation (See the Appendix of this chapter for a derivation.):

$$\sigma = (\sqrt{\pi/2})\text{MAD} = 1.25 \text{ MAD} \approx \sqrt{\text{MSE}}$$

Thus, in cases where the normal distribution is appropriate, one can estimate σ_1 via

$$\hat{\sigma}_1 = 1.25 \text{ MAD}$$

For several other common distributions, the theoretical conversion factor from the MAD to σ is not very different from 1.25. However, the factor can vary enough so that the safety stock obtained may not provide the required level of service. See Jacobs and Wagner (1989). Therefore, we recommend using the MSE.

3.5.1.3 Updating of MSE

In theory, Equation 3.58 could be used to update the estimate of MSE each time an additional period's information became available. Instead, we advocate a simple exponential smoothing form of updating

$$\text{MSE}_t = \omega(x_t - \hat{x}_{t-1,t})^2 + (1 - \omega)\text{MSE}_{t-1} \quad (3.62)$$

where MSE_t is the estimate of MSE at the end of period t and ω is a smoothing constant. (A small value of ω , namely between 0.01 and 0.10, should normally be used.)

There are two reasons for using Equation 3.62 instead of Equation 3.58. First, storage of an indefinitely long history is required for the use of Equation 3.58, whereas only the most recent MSE_{t-1} need be stored for Equation 3.62. Second, in using Equation 3.58, each period of history is given equal weight. In contrast, geometrically decaying weights occur through the use of Equation 3.62. In fact, these arguments are identical to those presented earlier in the chapter when we advocated exponential smoothing, instead of repeated least-squares regression, for updating the forecasts themselves.

3.5.1.4 Initialization of MSE

For each of the forecasting procedures discussed in Section 3.4, we suggested a method for estimating initial values of the parameters of the underlying model of demand, using sufficient available historical data. For such a situation, the initial estimate of MSE is obtained through the use of the following equation that is closely related to Equation 3.58:

$$\text{MSE}_0 = \sum_{t=-(n-1)}^0 (x_t - \tilde{x}_t)/(n - p) \quad (3.63)$$

where the n historical periods are numbered from 0 (the most recent period) back through $-1, -2, \dots, -(n-1)$, and \tilde{x}_t is the estimate of x_t resulting from the underlying demand model when the estimates of its p parameters are obtained from the historical data. (E.g., if the underlying model was a linear trend, $x_t = a + bt$, then p would be 2 and \tilde{x}_t would be $\hat{a}_0 + \hat{b}_0 t$.)

The denominator in Equation 3.63 is $n - p$, rather than n , to reflect that our estimate of MSE is based on an ex-post fit to the data, where the data have already been used to estimate the p parameters of the model.

The case where insufficient historical information exists will be treated in Section 3.8.1.

Table 3.11 Illustration of Computation of MSE_0 for the Example of Table 3.3

	$\hat{a}_0 = 27.19$		$\hat{b}_0 = 1.34$
t	x_t	$\tilde{x}_t = \hat{a}_0 + \hat{b}_0 t$	$(x_t - \tilde{x}_t)^2$
-5	20	20.48	0.23
-4	25	21.82	10.12
-3	21	23.16	4.67
-2	22	24.50	6.27
-1	27	25.85	1.33
0	28	27.19	0.66
		Total	23.28

3.5.1.5 Illustration

In Table 3.3, we showed the initialization of a trend model using six periods of historical information. The estimated values of the two parameters were $\hat{a}_0 = 27.19$ and $\hat{b}_0 = 1.34$. The original demand data and the \tilde{x}_t values are shown in Table 3.11. By using Equation 3.63, we obtain

$$\begin{aligned} MSE_0 &= \sum_{t=-5}^0 (x_t - \tilde{x}_t)^2 / (6 - 2) \\ &= (23.28) / 4 = 5.82 \end{aligned}$$

3.5.2 Estimating the Standard Deviation of Forecast Errors over a Lead Time

As mentioned earlier, the replenishment lead time is unlikely to be equal to the forecast update interval. Thus, we need a method of converting $\hat{\sigma}_1$ (our estimate of σ_1) into $\hat{\sigma}_L$ (the corresponding estimate of σ_L). The exact relationship between σ_L and σ_1 depends in a complicated fashion on the specific underlying demand model, the forecast-updating procedure, and the values of the smoothing constants used. See, for example, Brown (1963), Harrison (1967), and Johnston and Harrison (1986). One of the reasons for the complexity is that the smoothing procedure introduces a degree of dependence between the forecast errors in separate periods. Fortunately, however, we have found that for most inventory systems, the following model satisfactorily captures, empirically, the required relationship:

$$\hat{\sigma}_L = L^c \hat{\sigma}_1 \quad (3.64)$$

where

$\hat{\sigma}_L$ = estimate of the standard deviation of forecast errors over a lead time of duration L basic (forecast update) periods (L need not be an integer)

$\hat{\sigma}_1$ = estimate of the standard deviation of forecast errors over one basic (forecast update) period

c = coefficient that must be estimated empirically

Table 3.12 Forecasts Simulated to Estimate c

Length of L	Forecasts		Number of Forecasts for Each SKU
1	$\hat{x}_{t,t+1}$	$t = 0, 1, 2, \dots, 35$	36
2	$\sum_{\tau=1}^2 \hat{x}_{t,t+\tau}$	$t = 0, 2, 4, \dots, 34$	18
3	$\sum_{\tau=1}^3 \hat{x}_{t,t+\tau}$	$t = 0, 3, 6, \dots, 33$	12
\vdots	\vdots	\vdots	\vdots
6	$\sum_{\tau=1}^6 \hat{x}_{t,t+\tau}$	$t = 0, 6, 12, \dots, 30$	6

To check the reasonableness of Equation 3.64 for a particular firm, we proceed as follows. We select, say, 10 representative SKUs. We should pick, say, 6 from the A category and 4 from the B category, for which a number of years of historical monthly demand data are available. Assume that we have 7 years of data. Using the first 4 years of data, we initialize the Winters exponential smoothing model by the procedures described in Section 3.4.4. That is, for each of the 10 representative SKUs, we obtain the data needed to initialize the Winters forecast equation (Equation 3.47):

$$\hat{x}_{t,t+\tau} = (\hat{a}_t + \hat{b}_t\tau)\hat{F}_{t+\tau-p} \tag{3.65}$$

Then, for the fifth, sixth, and seventh years of the demand data, Equation 3.65, along with the Winters updating procedures described in Equations 3.44 through 3.46, is used to issue the forecasts listed in Table 3.12. In all cases, the model parameters are updated monthly. For each forecast in Table 3.12, the forecast figure is compared to the actual demand that resulted over the immediate period of duration L , and the associated forecast error is computed by using

$$\text{Forecast error} = e_t(L) = \sum_{\tau=1}^L \hat{x}_{t,t+\tau} - \sum_{\tau=1}^L x_{t+\tau} \tag{3.66}$$

For each value of L , the sample standard deviation of forecast errors is computed and used as an estimate of $\hat{\sigma}_L$

$$\hat{\sigma}_L = s_L = \left\{ \frac{1}{n-1} \sum_t [e_t(L) - \bar{e}(L)]^2 \right\}^{1/2}$$

where

- $\bar{e}(L) = \sum_t e_t(L)/n$ is the average error for the L under consideration
- $n =$ number of lead times of length L used (see the right-hand column of Table 3.12)

As a result of all the above calculations, there are 10 values (one for each of the SKUs) of $\hat{\sigma}_L$ for each of $L = 1, 2, 3, \dots, 6$. For each SKU, the 5 values of $\hat{\sigma}_L/\hat{\sigma}_1$ are computed for $L = 2, 3, 4, 5, 6$. The 50 values of $\log(\hat{\sigma}_L/\hat{\sigma}_1)$ are then plotted against $\log L$, as shown in Figure 3.8. Equation 3.64 can be written as

$$\hat{\sigma}_L/\hat{\sigma}_1 = L^c \text{ or } \log(\hat{\sigma}_L/\hat{\sigma}_1) = c \log L$$

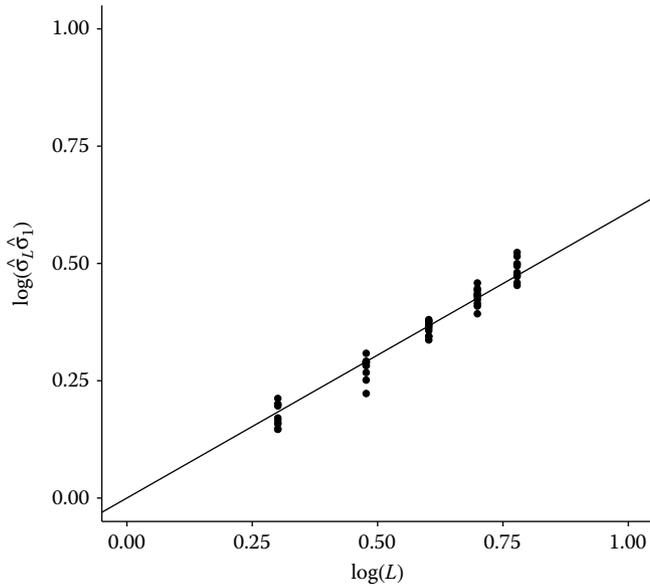


Figure 3.8 Relative standard deviations of forecast errors versus lead time.

Therefore, the slope of the line through the origin that gives a reasonable fit to the 50 data points is an estimate of c . The line with slope $c = 0.6$, plotted in Figure 3.8, gives a reasonable fit for the data shown. So, for these data, we choose to use this approximation, which gives

$$\begin{aligned}\log(\hat{\sigma}_L/\hat{\sigma}_1) &= 0.60 \log L \\ \hat{\sigma}_L &= L^{0.60} \hat{\sigma}_1 = \sqrt{L} \hat{\sigma}_1\end{aligned}\quad (3.67)$$

The result of Equation 3.67 could also be obtained by assuming that in an L period, the forecast errors in consecutive periods are independent and each has standard deviation σ_1 . In actual fact, both of these assumptions are violated to some extent. Nevertheless, the empirical behavior of Equation 3.67 is often quite reasonable. However, we hasten to recommend that an analysis, such as that leading to Figure 3.8, be carried out for any system under consideration. If a straight line through the origin is a reasonable fit, then, its slope, which need not be 0.6, gives the estimate of c in Equation 3.64.

3.5.3 Monitoring Bias

As noted above, we use the word *bias* to indicate that, on average, the forecasts are substantially above or below the actual demands. A bias signals some fundamental problem in the forecasting process. This could be the result of manual intervention in the forecasts or, if we are simply evaluating forecasts coming from a statistical model, bias would indicate that the parameters of the underlying demand model have been inaccurately estimated or that the model itself is incorrect. Where forecasts are used for production planning and inventory management decision making, either of these situations is undesirable. Because of the random component in a demand pattern, it is not easy to detect an underlying bias; the noise tends to camouflage it. In general, the forecast

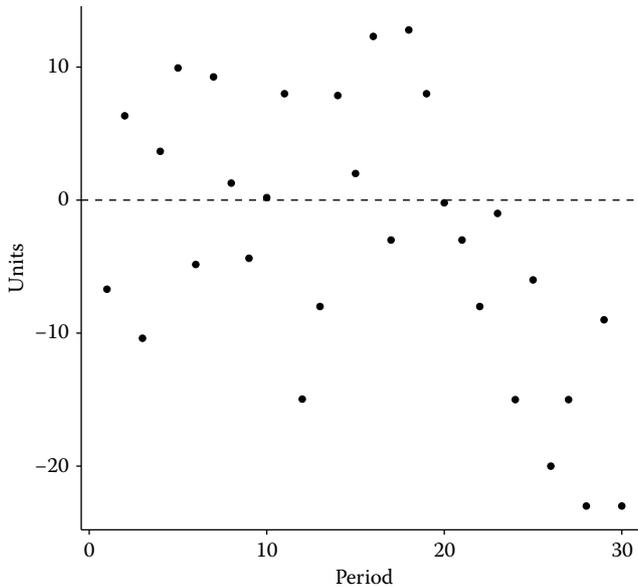


Figure 3.9 An illustration of bias in forecast errors.

error, e_t , should fluctuate around zero. A simple graph of forecast errors can indicate whether they are showing a trend on the high or low side. Figure 3.9 illustrates a case in which there is no bias until about month 20, and then the errors show a negative trend. In this section, we discuss three more rigorous measures of bias; then, in Section 3.5.4, we briefly present the subsequent corrective actions.

3.5.3.1 Cumulative Sum of Forecast Errors

Harrison and Davies (1964) suggest the use of cumulative sum techniques to monitor the bias in a forecasting procedure. Specifically, the cumulative sum of forecast errors is computed recursively using

$$C_t = C_{t-1} + e_t \tag{3.68}$$

where

- C_t = cumulative sum of forecast errors as of the end of period t , see Equation 3.57
- $e_t = x_t - \hat{x}_{t-1,t}$ = one-period-ahead forecast errors in period t

It is advisable to normalize with respect to the MSE; that is, use

$$U_t = C_t / \sqrt{\text{MSE}_t} \tag{3.69}$$

where U_t is the cumulative sum tracking signal, and MSE_t is the MSE at the end of period t .

U_t can take on any value, which is positive or negative. Nevertheless, if the forecasting procedure is unbiased, U_t should fluctuate around 0. An interesting property of the cumulative sum C_t is that on a graphical plot, if a straight line fit to it is reasonable over several periods, then, the slope of the straight line gives us an estimate of the average bias of a single-period forecast. More

advanced versions of the cumulative sum technique are based on graphical techniques that detect a certain amount of bias within a specific number of periods. Further details are available in Gardner (2006) and Chatfield (2013) among others.

3.5.3.2 Smoothed Error Tracking Signal

Trigg (1964) suggests the use of a *smoothed error tracking signal* employing the MAD, although it can be adjusted for the $\sqrt{\text{MSE}}$. This tracking signal has some nice intuitive properties, which will be discussed below. However, it also has a potential drawback in that there can be a significant lag before an actual change in the model (or its parameter values) is identified. Moreover, it is not easy to identify after the fact as to when exactly the change took place. The cumulative sum is more effective in both senses. The smoothed error tracking signal is defined by

$$T_t = z_t / \text{MAD}_t \quad (3.70)$$

where

- T_t = value of the smoothed error tracking signal at the end of period t
- MAD_t = mean absolute deviation of one-period-ahead forecast errors as of the end of period t (updated in a like manner to the MSE in Equation 3.62)
- z_t = smoothed forecast error at the end of period t

We update z in a fashion analogous to MSE. Note that the signs of the errors are retained

$$z_t = \omega(x_t - \hat{x}_{t-1,t}) + (1 - \omega)z_{t-1} \quad (3.71)$$

where ω is a smoothing constant.

For an unbiased forecasting procedure, the smoothed forecast error z_t should fluctuate about zero. Consider two consecutive forecast errors, one of which is positive, and the other is negative. We can see from Equation 3.71 that in the computation of z_t the effects of these two errors will tend to cancel one another. In contrast, in the computation of MAD_t , where the absolute value of e_t is taken, no canceling occurs. Continuing this line of reasoning, we can write that

$$-\text{MAD}_t \leq z_t \leq \text{MAD}_t$$

The limiting equalities are achieved if all errors are positive or if all errors are negative. From Equation 3.70, the equivalent limits on the tracking signal T_t are

$$-1 \leq T_t \leq 1$$

Moreover, T_t will be close to one of the limits if most of the errors are of the same sign, which is an indication that the forecasting system is biased (most of the forecasts are too high or most of the forecasts are too low).

3.5.3.3 Illustration

We illustrate both the cumulative sum and the smoothed tracking signal with the same example. Consider the item pictured in Figure 3.9 and let $t = 22$. At the end of period 21, suppose that we

have the following values:

$$\begin{aligned}\hat{x}_{21,22} &= 100 \text{ units of demand in period } t \\ \text{MAD}_{21} &= 6.25 \\ z_{21} &= 1.71\end{aligned}$$

and

$$\begin{aligned}\text{MSE}_{21} &= 57.95 \\ C_{21} &= 19.30\end{aligned}$$

Therefore, $T_{21} = 1.71/6.25 = 0.27$ and $U_{21} = 19.30/\sqrt{57.95} = 2.54$. Now, it happens that demand in period 22 is 92 units, so the forecast error is -8 . New values are computed, assuming that we use $\omega = 0.1$. For the cumulative sum, we update MSE, C , and U as follows:

$$\begin{aligned}\text{MSE}_{22} &= 0.1(92 - 100)^2 + 0.9(57.95) \\ &= 0.1(64) + 0.9(57.95) = 58.55 \text{ units} \\ C_{22} &= 19.30 - 8 = 11.30 \\ U_{22} &= 11.30/\sqrt{58.55} = 1.48\end{aligned}$$

Note that the MSE has increased due to the larger-than-usual error. The cumulative sum has decreased by 8, and the tracking signal has decreased, because the error was negative. The cumulative sums are also illustrated in Figure 3.10. (We have connected the points for clarity.) Note how they turn sharply negative due to the negative bias in the forecast errors after period 20.

Using a smoothing approach similar to that in Equations 3.62 and 3.71 is

$$\begin{aligned}\text{MAD}_{22} &= 0.1|92 - 100| + 0.9(6.25) \\ &= 0.1(8) + 0.9(6.25) = 6.43 \text{ units}\end{aligned}$$

Also Equation 3.71 gives

$$\begin{aligned}z_{22} &= 0.1(92 - 100) + 0.9(1.71) \\ &= 0.1(-8) + 0.9(1.71) = 0.74 \text{ units}\end{aligned}$$

So $T_{22} = 0.74/6.43 = 0.12$. Note again that the relatively large negative forecast error in period 22 has increased MAD slightly but has caused z to become closer to 0 because z_{21} was positive.

3.5.3.4 *Serial Correlation (or Autocorrelation) of Forecast Errors*

The serial correlation of a time series measures the degree of dependence between successive observations in the series. With an appropriate forecasting model, there should be negligible correlation between separate forecast errors. If significant positive (negative) correlation exists, then, this indicates that forecast errors of the same (opposite) sign tend to follow one another. This, in turn,

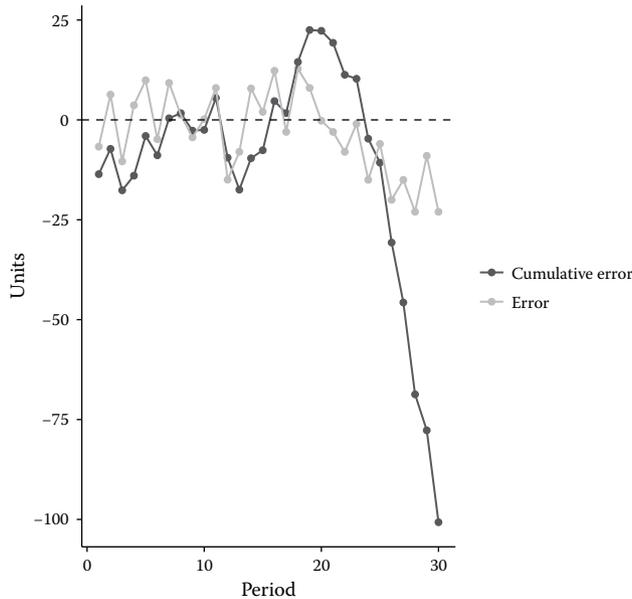


Figure 3.10 An illustration of cumulative sums with bias in the forecast errors.

signals the likelihood of errors in the specification of the demand model or its parameter values. Further details on the computation and properties of serial correlation can be found in several books on probability and statistics (see, e.g., Cox and Miller 1977).

3.5.4 Corrective Actions in Statistical Forecasting

As discussed above, there are at least three different measures of potential bias caused by incorrect specification of the underlying demand model or inaccurate estimates of the values of its parameters. We say “potential” because large absolute values of the tracking signals and substantial serial correlation can, from time to time, be simply generated from the random component in the demand pattern. In this regard, McKenzie (1978) develops statistical properties of the smoothed error tracking signal. The choice of threshold (or critical) values of the signals or serial correlation is completely analogous to the choice of the boundaries between acceptance and rejection in statistical hypothesis testing. There are two types of potential errors. First, we can take a corrective action when nothing has really changed. Second, we can do nothing when, in fact, a fundamental change has taken place. The choice of the threshold (or critical) values of the bias measures involves a, usually implicit, trade-off between these two types of errors.

There are two general types of corrective actions that can be used. The first, which tends to be utilized in a more automated mode, is to increase the values of the smoothing constants in the hope that the form of the underlying model is still correct and all that is needed is to more quickly smooth the parameter estimates closer to their true, but unknown, values. The second type of corrective action is more substantive. It typically involves human intervention to make significant changes in the parameter estimates or in the form of the demand model itself. We discuss each in turn.

3.5.4.1 *Changing the Smoothing Constants: Adaptive Forecasting*

The general idea of adaptive smoothing is that the smoothing constants are increased (i.e., smoothing becomes faster) when the tracking signal, T or U , gets too far away from 0, while smoothing is decreased as T or U returns closer to 0. We discuss one of the better known adaptive-forecasting techniques, and then reference others.

Eilon and Elmaleh (1970) develop an approach to adaptive smoothing that consists of setting a smoothing constant to a fixed value, and then making forecasts for a specified number of periods (called the evaluation interval). At the end of the evaluation interval, forecast errors are computed using the demand time series that actually occurred. Forecast errors are also computed for a range of permitted values of the smoothing constant (0.1–0.9) to see what size forecast errors would have resulted if one of these values had been used over the evaluation interval. The permitted value of the constant that would have minimized the forecast error variance over the historical evaluation interval is then used until the next evaluation interval. The procedure is repeated periodically. Alternative procedures and enhancements can be found in Chow (1965), Trigg and Leach (1967), Roberts and Reed (1969), Montgomery (1970), Whybark (1973), Flowers (1980), Gardner (1983), Fliedner et al. (1986), and Williams (1987).

Unfortunately, although all these procedures have an intuitive appeal, substantial research findings (see, e.g., Chatfield 1978; Ekern 1981; Flowers 1980; Gardner and Dannenbring 1980) suggest that adaptive methods are not necessarily better than regular, nonadaptive smoothing. In particular, there are dangers of instabilities being introduced by the attempt at adaptation. Consequently, we suggest that careful testing be undertaken with actual time series from the organization under study before any automatic adaptive procedure is adopted.

3.5.4.2 *Human Intervention*

Where sufficient bias exists or where smoothing constants remain at high levels, it may be appropriate to call for human intervention. Sufficient bias could be defined by

$$|T_t| > f$$

where f is a threshold fraction, or by

$$|C_t| > k\sqrt{\text{MSE}}$$

where k is some threshold multiple.

The choice of f or k implicitly involves the trade-off between the two types of errors mentioned above—namely (1) not intervening when such an action is necessary, and (2) intervening when it is unnecessary. Lower thresholds would be used for A items than for B items, because the first type of error is relatively more expensive for A items; for B items, we would like to keep interventions on an exception basis. In actual practice, a reasonable way to establish f or k is by trial and error in order to provide a reasonable intervention workload in an aggregate sense. In other words, if f or k is too low, managers will be overwhelmed by having to intervene so frequently. Plausible initial values are $f = 0.4$ and $k = 4$.

Significant assistance can be provided to the intervener through a display of several periods of recent actual demands and forecasts made by the most current version of the model. The individual should attempt to ascertain the reason for the recent poor performance. There may be factors, external to the model, that have changed recently, such as

1. A temporary effect—for example, a sales promotion.
2. A more permanent effect—for example, a new competitive product or a new market penetration.

A temporary effect could lead to human override of forecasts for a few periods. On the other hand, a more permanent effect could be handled by manually adjusting parameter values or even by changing the nature of the demand model. If the visual display indicates that the change occurred for several periods in the past, it may be appropriate to reinitialize the model using only data since the suspected time of the change. See Hogarth and Makridakis (1981), Ashton and Ashton (1985), Lawrence et al. (1986), and Heath and Jackson (1994). Some of these papers provide forecasting models that capture both historical demand and human expertise. In Section 3.7, we provide some guidelines for incorporating human judgment.

3.5.5 Probability Distributions of Forecast Errors

As we will see in Chapter 6, the selection of safety stocks to provide adequate customer service will depend not just on the standard deviation of forecast errors over the replenishment lead time but, instead, on the whole probability distribution of these errors. Of particular concern will be the probabilities of large positive values of the errors (demand minus forecasts) because such errors can lead to serious stockout situations.

Through most of this book, we recommend the use of a normal distribution of forecast errors for three reasons. First, empirically the normal distribution usually provides a better fit to data than most other suggested distributions. Second, particularly if the lead time (or lead time plus review interval) is long, forecast errors in many periods are added together, so that we would expect a normal distribution through the Central Limit Theorem. Finally, the normal distribution leads to analytically tractable results. Nevertheless, particularly for expensive, slow-moving items, it may be appropriate to use an alternative distribution such as the exponential, Gamma, Poisson, or negative binomial. In any event, before making a choice, one should at least perform a relatively simple test of the fit of the normal distribution to the observed forecast errors for each of a number of representative items. There is a special normal probability paper such that the cumulative distribution of a normal distribution will appear as a straight line on it. Thus, one can plot the empirical cumulative distribution and observe how closely it is fit by a straight line. In addition, many statistical software packages are available for personal computers. These packages usually contain procedures, including the goodness-of-fit (or chi-square) test and the Kolmogorov–Smirnov test, for testing the fit of data to a hypothesized distribution. See, for example, Marx and Larsen (2006).

3.6 Handling Anomalous Demand

In general, one should have a healthy suspicion of data that are purported to represent actual customer demand period by period. First, in most organizations, sales or even just shipments to customers are recorded as opposed to demand. If not properly interpreted, a series of periods in which a stockout situation existed would imply no demand in each of those periods. Furthermore, if backordering is permitted, the periods immediately following the stockout may show sales or shipments well above the actual demands in those periods. For approaches to forecasting when stockouts distort the demand data, see Wecker (1978), Bell (1981), Taylor and Thomas (1982), Lovejoy (1990), Agrawal and Smith (1996), Nahmias (1994), Lariviere and Porteus (1999), and Lau and Lau (1996a).

Furthermore, recording and transmission errors can distort the demand picture. Certain types of demand also may not be normally handled on a routine basis by the production/stocking point under consideration and thus should not be incorporated in the forecast updating. For example, unusually large customer orders may be shunted back to the higher level of supply. Orders larger than a certain size may be directly shipped from a central warehouse rather than by the usual mode of transshipment through a regional branch warehouse. Finally, there may be certain temporary changes in the demand pattern (e.g., effects of promotions and price changes) that should be filtered out prior to updating.

A useful aid in monitoring anomalous demand is to have an exception signal trigger human investigation when the “demand” stated for a particular period differs from the forecast by more than $k\sqrt{\text{MSE}}$ where a k value of 3–4 is not unreasonable. (Note that this signal is for a single demand value, whereas the thresholds in Section 3.5.4 were for a tracking signal and a cumulative sum.)

3.7 Incorporation of Human Judgment

An overall framework for forecasting was shown in Figure 3.1. In it, the input of human judgment played a key role. Such intervention is quite common in practice. For example, both Sanders and Manrodt (2003a) and Fildes and Goodwin (2007) report results from surveys of forecasters indicating that judgment plays a substantial role in sales forecasting in practice. Lawrence et al. (2006) provide an excellent discussion of research on judgmental input to forecasts, and Armstrong (2001b) is an excellent, in-depth resource for several key concepts in sales forecasting, including concepts for effectively incorporating human judgment.

In Sections 3.5.4 and 3.6, we have already seen the evidence of manual intervention in an essentially reactive fashion, that is, responding to signals of forecast bias or unusual demand levels. Now, we concentrate on the proactive use of human judgment. However, clearly there is considerable overlap in that a reactive response will often initiate a proactive input.

Obviously managerial judgment as an input is essential in medium-range, aggregate forecasting. The decisions concerning work force levels, operating hours, seasonal inventory levels, and so forth, are of major importance to an organization and thus justify the devotion of a considerable amount of management time. Moreover, there are relatively few time series to be forecasted because of the aggregation. One approach that has proved to be helpful is the Delphi method (see Basu and Schroeder (1977) for an illustration). Most of the remaining discussion will concentrate on shorter-range forecasting, but many of the ideas are also applicable in the medium range.

3.7.1 Factors Where Judgment Input Is Needed

There are a number of factors, normally not included in a statistical forecasting model, where judgment input is clearly needed. These factors can be categorized along two different dimensions: first, factors that are external to the organization versus those that are internally influenced; second, factors that have a temporary effect versus a more or less permanent effect. Our discussion will concentrate on the first dimension.

Factors external to the organization include

1. The general economic situation (e.g., the inflation rate, the cost of borrowing capital, and the unemployment rate).

2. Government regulations such as subsidies, import duties and/or quotas, pollution restrictions, safety standards, etc.
3. Competitor actions.
4. Potential labor stoppages at the customer level (users of our products).
5. Consumer preferences (particularly with regard to style goods).

The important factors internal to the organization include

1. Price changes.
2. Promotions.
3. Advertising.
4. Engineering changes that, for example, will improve the reliability of a product, thus reducing the demand for service parts.
5. Introduction of substitute products. A classic example was the effect on the sales of slide rules that resulted from the introduction of hand calculators.
6. Opening a new distribution outlet for an existing product.
7. The pipeline-filling effect associated with the introduction of a new product. A temporary surge of demand will occur simply to initially place appropriate levels of the inventory at the various stocking points.

Related to internal factors is the need to ensure that the total (in dollars) of statistical forecasts of individual items agrees reasonably well with the aggregate financial plan of the organization.

Some success has been achieved in introducing some of the above judgmental factors into a form of the statistical forecasting model known as Bayesian forecasting. The general approach is discussed by Harrison and Stevens (1971), Harrison and Stevens (1976), Johnston (1980), Fildes (1983), Azoury (1985), Kaplan (1988), Bradford and Sugrue (1991), Karmarkar (1994), and Eppen and Iyer (1997).

Fildes and Goodwin (2007) discuss four case studies and go on to suggest practices for maximizing the value of judgmental input. Their recommendations, with which we wholeheartedly agree, include limiting adjustments to statistical forecasts, requiring the forecaster to document their reasons and assumptions when an adjustment is made, and tracking the impact of behavioral adjustments. Fildes et al. (2009) provide an example of the insights from tracking behavioral adjustments in practice. Their work shows that small adjustments to forecasts tend to hurt forecasting performance while large adjustments improve the performance. This is primarily due to the fact that large adjustments are made only when a manager has significant and relevant information.

3.7.2 Guidelines for the Input and Monitoring of Judgment

It is advisable to think in terms of a small committee being responsible for the judgmental input. Ideally, the membership should include an individual with a broad perspective (e.g., from strategic planning), as well as representatives from marketing, production/inventory management, and engineering/design. Meetings should be held on the order of once per month.

This committee is likely to be concerned with groups or families of items, perhaps by geographic region. The starting point in each case should be the objective (statistical) forecast, aggregated across the appropriate group of items. In addition, recent demand history and the associated forecasts should be available. The committee should be able to easily adjust the aggregate forecast. Then, a computer routine should simply prorate a family adjustment to the individual-item forecasts according to the recent fractions of family demand represented by the individual items.

It is often helpful to ensure that the committee develop and record their forecasts individually before talking as a group. Otherwise, the more influential members of the committee might sway the others. The committee can then take the average and variance of the independent forecasts. Further discussion on this topic can be found in Fisher et al. (1994) and Fisher and Ittner (1999). Gaur et al. (2007) show how judgmental forecasts can be effectively used to develop estimates for the mean and variance of demand.

It is important to provide feedback on the performance of the judgmental input. This is accomplished by retaining both the original statistical forecast and the revised (through judgment) forecast. When the actual demand materializes, one can then show the statistical forecast error and the error of the judgmental forecast. In fact, a report can be developed listing the difference of the two errors (likely normalized by dividing by $\sqrt{\text{MSE}}$) for each family of items forecasted. The listing can be ranked by relative performance of the two methods. The intention should not be to externally evaluate the human input. Rather, the objective should be to provide a mechanism for the committee members themselves to learn from their earlier successes and failures.

The above discussion focuses on trying to maximize the value of judgment in improving forecast accuracy. In many instances, there are other motives that may lead to judgmental influence on the forecast. Galbraith and Merrill (1996) describe a survey of executives that examines why forecasts are changed. They find that some executives change the forecast to influence the firm's stock price, while others are attempting to influence internal resource allocations. In some cases, of course, they are simply trying to improve the forecast with better information. This is an interesting article with good warnings for users and developers of forecasts. Oliva and Watson (2011) discuss a case study for a high-tech manufacturer where the organization structure and forecasting process led to poor forecasting performance and poor resource allocation decisions based on the forecasts. Oliva and Watson (2009) further document how such poor alignment can be addressed by creating a common information technology (IT) system that contains all relevant information such as sales, forecasts, and assumptions as well as adopting what they call a consensus-forecasting process where a series of meetings are held with relevant stakeholders in the organization to come to an agreement on the forecast. In their case, and in most cases, the relevant stakeholders include representatives from operations planning, marketing, and finance. Such a process is commonly a part of what is known as the Sales and Operations Planning (S&OP) process. The S&OP process is discussed further in Section 12.2.

3.8 Dealing with Special Classes of Individual Items

The procedures discussed to this stage in the chapter should be of use for a major portion of the items handled by a typical organization. However, there are special classes of items where caution must be exercised. The intention of this section is to provide some insights concerning each of the several special classes. Additional information about forecasting style goods items will be covered in Chapter 9.

3.8.1 *Items with Limited History*

Here, we finally address the issue of initialization of the exponential smoothing procedures of Section 3.4, when there is insufficient historical data to use the methods suggested earlier for obtaining estimates of the parameters a , b , the F s, and MSE. There are two reasons for an item having insufficient historical data. First, the item may be a relatively new product (i.e., early in its product

life cycle), as discussed in Section 2.5.1. Second, when a new control system is introduced in an organization, there may simply not be enough historical records even though the item has had sales for a considerable period of time.

For an item with insufficient history, a knowledgeable individual (likely a member of the judgmental input committee) should estimate

1. Total demand in the first year
2. Total demand in the second year
3. Seasonal pattern (if appropriate)

The first two give an estimate of the level and trend. The third most likely would be provided by saying that the item's seasonal behavior should be the same as an existing individual item or group of items.

It is important to provide a graph of the implied 2-year demand pattern so that adjustments can be made, if necessary. Higher-than-normal smoothing should be used in the first several periods because of the relative inaccuracy of the initial estimates of the parameters.

The MSE (or σ) of an item with insufficient history is estimated using an empirical relationship between σ and the demand level, a , found to hold for more established items in the same inventory population. In particular, a relationship that has been observed to give a reasonable fit for many organizations is of the form

$$\sigma_1 = c_1 a^{c_2} \quad (3.72)$$

or equivalently

$$\log \sigma_1 = \log c_1 + c_2 \log a \quad (3.73)$$

where c_1 and c_2 are regression coefficients typically with $0.5 < c_2 < 1$. The relationship of Equation 3.73 will plot as a straight line on log-log graph paper.

Sometimes, Equation 3.72 is modified to estimate the standard deviation on a dollar basis

$$\sigma_1 v = c_1 (av)^{c_2} \quad (3.74)$$

where v is the unit variable cost of an item.

This relationship can be used on an on-going basis as needed to estimate the standard deviation of any C items because these items may not be statistically forecasted, and therefore their errors would not be recorded. Relationships other than Equation 3.72 have been suggested in the literature. See, for example, Hausman and Kirby (1970), Stevens (1974), Shore (1995, 1998), and Kurawarwala and Matsuo (1996). The papers by Shore provide a simple procedure for approximating the distribution when data are extremely limited.

3.8.1.1 Illustration

Consider a new item with an estimated demand of 300 units per year. The person responsible for forecasting this item used several hundred other items similar to this one to find c_1 and c_2 by regression, with the following result:

$$\sigma_1 = 1.04(D/12)^{0.65}$$

where D is the annual demand. Therefore, the initial estimate of σ_1 is

$$\sigma_1 = 1.04(300/12)^{0.65} = 8.43 \text{ units}$$

and the initial estimate of MSE is

$$\text{MSE} = (8.43)^2 = 71.02$$

3.8.2 Intermittent and Erratic Demand

The exponential smoothing forecast methods described earlier have been found to be ineffective where transactions (not necessarily unit sized) occur on a somewhat infrequent basis. To illustrate this point, let us consider a particularly simple situation where only transactions of a single size j occur exactly every n periods of time. Suppose that simple exponential smoothing (level only) is used and the updating is done every unit period. Note that immediately after a demand transaction occurs, the forecast will exceed the average demand per period, because of the weight given to the last demand of size j . Because there is zero demand for the next $n - 1$ periods, the forecast decreases thereafter, dropping below the average demand by the end of the cycle. In other words, for a deterministic demand pattern, exponential smoothing produces a sawtooth-type forecast.

This illustration is a symptom of an undesirable performance of exponential smoothing under the more general situation of infrequent transactions, not necessarily all of the same size. For such a situation, it is preferable to forecast two separate components of the demand process—namely, the time between consecutive transactions and the magnitude of individual transactions. That is, we let

$$x_t = y_t z_t \tag{3.75}$$

where

$$y_t = \begin{cases} 1 & \text{if a transaction occurs} \\ 0 & \text{otherwise} \end{cases}$$

and z_t is the size of the transaction.

If we define a quantity n as the number of periods between transactions and if demands in separate periods are considered to be independent, then, the occurrence or nonoccurrence of a transaction in a period can be considered as a Bernoulli process with the probability of occurrence being $1/n$. That is

$$\Pr\{y_t = 1\} = 1/n$$

and

$$\Pr\{y_t = 0\} = 1 - 1/n$$

Within this framework and assuming that transaction sizes are normally distributed, Croston (1972) recommends the following reasonable updating procedure. If $x_t = 0$ (i.e., no demand occurs),

1. Transaction size estimates are not updated, that is, $\hat{z}_t = \hat{z}_{t-1}$
2. $\hat{n}_t = \hat{n}_{t-1}$

If $x_t > 0$ (i.e., a transaction occurs),

1. $\hat{z}_t = \alpha x_t + (1 - \alpha)\hat{z}_{t-1}$
2. $\hat{n}_t = \alpha n_t + (1 - \alpha)\hat{n}_{t-1}$

where

n_t = number of periods since the last transaction

\hat{n}_t = estimated value of n at the end of period t

\hat{z}_t = estimate, at the end of period t , of the average transaction size

Croston argues that forecasts for replenishment purposes will usually be needed immediately after a transaction. He shows that the forecast at that time, namely,

$$\hat{z}_t / \hat{n}_t$$

is preferable to that obtained by simple exponential smoothing for two reasons:

1. It is unbiased, whereas the simple smoothing forecast is not.
2. It has a lower variance than does the simple smoothing forecast.

However, Croston warns that the infrequent updating (only when a transaction occurs) introduces a marked lag in responding to actual changes in the underlying parameters. Therefore, he rightfully stresses the importance of control signals to identify deviations.

A key quantity to use in establishing the safety stock for an intermittently demanded item is $MSE(z)$, the mean square error of the sizes of *nonzero*-sized transactions. This quantity is only updated each time a transaction occurs. The updating equation is of the form

$$\text{New MSE}(z) = \omega(x_t - \hat{z}_{t-1})^2 + (1 - \omega)\text{Old MSE}(z) \quad x_t > 0 \quad (3.76)$$

In establishing the safety stock, one must also take into account of the fact that there is a nonzero chance of no transaction occurring in a period. Willemain et al. (1994) show that Croston's method is superior to exponential smoothing, and also note that the lognormal distribution is preferable to the normal for transaction sizes. Further discussion on Croston's method can be found in Hyndman et al. (2008). Johnston (1980) suggests an alternate method for directly estimating σ_1 for an intermittent demand item. See also Wright (1986), Johnston (1993), Johnston and Boylan (1994, 1996), Anderson (1994), Walton (1994), and Syntetos and Boylan (2005).

An item is said to have an erratic demand pattern if the variability is large relative to the mean. Brown (1977) operationalizes this by saying that demand is erratic when $\hat{\sigma}_1 > \hat{a}$ where $\hat{\sigma}_1$ is obtained by a fit to history, as in Section 3.5.1. In such a case, he suggests restricting the underlying model to the case of only a level and a random component. See also Muir (1980), Watson (1987), and Willemain (2004).

3.8.3 Replacement or Service Parts

In the case of replacement or service parts, we have more information than is usually available in a demand-forecasting situation. In particular, the requirements for replacement parts depend on the failure rate of operating equipment, which, in turn, depends on the time stream of the introduction of new (or repaired) pieces of equipment as well as the distribution of the service life of the equipment. Although it is relatively easy to state the above relationship, it is quite another thing to operationalize it. In particular, the mathematics become extremely involved. Also, it is neither easy to maintain the required records nor to estimate the service-life distribution. Keilson and Kubat (1984) develop tractable results for the special case where new equipment is introduced

according to a time-varying Poisson distribution and failures also occur according to a Poisson process. A closely related problem is the forecasting of the returns of rented items or containers (e.g., soft drink bottles, beer kegs, etc.). See Goh and Varaprasad (1986), Kelle and Silver (1989), Fleischmann et al. (1997), Guide and Wassenhove (2001), and the references in Section 11.5.

3.8.4 Terminal Demand

Here, we are dealing with exactly the opposite situation from that in Section 3.8.1. Now, we are nearing the end of the life cycle of the item. It has been empirically observed (see, e.g., Brown et al. 1981; Fortuin 1980) that in the terminal phase of demand, the pattern tends to decrease geometrically with time. That is, the demand in period $t + 1$ is a constant fraction (f) of the demand in period t . Mathematically, we then have

$$x_{t+1} = fx_t$$

or, more generally,

$$x_t = f^t x_0 \quad t = 0, 1, 2, \dots \quad (3.77)$$

In this last equation, the time origin has been placed at the beginning of, or somewhere during, the terminal phase of the pattern.

Taking the logarithms of Equation 3.77 reveals that

$$\log x_t = t \log f + \log x_0 \quad (3.78)$$

Thus, a plot of x_t versus t on a semilogarithmic graph should produce a straight line with slope $\log f$. Such a line can be fit by standard regression techniques. Brown (1981) develops an expression for the variance of an extrapolated forecast into the future as a function of the residual variance of the historical fit, the number of historical data points used in the fit, and how far into the future the projection is made.

If the origin is placed at the current period, and the current level is x_0 , then, an estimate of the all-time future requirements using Equation 3.77 is given by

$$\begin{aligned} ATR &= \sum_{\tau=1}^{\infty} \hat{x}_{0,\tau} \\ &= \sum_{\tau=1}^{\infty} f^{\tau} x_0 = \frac{f}{1-f} x_0 \end{aligned} \quad (3.79)$$

3.8.4.1 Illustration

Consider a particular spare part for a late-model car. Suppose that the demand has begun dropping off geometrically with a factor $f = 0.81$ and the demand level in the current year is 115 units. From Equation 3.79, the all-time future requirements are estimated to be

$$ATR = \frac{0.81}{0.19} (115) \approx 490 \text{ units}$$

For other information on this topic, including a discussion of the sizing of replenishment batches, see Hill et al. (1989), and Section 8.5.

3.9 Assessing Forecasting Procedures: Tactics and Strategy

In this section, we first address a tactical question—namely, measuring and comparing the statistical accuracy of different forecasting procedures. Then, we return to the issues of a more strategic nature.

3.9.1 *Statistical Accuracy of Forecasts*

The first point to note is that there are several possible measures of forecast accuracy—for example, the MSE, the MAD, the sample variance, the sample bias, any of these normalized by the mean demand level, and so forth. No single measure is universally best. In fact, forecast accuracy is only a surrogate for overall production/inventory system performance (encompassing the uses of the forecasts). The surrogate is usually justified in terms of keeping the evaluation procedure of a manageable size. We return to this point in Section 3.9.2.

Besides the choice of several possible statistical measures of forecast accuracy, there is the added problem of not even being able to make valid statistical claims about the relative sizes of the same measure achieved by two different forecasting procedures. The reason for this is that the two series of forecast errors are correlated to an unknown extent. (See Peterson 1969.)

To complicate matters, it can be shown that a linear combination of forecasts from two or more procedures can outperform all of those individual procedures. Some research has even shown that attempts to find optimal weights when combining different forecasts are not worth the effort. A simple average provides very good results. (See Kang 1986.) In theory, the best combination of procedures depends on the value of the measure of variability (e.g., the MSE) for each individual procedure, as well as the correlation between the results of the procedures. However, it is difficult to accurately estimate these quantities. Hence, if combinations of forecasts are to be considered, we recommend trying combinations of only two at a time, and starting with equal weights. References on combinations of forecasts include Bates and Granger (1969), Dickinson (1975), Lewis (1978), Makridakis and Winkler (1983), Armstrong (1984, 2001a), Ashton and Ashton (1985), Flores and Whybark (1986), Georgoff and Murdick (1986), Lawrence et al. (1986), Gupta and Wilton (1987), Bunn (1988), Sanders and Ritzman (1989), Schmittlein et al. (1990), and Timmermann (2006).

In developing the relative measures of accuracy, it is crucial to use an *ex-ante*, as opposed to an *ex-post*, testing procedure. In an *ex-post* approach, the forecast model and updating procedure are tested on the same data from which the model form or smoothing constant values were selected. In contrast, the *ex-ante* method involves testing with an entirely different set of data. This is usually accomplished by using a first part of the historical information for fitting purposes, and the remaining portion for testing. Performance on *ex-post* tests can be a very poor indicator of *ex-ante* performance, and it is the latter that is of practical interest. See Fildes and Howell (1979), and Makridakis (1990) for further discussion and references.

Numerous tests that have measured the statistical accuracy of different forecasting procedures have been reported in the literature. However, of primary interest is the comprehensive study undertaken by Makridakis et al. (1982), subsequently followed by Makridakis et al. (1993) and Makridakis and Hibon (2000). It involved some 1,001 time series including both micro- and macrolevel data as well as monthly, quarterly, and annual forecast update intervals. All the common statistical forecasting procedures (from very simple to quite complex) were tested, and a number of measures of statistical accuracy were obtained for each combination of procedure and time series. Some of the salient findings were the following:

1. Significant differences in forecast accuracy occurred as a function of the micro-/macroclassification. On microdata (the case in individual-item forecasting), the relatively simple methods did much better than the statistically sophisticated methods. The latter were better, however, for macrodata (the case for medium or even longer-term forecasting).
2. For horizons of 1–6 periods ahead, simple exponential smoothing, Holt and Brown trend procedures, and the Winters method performed well. For horizons of 7 or more periods, a more complicated procedure, developed by Lewandowski (1979), showed the best performance.
3. Deseasonalization of the data by the ratio to moving average method (which we suggested in Section 3.4.4) did as well as much more sophisticated deseasonalization procedures.
4. Adaptive methods did not outperform nonadaptive methods.
5. A simple average of six of the simpler statistical methods did somewhat better than any of the individual methods, and also outperformed a combination of the methods that in theory was supposed to give the best results.
6. In general, sophisticated methods did not produce more accurate results than did the simpler methods.

In summary, the results of this major international study lend strong support to the forecasting procedures that we have advocated in this chapter. See also Fildes (1989) and Armstrong (2001b).

3.9.2 Some Issues of a More Strategic Nature

As Fildes (1979) points out, two questions dominate any assessment of a forecasting procedure. First, are the results statistically satisfactory (the topic covered in Section 3.9.1)? Second, will the procedure, once developed, be used and perform in a cost-effective fashion? In Fildes's words, "it is the latter that is often of more concern to the practicing forecaster." This view is shared by the authors of the Study Guide for the Certification Program of the American Production and Inventory Control Society (see Blagg et al. 1980):

Forecasting is not done for its own sake; it is meaningful only as it relates to and supports decision making within the production system. Hence, managerial considerations regarding cost, effectiveness, and appropriateness of the forecasting function must be a part of the domain of forecasting.

Earlier, we mentioned that it has often been advocated that the statistical accuracy of a forecasting method should be used as a surrogate for the portion of system costs associated with forecast errors. Unfortunately, studies have shown that traditional error measures do not necessarily relate to overall system costs. Thus, where possible, a choice between forecasting procedures should be based partly on a comparative simulation test of system costs using historical data or a trial run on new data.

Closely related to this point is the desirable property of robustness (Fildes 1979). The choice of a particular demand model and updating procedure should again be at least partially based on an insensitivity to the data quality, missing observations, outliers, and so forth. We hope that the suggestions made in Section 3.6 would benefit any selected procedure from a standpoint of robustness.

A study by Adam and Ebert (1976) (see also Hogarth and Makridakis 1981) compares objective (statistical model) forecasts with judgmental forecasts. Interestingly enough, the objective forecasts,

on the average, outperformed the subjective forecasts. However, Bunn and Wright (1991) note that expert, informed judgmental forecasts have higher quality than many researchers have previously asserted. Therefore, it is essential to solicit human input to the forecasting process. To have meaningful subjective input, the manager must be able to understand the underlying assumptions and general idea of the statistical model. This is an important reason for why we have placed so much emphasis on the relatively simple exponential smoothing procedures. Moreover, in Chapters 10 and 15, where there will be a more explicit dependence of demand among different SKUs (multi-echelon inventory systems and multistage production processes), we will emphasize a different type of forecasting that involves the explosion of known future usage.

Related to the usefulness of a forecasting system, Makridakis and Wheelwright (1979) list a number of potential organizational blocks to improved forecasting. Of specific interest here are (1) bias and (2) credibility and communication. By bias, we mean that in many organizations, there is incentive for a forecast to not represent the most likely outcome but, instead, to serve as a self-fulfilling goal. For example, sales personnel may deliberately provide low forecasts if they are rewarded for exceeding the target levels based, at least in part, on the forecasted values. By credibility, we mean that the forecasts may lack relevance, or appear to lack relevance, which often is a result of a lack of proper communication between the forecast analyst and the decision maker. Additional information is provided by Sanders and Manrodt (1994, 2003b) who find that the lack of relevant data and low organizational support are the primary reasons that formal methods are not used as extensively in U.S. corporations.

In conclusion, we recall the experience of one of the authors who was once asked to mediate between competing forecasting groups at a large international corporation. Each group was obviously very sophisticated in mathematical time-series analysis and very clever in coaxing the most obscure of patterns out of historical data through spectral analysis. However, none of the nine highly trained mathematician forecasters involved had ever really closely examined a retail outlet or wholesaler through which the products that they forecasted were being sold. To most of them, products were characterized by amplitudes and phase angles.

Top managers for their part were somewhat intimidated by all the apparent sophistication (as was the author initially). Little collaboration existed between the users and the producers of forecasts. Both management and analysts ignored the fact that one really cannot foretell the future with great accuracy. Tactics without any overall strategy prevailed. On closer examination, it became clear that the management was not taking the mathematicians' forecasts seriously because they did not understand how they were being generated. The problem that brought matters to a head was the entrance of an aggressive competitor into the market who claimed through massive advertising that the company's basic product concept was technologically obsolete. It took for top management several weeks to react because they were unable to determine the extent of the impact on their sales until wholesalers and retailers started to refuse previously planned shipments.

Eventually the situation was resolved as follows. Instead of trying to foretell the future, top management agreed to set up better procedures that would allow the company to quickly react to forecast errors as soon as their magnitude was determined. This involved a basic change in management philosophy. No longer were forecast errors a surprise event. They were fully anticipated, and everyone in the decision hierarchy prepared contingency plans in response. All agreed that it was at least equally as important to develop techniques for responding to errors, as it was to try to measure and predict them. In the process, six of the high-powered time-series analysts were laid off, resulting in a substantial reduction of overall costs associated with forecasting as well as in a better forecasting strategy.

Problems

- 3.1 Suppose that the following were sales of beer in the Kitchener-Waterloo area by month for the 3 years shown:

	<i>Jan.</i>	<i>Feb.</i>	<i>Mar.</i>	<i>Apr.</i>	<i>May</i>	<i>June</i>	<i>July</i>	<i>Aug.</i>	<i>Sept.</i>	<i>Oct.</i>	<i>Nov.</i>	<i>Dec.</i>
2012	73	69	68	64	65	98	114	122	74	56	72	153
2013	80	81	83	69	91	140	152	170	97	122	78	177
2014	95	66	81	100	93	116	195	194	101	197	80	189

- Plot the data roughly to scale.
 - Is there a trend present?
 - Is there any seasonality? Discuss.
 - Are there any anomalies? Discuss.
 - What is your forecast of demand for January 2015? For June 2015?
- 3.2 Southern Consolidated manufactures gas turbines. Some of the parts are bought from outside, while the important ones are made in the company's plant. CZ-43 is an important part that wears out in approximately 2 years and has to be replaced. This means CZ-43 has to be manufactured not only to meet demand at the assembly line but also to satisfy its demand as a spare part. The gas turbine was first sold in 2002 and the demand for CZ-43 for 2002–2015 is given below. Forecast the total demand for CZ-43 for 2016.

<i>Year</i>	<i>Spare Part</i>	<i>Assembly</i>	<i>Demand</i>
2002	3	87	90
2003	6	48	54
2004	14	43	57
2005	30	31	61
2006	41	49	90
2007	46	31	77
2008	42	37	79
2009	60	42	102
2010	58	47	105
2011	67	32	99
2012	70	25	95
2013	58	29	87
2014	57	30	87
2015	85	23	108

3.3 Under the assumption of a level model, that is,

$$x_t = a + \epsilon_t$$

show that the choice of a to minimize

$$S = \sum_{j=t-N+1}^t (x_j - a_t)^2$$

where x_j is the actual demand in period j , is given by

$$\hat{a}_t = (x_t + x_{t-1} + \dots + x_{t-N+1})/N$$

- 3.4 Prove that the average age of the data used in the estimate of Equation 3.25 is given by $1/\alpha$.
 3.5 The number of violent crimes in each of the last 6 weeks in the Hill Road Precinct has been:

Week j	1	2	3	4	5	6
Incidents x_j	83	106	95	91	110	108

- a. Use a 2-week moving average to forecast the incidents in each of weeks 7 and 10. Also compute the MSE based on the incidents and forecasts for periods 1–6. To initialize, assume $x_{-1} = x_0 = 100$.
 b. Detective Bick Melker suggests the use of simple (single) exponential smoothing with $\alpha = 0.1$. Use such a procedure to develop forecasts of incidents in weeks 7 and 10 for Hill Road. Initialize with $\hat{a}_0 = 100$. Also compute the MSE and MAD.
 c. Without redoing all the computations, what would be the forecast of incidents in week 7 using exponential smoothing if \hat{a}_0 were 90 instead of 100?
 3.6 Brown’s updating equations, developed from discounted least squares, for the case of a trend model are actually given by

$$\hat{a}_t = x_t + (1 - \alpha)^2(\hat{a}_{t-1} + \hat{b}_{t-1} - x_t)$$

and

$$\hat{b}_t = \hat{b}_{t-1} - \alpha^2(\hat{a}_{t-1} + \hat{b}_{t-1} - x_t)$$

Show that these two equations are equivalent to Equations 3.32 and 3.33.

- 3.7 The demand for a particular model of tablet at a particular outlet of Great Buy has had the following pattern during the past 15 weeks:

Week t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Demand x_t	34	40	34	36	40	42	46	50	55	49	49	53	63	69	65

A systems analyst at Great Buy believes that demand should be trending upward; thus, she wishes to evaluate the use of exponential smoothing with a trend model.

- a. Use the first 5 weeks of data in Equations 3.36 and 3.37 to obtain estimates of \hat{a} and \hat{b} as of the end of week 5.
 - b. Using an α value of 0.15, employ exponential smoothing to update \hat{a} and \hat{b} and show the forecasts for weeks 6, 7, . . . , 16 made at the start of each of these periods. Also, what is the forecast, made at the end of week 15, of the demand in week 20?
- 3.8 Union Gas reports the following (normalized) figures for gas consumption by households in a metropolitan area of Ontario:

	Jan.	Feb.	Mar.	Apr.	May	June	July	Aug.	Sept.	Oct.	Nov.	Dec.
2012	196	196	173	105	75	39	13	20	37	73	108	191
2013	227	217	197	110	88	51	27	23	41	90	107	172
2014	216	218	193	120	99	59	33	37	59	95	128	201
2015	228											

- a. Using a simple (single) exponential smoothing model (with $\alpha = 0.10$), determine the forecast for consumption in (1) February 2015 and (2) June 2016.
 - b. Do the same using the Winters seasonal smoothing model with $\alpha_{HW} = 0.20$, $\beta_{HW} = 0.05$, and $\gamma_{HW} = 0.10$. Use the first 2 years of data for initialization purposes.
 - c. Plot the demand data and the two sets of forecasts roughly to scale. Does there appear to be a point of abrupt change in the demand pattern? Discuss your results briefly.
- 3.9 For purposes of calculating the safety stock of a particular item, we are interested in having an estimate of the MAD or MSE of forecast errors. No historical forecasts are available, but historical monthly demand data for 3 years are available. An analyst on your staff has proposed using the MAD with the following procedure for estimating it:

Step 1 Plot the 3 years of data.

Step 2 Fit the best straight line to it (either by eye or through a statistical fit).

Step 3 For each month find the absolute deviation

$$|\epsilon_t| = |x_t - \tilde{x}_t|$$

where x_t is the actual demand in month t and \tilde{x}_t is the value read off the straight line (found in Step 2) at month t .

Step 4 MAD = average value of $|\epsilon_t|$ over the 36 months.

- a. Discuss why the MAD found by the above procedure might be lower than that actually achievable by statistical forecasting.
 - b. Discuss why the MAD found by the above procedure might be substantially higher than that actually achievable by a forecasting procedure. Hint: The use of diagrams may be helpful.
 - c. Why might the MSE be a better choice?
- 3.10 Demand for a new breakfast cereal at a medium-sized grocery store is given in the table below:

Month	Demand	Month	Demand	Month	Demand
1	165	9	357	17	585
2	201	10	410	18	702
3	239	11	603	19	775
4	216	12	449	20	652
5	251	13	479	21	586
6	269	14	569	22	760
7	299	15	623	23	788
8	332	16	591	24	817

- a. Using a simple (single) exponential smoothing model (with $\alpha = 0.15$), determine the forecast for consumption in (1) month 25 and (2) month 30.
 - b. Do the same using the trend model with $\alpha_{HW} = 0.20$ and $\beta_{HW} = 0.05$. Use the first year of data for initialization purposes.
 - c. Plot the demand data and the two sets of forecasts roughly to scale. Discuss your results briefly.
- 3.11 In current times, give an illustration of an item fitting into each of the following special classes not using the specific examples in the book (use your everyday experiences at work or elsewhere):
- a. Limited history
 - b. Intermittent demand
 - c. Erratic demand
 - d. Replacement parts
 - e. Terminal demand
- 3.12 Suppose that the demand rate for an item could be modeled as a continuous linearly decreasing function of time

$$x_t = a - bt \quad a > 0, \quad b > 0$$

- where $t = 0$ is the present time. Compute an estimate of the all-time future requirements.
- 3.13 Suppose you have been hired by a company to install a forecasting system for about 5,000 B items. You suggest the use of exponential smoothing and the manager to whom you report is not happy about this. He says “Exponential smoothing is nothing but a massaging of historical data. It takes no account of our pricing strategy (e.g., special sales at certain times of the year), our competitor’s actions, etc.”
- a. How would you respond to this criticism?
 - b. If the manager insisted on having pricing effects built into the forecasting model, briefly outline how you would proceed to do this, including what data you would attempt to obtain.
- 3.14 Develop a spreadsheet that computes the simple 5-period moving average and uses it for forecasting. Include the computation of the forecast error.

- 3.15 Develop a spreadsheet that performs the simple level exponential smoothing operation and uses it for forecasting. Include the computation of the forecast error.
- 3.16 Develop a spreadsheet that performs exponential smoothing with trend and uses it for forecasting. Include the computation of the forecast error.
- 3.17 Develop a spreadsheet that performs exponential smoothing with seasonality and uses it for forecasting. Include the computation of the forecast error.
- 3.18 For a local business, find several reasons why demand in the next few periods might be very different from demand in the recent past. How should the forecast be modified?
- 3.19 For a local business, find several factors external to the firm that might influence the forecast of demand. How should the forecast be modified?
- 3.20 A summer student working at a small local firm has been assigned the task of forecasting. The president tells her that for some products, there is not enough relevant historical demand to use any of the traditional forecasting techniques described in this chapter. What are the possible reasons the president could give? What should the summer student do? If the student wanted to use historical demand of a related item to forecast one of the items that does not have enough relevant history, what factors should she consider?
- 3.21 One of the major advantages of the new computer-aided manufacturing technology is that it permits a shorter design and manufacturing lead time, that is, a faster market response. Describe the impact of the faster market response on (i) forecasting requirements and (ii) the required inventory levels.
- 3.22 A common way of testing forecasting procedures consists of monitoring how a particular procedure responds to a step function in demand or a one period transient change in demand level. (Needless to say, the step and transient functions are only two of the many functions that can be used.)

<i>Time</i>	<i>Step Function</i>	<i>Transient Change</i>
1	100	100
2	100	100
3	150	150
4	150	100
5	150	100
6	150	100
7	150	100
8	150	100
9	150	100
10	150	100

- a. Use both a simple and a trend exponential smoothing model (trying both $\alpha = 0.1$ and $\alpha = 0.3$) to track both of the above demand patterns. For initial conditions, assume at the start of time 1 (end of period 0) that $\hat{a} = 100, \hat{b} = 0$.
- b. Plot the corresponding forecast errors under each procedure.

- 3.23 Demand for a given product has been tracked for 5 years, and the results are shown below:
- Use the first 4 years of demand to establish a simple (single) exponential smoothing model (with $\alpha = 0.10$). At the end of month 48, what is your forecast for month 49? Month 60?
 - Employ exponential smoothing with trend (with $\alpha_{HW} = 0.15$ and $\beta_{HW} = 0.05$) to update \hat{a} and \hat{b} and show the forecasts for weeks 49, 50, . . . , 60 made at the start of each of these periods.
 - Plot the demand data and the forecasts roughly to scale. Does there appear to be a point of abrupt change in the demand pattern? Discuss your results briefly.

<i>Month</i>	<i>Demand</i>	<i>Month</i>	<i>Demand</i>	<i>Month</i>	<i>Demand</i>
1	239	21	261	41	165
2	186	22	223	42	119
3	164	23	368	43	176
4	179	24	231	44	177
5	229	25	219	45	200
6	219	26	142	46	173
7	259	27	179	47	260
8	208	28	188	48	192
9	201	29	237	49	136
10	186	30	126	50	184
11	166	31	237	51	239
12	196	32	234	52	219
13	191	33	300	53	174
14	229	34	174	54	208
15	169	35	187	55	241
16	254	36	154	56	220
17	155	37	211	57	194
18	158	38	129	58	197
19	210	39	207	59	218
20	143	40	175	60	252

- 3.24 A company sells a group of items having a seasonal demand pattern. It is reasonable to assume that all items in the group have the same seasonal pattern. The forecasting procedure currently in use is as follows: From historical data on the whole group, reasonably

accurate estimates are available for the fraction of annual demand that occurs in each of the 12 months. Let us call these F_1, F_2, \dots, F_{12} and let

$$C_k = \sum_{j=1}^k F_j$$

be the cumulative fraction of annual demand historically experienced through month $k, k = 1, 2, \dots, 12$.

Consider a particular item i and suppose that its actual demand in month j is D_{ij} . At the end of month k , the annual demand projection for item i , denoted by P_{ik} , is given by

$$P_{ik} = \frac{\sum_{j=1}^k D_{ij}}{C_k}$$

and the forecast for a single future month m in the same year is

$$P_{ikm} = p_{ik} F_m$$

- a. Illustrate the above forecasting method with annual rate projections and 1-month-ahead forecasts made at the end of each of Jan., Feb., . . . , Dec. and Jan. for an item having the following monthly demands in a particular 13-month period:

	<i>Jan</i>	<i>Feb</i>										<i>Dec</i>	<i>Jan</i>
Month, j	1	2	3	4	5	6	7	8	9	10	11	12	13
D_j	20	20	81	87	106	215	147	125	84	128	122	28	77

The historically determined F'_j 's for the associated group of items are

<i>Month, j</i>	1	2	3	4	5	6	7	8	9	10	11	12	13
F_j	0.03	0.04	0.06	0.07	0.09	0.16	0.14	0.12	0.11	0.09	0.06	0.03	

- b. The inventory manager claims that the method is very poor in the early months of a year, but gets better and better as one approaches the end of the year. Does this make sense? If so, how would you suggest modifying the method so as to improve matters early in the year? (There is no need to do a lot of numerical work here.)

3.25 Monthly demand at a local auto parts store for automobile batteries is shown in the table below:

Month	Demand	Month	Demand	Month	Demand
1	34	9	38	17	58
2	44	10	44	18	54
3	42	11	36	19	43
4	30	12	46	20	48
5	46	13	42	21	40
6	44	14	30	22	46
7	56	15	52	23	52
8	50	16	48	24	54

- Using a simple (single) exponential smoothing model (with $\alpha = 0.15$), determine the forecast for consumption in (1) month 25 and (2) month 30.
- Do the same using the Winters seasonal smoothing model with $\alpha_{HW} = 0.20$, $\beta_{HW} = 0.05$, and $\gamma_{HW} = 0.10$. Use the first year of data for initialization purposes.
- Plot the demand data and the two sets of forecasts roughly to scale. Discuss your results briefly.

Appendix 3A: Derivations

3A.1 Derivation of Simple Exponential Smoothing

From Equation 3.22, we know that we wish to select \hat{a}_t in order to minimize

$$S' = \sum_{j=0}^{\infty} d^{j+1} (x_{t-j} - \hat{a}_t)^2$$

By setting $dS'/d\hat{a}_t = 0$, we obtain

$$-2d \sum_{j=0}^{\infty} d^j (x_{t-j} - \hat{a}_t) = 0$$

or

$$\hat{a}_t \sum_{j=0}^{\infty} d^j = \sum_{j=0}^{\infty} d^j x_{t-j} \quad (3A.1)$$

Now

$$\sum_{j=0}^{\infty} d^j = \frac{1}{1-d}$$

Therefore, Equation 3A.1 becomes

$$\hat{a}_t = (1 - d) \sum_{j=0}^{\infty} d^j x_{t-j} \tag{3A.2}$$

$$\hat{a}_t = (1 - d) \left(x_t + \sum_{j=1}^{\infty} d^j x_{t-j} \right)$$

By substituting $r = j - 1$ in the summation, we obtain

$$\hat{a}_t = (1 - d)x_t + (1 - d)d \sum_{j=0}^{\infty} d^j x_{t-1-j} \tag{3A.3}$$

From Equation 3A.2, we see that

$$\hat{a}_{t-1} = (1 - d) \sum_{r=0}^{\infty} d^r x_{t-1-r}$$

By substituting into Equation 3A.3

$$\begin{aligned} \hat{a}_t &= (1 - d)x_t + d\hat{a}_{t-1} \\ &= \alpha x_t + (1 - \alpha)\hat{a}_{t-1} \end{aligned}$$

where $\alpha = 1 - d$.

3A.2 Relationship between MAD and Standard Deviation for Unbiased Forecasts

Let MAD represent mean absolute deviation of an unbiased forecast and let σ represent the standard deviation of forecast errors (also for an unbiased forecast). When forecast errors are normally distributed, the following relationship holds:

$$\sqrt{\frac{\pi}{2}} \text{MAD} = \sigma$$

To derive this relationship, note that the unbiased forecast would be mean demand, \hat{x} . So, mean absolute deviation can be expressed as

$$\text{MAD} = \int_{-\infty}^{\hat{x}} (\hat{x} - x_0) f_x(x_0) dx_0 + \int_{\hat{x}}^{\infty} (x_0 - \hat{x}) f_x(x_0) dx_0$$

where $f_0(x_0)$ is the probability density function of the normal distribution

$$f_x(x_0) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_0-\hat{x})^2}{2\sigma^2}}$$

Note the following:

$$f'_x(x_0) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x_0-\hat{x})^2}{2\sigma^2}} \left(\frac{-(x_0-\hat{x})}{\sigma^2} \right) = \left(\frac{\hat{x}-x_0}{\sigma^2} \right) f_x(x_0)$$

or equivalently,

$$\sigma^2 f'_x(x_0) = (\hat{x} - x_0) f_x(x_0)$$

which implies

$$\int (\hat{x} - x_0) f_x(x_0) dx_0 = \int \sigma^2 f'_x(x_0) dx_0 = \sigma^2 f_x(x_0)$$

Using this, we can rewrite MAD as

$$\begin{aligned} \text{MAD} &= \int_{-\infty}^{\hat{x}} \sigma^2 f'_x(x_0) dx_0 - \int_{\hat{x}}^{\infty} \sigma^2 f'_x(x_0) dx_0 \\ &= \sigma^2 \int_{-\infty}^{\hat{x}} f'_x(x_0) dx_0 - \sigma^2 \int_{\hat{x}}^{\infty} f'_x(x_0) dx_0 \\ &= \sigma^2 [(f_x(\hat{x}) - 0) - (0 - f_x(\hat{x}))] \\ &= \sigma^2 \frac{2}{\sigma\sqrt{2\pi}} \\ \text{MAD} &= \sigma \sqrt{\frac{2}{\pi}} \end{aligned}$$

Rearranging, this gives the relationship

$$\text{MAD} \sqrt{\frac{\pi}{2}} = \sigma \approx 1.25 \text{MAD}$$

References

- Adam, E. E. and R. J. Ebert. 1976. A comparison of human and statistical forecasting. *AIIE* 8, 120–127.
- Adshead, N. S. and D. H. R. Price. 1987. Demand forecasting and cost performance in a model of a real manufacturing unit. *International Journal of Production Research* 25(9), 1251–1265.
- Agrawal, N. and S. A. Smith. 1996. Estimating negative binomial demand for retail inventory management with unobservable lost sales. *Naval Research Logistics* 43(6), 839–861.
- Anderson, J. R. 1994. Simpler exponentially weighted moving averages with irregular updating periods. *Journal of the Operational Research Society* 45(4), 486–487.
- Anderson, O. D. 1976. *Times Series Analysis and Forecasting: The Box–Jenkins Approach*. London: Butterworths.
- Armstrong, J. S. 1984. Forecasting by extrapolation: Conclusions from 25 years of research. *Interfaces* 14(6), 52–66.
- Armstrong, J. S. 2001a. Combining forecasts. In J. S. Armstrong (Ed.), *Principles of Forecasting*, pp. 417–439. New York: Springer.

- Armstrong, J. S. 2001b. *Principles of Forecasting: A Handbook for Researchers and Practitioners*, Volume 30. New York: Springer Science & Business Media.
- Ashton, A. H. and R. H. Ashton. 1985. Aggregating subjective forecasts: Some empirical results. *Management Science* 31(12), 1499–1508.
- Azoury, K. S. 1985. Bayes solution to dynamic inventory models under unknown demand distribution. *Management Science* 31(9), 1150–1160.
- Basu, S. and R. C. Schroeder. 1977. Incorporating judgments in sales forecasts: Application of the Delphi method at American Hoist & Derrick. *Interfaces* 7(3), 18–27.
- Bates, J. and C. Granger. 1969. The combination of forecasts. *Operational Research Quarterly* 20(4), 451–468.
- Bell, P. C. 1981. Adaptive forecasting with many stockouts. *Journal of the Operational Research Society* 32(10), 865–873.
- Berger, J. O. 2013. *Statistical Decision Theory and Bayesian Analysis*. New York: Springer Science & Business Media.
- Blagg, B., G. Brandenburg, R. Conti, D. Fogarty, T. Hoffman, R. Martins, J. Muir, P. Rosa, and T. Vollmann. 1980. APICS Certification Program Study Guide and Review Course Outline.
- Box, G. E., G. M. Jenkins, and G. C. Reinsel. 2011. *Time Series Analysis: Forecasting and Control*, Volume 734. Hoboken, NJ: John Wiley & Sons.
- Bradford, J. W. and P. K. Sugrue. 1991. Inventory rotation policies for slow moving items. *Naval Research Logistics* 38, 87–105.
- Brown, G. G., A. Geoffrion, and G. Bradley. 1981. Production and sales planning with limited shared tooling at the key operation. *Management Science* 27(3), 247–259.
- Brown, R. G. 1963. *Smoothing, Forecasting and Prediction*. Englewood Cliffs, NJ: Prentice-Hall.
- Brown, R. G. 1977. *Materials Management Systems*. New York: Wiley-Interscience.
- Brown, R. G. 1981. *Confidence in All-Time Supply Forecasts*. Norwich, VT: Materials Management Systems, Inc.
- Bunn, D. and G. Wright. 1991. Interaction of judgemental and statistical forecasting methods: Issues and analysis. *Management Science* 37(5), 501–518.
- Bunn, D. W. 1988. Combining forecasts. *European Journal of Operational Research* 33, 223–229.
- Chatfield, C. 1978. The Holt-Winters forecasting procedure. *Applied Statistics* 27(3), 264–279.
- Chatfield, C. 2013. *The Analysis of Time Series: An Introduction*. London: CRC Press.
- Chow, W. M. 1965. Adaptive control of the exponential smoothing constant. *Journal of Industrial Engineering* 16(5), 314–317.
- Cox, D. R. and H. D. Miller. 1977. *The Theory of Stochastic Processes*, Volume 134. Boca Raton, FL: CRC Press.
- Dickinson, J. P. 1975. Some comments on the combination of forecasts. *Operational Research Quarterly* 24(2), 253–260.
- Eilon, S. and J. Elmaleh. 1970. Adaption limits in inventory control. *Management Science* 16(18), B533–B548.
- Ekern, S. 1981. Adaptive exponential smoothing revisited. *Journal of the Operational Research Society* 32(9), 775–782.
- Eppen, G. D. and A. V. Iyer. 1997. Improved fashion buying using Bayesian updates. *Operations Research* 45(6), 805–819.
- Fildes, R. 1979. Quantitative forecasting—The state of the art: Extrapolative models. *Journal of the Operational Research Society* 30(8), 691–710.
- Fildes, R. 1983. An evaluation of Bayesian forecasting. *Journal of Forecasting* 2, 137–150.
- Fildes, R. 1989. Evaluation of aggregate and individual forecast method selection rules. *Management Science* 35(9), 1056–1065.
- Fildes, R. and P. Goodwin. 2007. Against your better judgment? How organizations can improve their use of management judgment in forecasting. *Interfaces* 37(6), 570–576.
- Fildes, R., P. Goodwin, M. Lawrence, and K. Nikolopoulos. 2009. Effective forecasting and judgmental adjustments: An empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting* 25(1), 3–23.

- Fildes, R. and S. Howell. 1979. On selecting a forecasting model. In S. Makridakis and S. C. Wheelwright (Eds.), *TIMS Studies in the Management Sciences*, Volume 12, pp. 297–312. Amsterdam: North Holland.
- Fisher, M. and C. D. Ittner. 1999. The impact of product variety on automobile assembly operations: Empirical evidence and simulation analysis. *Management Science* 45(6), 771–786.
- Fisher, M. L., J. H. Hammond, W. R. Obermeyer, and A. Raman. 1994. Making supply meet demand in an uncertain world. *Harvard Business Review* May–June, 83–93.
- Fleischmann, M., J. M. Bloemhof-Ruwaard, R. Dekker, E. Van der Laan, J. A. Van Nunen, and L. N. Van Wassenhove. 1997. Quantitative models for reverse logistics: A review. *European Journal of Operational Research* 103(1), 1–17.
- Fliedner, E., B. Flores, and V. Mabert. 1986. Evaluating adaptive smoothing models: Some guidelines for implementation. *International Journal of Production Research* 24(4), 955–970.
- Flores, B. and D. C. Whybark. 1986. A comparison of focus forecasting with averaging and exponential smoothing. *Production and Inventory Management Journal* 27(3), 96–103.
- Flowers, A. D. 1980. A simulation study of smoothing constant limits for an adaptive forecasting system. *Journal of Operations Management* 1(2), 85–94.
- Fortuin, L. 1980. The all-time requirement of spare parts for service after sales—Theoretical analysis and practical results. *International Journal of Operations and Production Management* 1(1), 59–70.
- Galbraith, C. S. and G. B. Merrill. 1996. The politics of forecasting: Managing the truth. *California Management Review* 38(2), 29–43.
- Gardner, E. S. 1983. Approximate decision rules for continuous review inventory systems. *Naval Research Logistics Quarterly* 30, 59–68.
- Gardner, E. S. 1990. Evaluating forecast performance in an inventory control system. *Management Science* 36(4), 490–499.
- Gardner, E. S. 2006. Exponential smoothing: The state of the art—Part II. *International Journal of Forecasting* 22(4), 637–666.
- Gardner, E. S. and D. Dannenbring. 1980. Forecasting with exponential smoothing: Some guidelines for model selection. *Decision Sciences* 11(2), 370–383.
- Gardner, E. S. and E. McKenzie. 1985. Forecasting trends in time series. *Management Science* 31(10), 1237–1246.
- Gardner, E. S. and E. McKenzie. 1989. Seasonal exponential smoothing with damped trends. *Management Science* 35(3), 372–376.
- Gardner, E. S. J. 1988. A simple method of computing prediction intervals for time series forecasts. *Management Science* 34(4), 541–546.
- Gaur, V., S. Kesavan, A. Raman, and M. L. Fisher. 2007. Estimating demand uncertainty using judgmental forecasts. *Manufacturing and Service Operations Management* 9(4), 480–491.
- Georgoff, D. and R. Murdick. 1986. Manager's guide to forecasting. *Harvard Business Review* 62(1), 110–120.
- Goh, T. N. and N. Varaprasad. 1986. A statistical methodology for the analysis of the life-cycle of reusable containers. *IIE Transactions* 18(1), 42–47.
- Goodman, D. 1974. A goal programming approach to aggregate planning of production and work force. *Management Science* 20(12), 1569–1575.
- Granger, C. W. J. and P. Newbold. 2014. *Forecasting Economic Time Series*. San Diego, CA: Academic Press.
- Guide, V. D. R. and L. N. Wassenhove. 2001. Managing product returns for remanufacturing. *Production and Operations Management* 10(2), 142–155.
- Gupta, S. and P. Wilton. 1987. Combination of forecasts: An extension. *Management Science* 33(3), 356–372.
- Hamburg, M. and P. Young. 1994. *Statistical Analysis for Decision Making*. Belmont, CA: Wadsworth Publishing Company.
- Harrison, P. and C. Stevens. 1971. A Bayesian approach to short-term forecasting. *Operational Research Quarterly* 22(4), 341–362.
- Harrison, P. and C. Stevens. 1976. Bayesian forecasting. *Journal of the Royal Statistical Society B38*, 205–228.
- Harrison, P. J. 1967. Exponential smoothing and short-term sales forecasting. *Management Science* 13(11), 821–842.

- Harrison, P. J. and O. Davies 1964. The use of cumulative sum (CUSUM) techniques for the control of product demand. *Operations Research* 12(2), 325–333.
- Hausman, W. H. and R. M. Kirby. 1970. Estimating standard deviations for inventory control. *AIIE Transactions* 11(1), 78–81.
- Heath, D. and P. Jackson. 1994. Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution systems. *IIE Transactions* 26(3), 17–30.
- Hill, A., V. Giard, and V. Mabert. 1989. A decision support system for determining optimal retention stocks for service parts inventories. *IIE Transactions* 21(3), 221–229.
- Hogarth, R. and S. Makridakis. 1981. Forecasting and planning: An evaluation. *Management Science* 27(2), 115–138.
- Holt, C. C. 2004. Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting* 20(1), 5–10.
- Hyndman, R., A. B. Koehler, J. K. Ord, and R. D. Snyder. 2008. *Forecasting with Exponential Smoothing: The State Space Approach*. Berlin: Springer Science & Business Media.
- Jacobs, R. A. and H. M. Wagner. 1989. Lowering inventory systems costs by using regression-derived estimators of demand variability. *Decision Sciences* 20(3), 558–574.
- Johnson, L. A. and D. C. Montgomery. 1974. *Operations Research in Production Planning, Scheduling and Inventory Control*, Vol. 6. New York: John Wiley & Sons.
- Johnston, F. R. 1980. An interactive stock control system with a strategic management role. *Journal of the Operational Research Society* 31(12), 1069–1084.
- Johnston, F. R. 1993. Exponentially weighted moving average (EWMA) with irregular updating periods. *Journal of the Operational Research Society* 44(7), 711–716.
- Johnston, F. R. and J. E. Boylan. 1994. How far ahead can an EWMA model be extrapolated. *Journal of the Operational Research Society* 45(6), 710–713.
- Johnston, F. R. and J. E. Boylan. 1996. Forecasting for items with intermittent demand. *Journal of the Operational Research Society* 47, 113–121.
- Johnston, F. R. and P. J. Harrison. 1986. The variance of lead-time demand. *Journal of the Operational Research Society* 37(3), 303–308.
- Kang, H. 1986. Unstable weights in the combination of forecasts. *Management Science* 32(6), 683–695.
- Kaplan, A. J. 1988. Bayesian approach to inventory control of new parts. *IIE Transactions* 20(2), 151–156.
- Karmarkar, U. 1994. A robust forecasting technique for inventory and leadtime management. *Journal of Operations Management* 12(1), 45–54.
- Keilson, J. and P. Kubat. 1984. Parts and service demand distribution generated by primary production growth. *European Journal of Operational Research* 17(2), 257–265.
- Kelle, P. and E. A. Silver. 1989. Purchasing policy of new containers considering the random returns of previously issued containers. *IIE Transactions* 21, 349–354.
- Kurawarwala, A. A. and H. Matsuo. 1996. Forecasting and inventory management of short life cycle products. *Operations Research* 44(1), 131–150.
- Lariviere, M. A. and E. L. Porteus. 1999. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science* 45(3), 346–363.
- Larsen, R. J. and M. L. Marx. 2012. *An Introduction to Mathematical Statistics and Its Applications*. Boston: Prentice Hall.
- Lau, H. and A. Lau. 1996. Estimating the demand distributions of single-period items having frequent stockouts. *European Journal of Operational Research* 92, 254–265.
- Lawrence, M., P. Goodwin, M. O'Connor, and D. Önkal. 2006. Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting* 22(3), 493–518.
- Lawrence, M. J., R. H. Edmurdson, and M. J. O'Connor. 1986. Accuracy of combining judgemental and statistical forecasts. *Management Science* 32(12), 1521–1532.
- Lewandowski, R. 1979. *La Prevision a Court Terme*. Paris: Dunod.
- Lewandowski, R. 1982. Sales forecasting by FORSYS. *Journal of Forecasting* 1, 205–214.
- Lewis, C. D. 1978. The versatility of the exponential smoother's portfolio of skills. *Production and Inventory Management Journal* 19(2), 53–66.

- Lovejoy, W. S. 1990. Myopic policies for some inventory models with uncertain demand distributions. *Management Science* 36(6), 724–738.
- Maddala, G. S. and K. Lahiri. 1992. *Introduction to Econometrics*, Volume 2. New York: Macmillan.
- Makridakis, S. 1990. Sliding simulation: A new approach to time series forecasting. *Management Science* 36(4), 505–512.
- Makridakis, S., A. Andersen, R. Carbone, R. Filder, M. Hibon, R. Lewandowski, J. Newton, E. Parzen, and R. L. Winkler. 1982. The accuracy of extrapolation (time series) methods: Results of a forecasting competition. *Journal of Forecasting* 1, 111–153.
- Makridakis, S., C. Chatfield, M. Hibon, M. Lawrence, T. Mills, K. Ord, and L. F. Simmons. 1993. The M2-competition: A real-time judgmentally based forecasting study. *International Journal of Forecasting* 9(1), 5–22.
- Makridakis, S. and M. Hibon. 2000. The M3-competition: Results, conclusions and implications. *International Journal of Forecasting* 16(4), 451–476.
- Makridakis, S. and S. C. Wheelwright. 1979. Forecasting: Framework and overview. In S. Makridakis and S. C. Wheelwright (Eds.), *TIMS Studies in the Management Sciences*, Volume 12, pp. 1–15. Amsterdam: North Holland.
- Makridakis, S., S. C. Wheelwright, and R. J. Hyndman. 2008. *Forecasting Methods and Applications*. Hoboken, NJ: John Wiley & Sons.
- Makridakis, S. and R. L. Winkler. 1983. Averages of forecasts: Some empirical results. *Management Science* 29(9), 987–996.
- McClain, J. and L. J. Thomas. 1973. Response-variance tradeoffs in adaptive forecasting. *Operations Research* 21(2), 554–568.
- McKenzie, E. 1978. The monitoring of exponentially weighted forecasts. *Journal of the Operational Research Society* 29(5), 449–458.
- McLeavey, D. W., T. S. Lee, and E. E. Adam. 1981. An empirical evaluation of individual item forecasting models. *Decision Sciences* 12(4), 708–714.
- Montgomery, D. C. 1970. Adaptive control of exponential smoothing parameters by evolutionary operation. *AIIE Transactions* 2(3), 268–269.
- Muir, J. W. 1980. Forecasting items with irregular demand. In *American Production and Inventory Society Conference Proceedings*, Vol. 23, pp. 143–145, Los Angeles.
- Nahmias, S. 1994. Demand estimation in lost sales inventory systems. *Naval Research Logistics* 41, 739–757.
- Oliva, R. and N. Watson. 2009. Managing functional biases in organizational forecasts: A case study of consensus forecasting in supply chain planning. *Production and Operations Management* 18(2), 138–151.
- Oliva, R. and N. Watson. 2011. Cross-functional alignment in supply chain planning: A case study of sales and operations planning. *Journal of Operations Management* 29(5), 434–448.
- Peterson, R. 1969. A note on the determination of optimal forecasting strategy. *Management Science* 16(4), B165–B169.
- Rhodes, P. 1996. Forecasting seasonal demand. *APICS—The Performance Advantage April*, 50–54.
- Roberts, A. and R. Reed. 1969. The development of a self-adaptive forecasting technique. *AIIE Transactions* 1(4), 314–322.
- Sanders, N. and K. Manrodt. 1994. Forecasting practices in US corporations: Survey results. *Interfaces* 24(2), 92–100.
- Sanders, N. R. and K. B. Manrodt. 2003a. The efficacy of using judgmental versus quantitative forecasting methods in practice. *OMEGA* 31(6), 511–522.
- Sanders, N. R. and K. B. Manrodt. 2003b. Forecasting software in practice: Use, satisfaction, and performance. *Interfaces* 33(5), 90–93.
- Sanders, N. R. and L. P. Ritzman. 1989. Some empirical findings on short-term forecasting: Technique complexity and combinations. *Decision Sciences* 20(3), 635–640.
- Schmittlein, D. C., J. Kim, and D. G. Morrison. 1990. Combining forecasts: Operational adjustments to theoretically optimal rules. *Management Science* 36(9), 1044–1056.
- Shore, H. 1995. Identifying a two-parameter distribution by the first two sample moment (partial and complete). *Journal of Statistical Computation and Simulation* 52(1), 17–32.

- Shore, H. 1998. Approximating an unknown distribution when distribution information is extremely limited. *Communications in Statistics—Simulation and Computation* 27(2), 501–523.
- Steece, B. and S. Wood. 1979. An ARIMA-based methodology for forecasting in a multi-item environment. *Management Science* 25, 167–187.
- Stevens, C. 1974. On the variability of demand for families of items. *Operational Research Quarterly* 25(3), 411–419.
- Syntetos, A. A. and J. E. Boylan. 2005. The accuracy of intermittent demand estimates. *International Journal of Forecasting* 21(2), 303–314.
- Taylor, P. F. and M. E. Thomas. 1982. Short term forecasting: Horses for courses. *Journal of the Operational Research Society* 33(8), 685–694.
- Timmermann, A. 2006. Forecast combinations. *Handbook of Economic Forecasting* 1, 135–196.
- Trigg, D. 1964. Monitoring a forecasting system. *Operational Research Quarterly* 15, 271–274.
- Trigg, D. and A. Leach. 1967. Exponential smoothing with an adaptive response rate. *Operational Research Quarterly* 18(1), 53–59.
- Walton, J. H. D. 1994. Inventory control—A soft approach? *Journal of the Operational Research Society* 45(4), 485–486.
- Watson, R. 1987. The effects of demand-forecast fluctuations on customer service and inventory cost when demand is lumpy. *Journal of the Operational Research Society* 38(1), 75–82.
- Wecker, W. 1978. Predicting demand from sales data in the presence of stockouts. *Management Science* 24(10), 1043–1054.
- Whybark, D. C. 1973. A comparison of adaptive forecasting techniques. *Logistic Transportation Review* 9(1), 13–26.
- Willemain, T. R., C. N. Smart, and H. F. Schwarz. 2004. A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting* 20, 375–387.
- Willemain, T. R., C. N. Smart, J. H. Shockor, and P. A. DeSautels. 1994. Forecasting intermittent demand in manufacturing: A comparative evaluation of Croston's method. *International Journal of Forecasting* 10, 529–538.
- Williams, T. M. 1987. Adaptive Holt-Winters forecasting. *Journal of the Operational Research Society* 38(6), 553–560.
- Winters, P. 1960. Forecasting sales by exponentially weighted moving averages. *Management Science* 6, 324–342.
- Wright, D. J. 1986. Forecasting data published at irregular time intervals using an extension of Holt's method. *Management Science* 32(4), 499–510.



REPLENISHMENT SYSTEMS FOR MANAGING INDIVIDUAL ITEM INVENTORIES WITHIN A FIRM

In the first part of this book, we have laid groundwork by specifying the environment, goals, constraints, relevant costs, and forecast inputs of production/inventory decisions. Now, we turn our attention to the detailed logic for short-range decision making in situations of a strictly inventory (i.e., nonproduction) nature. We are concerned primarily with individual items or SKUs. However, as discussed in Chapter 2, aggregate considerations will often still play a crucial role by examining the consequences across a population of inventoried items of using different values of certain policy parameters. Management can select the aggregate operating point on an exchange curve, thus implying values of the parameters that are to be used in the item by item decision rules.

Within this part of the book, we deal with the so-called B items, which, as we have seen earlier, comprise the majority of items in a typical inventory situation. Unlike for C items, which will be handled in Section III, there are usually significant savings to be derived from the use of a reasonably sophisticated control system. On the other hand, the potential savings per item are not as high as those for individual A items. The combination of these two factors suggests the use of a hybrid system that combines the relative advantages of human and computer. The management-by-exception control system that we describe is reasonably sophisticated, and usually automated, but requires human intervention on a relatively infrequent basis. Such automated control systems, when applied in practice, have resulted in significant improvements in the control of B items.

In Chapter 4, we first answer the question of how much to replenish in the context of an approximately level demand pattern. Next, in Chapter 5, we consider the more general situation of a demand pattern, such as that seen with seasonal items, where the mean or forecast demand changes appreciably with time. This more general case considerably complicates the analysis. Therefore, the suggested decision system is based on a reasonable heuristic decision rule.

Chapter 6 is concerned, for the most part, with the situation where the average demand rate stays approximately level with time, but there is a random component present. That is, the mean or forecast demand stays constant or changes slowly with time, but there definitely are appreciable forecast errors. When demand is known only in a probabilistic sense, two additional questions must be answered: “How often should we review the stock status?” and “When should we place a replenishment order?”

All the material in Chapters 4 through 6 ignores the possible benefits of coordinating the replenishments of two or more items. We return to this important topic in Chapters 10 and 11 of Section IV.

Chapter 4

Order Quantities When Demand Is Approximately Level

In this chapter, we address the question of how large a replenishment quantity to use for a single item when conditions are rather stable (any changes in parameters occur slowly over time), and there is relatively little or no uncertainty concerning the level of demand. This simplified situation is a reasonable approximation of reality on certain occasions. More important, however, the results obtained turn out to be major components of the decision systems when parameters change with time, demand is probabilistic or multiple items are managed jointly. In addition, the goal is not to have a completely automated replenishment system, but to provide decision support to a planner. This is essential for several reasons, including (1) giving the planner the ability to incorporate factors not included in the underlying mathematical model, and (2) cultivating a sense of accountability and responsibility on the part of the decision makers.

In Section 4.1, we discuss the rather severe assumptions needed to derive the economic order quantity (EOQ). This is followed in Section 4.2 by the actual derivation. In Section 4.3, we show that the total relevant costs are not seriously affected by somewhat large deviations of the replenishment quantity away from the minimizing value. This robust property of the EOQ is important from an operational standpoint as there are many factors in practice that may make a quantity other than the calculated EOQ convenient. Section 4.4 introduces a tabular aid that groups items into broad categories to make implementation easier, particularly in a noncomputerized system. The important situation of quantity discounts is handled in Section 4.5. This is followed, in Section 4.6, by a discussion of the effects of inflation on the order quantity. Section 4.7 deals with several types of limitations on replenishment quantities ignored in the earlier development of the EOQ. In Section 4.8, we address the production situation in which the entire replenishment batch does not arrive at one time. Other important factors in selecting the best reorder quantity are included in Section 4.9. Finally, the derivation of the basic EOQ includes two cost parameters: (1) the carrying charge r , and (2) the fixed cost per replenishment A . As discussed in Chapter 2, one or both of the values of these parameters may be explicitly evaluated or implicitly specified. They

may be implicitly specified by looking at the aggregate consequences of using different values of r (or A/r), and then selecting a suitable aggregate operating point, which implies an r (or A/r) value. Section 4.10 shows how to develop the exchange curves, a crucial step in this implicit approach.

4.1 Assumptions Leading to the Basic EOQ

First, we will state the assumptions. Some of these assumptions are quite strong and may initially appear to be quite limiting but, as we will see, the EOQ forms an important building block in the majority of decision systems that we advocate. In addition, Zheng (1992) has shown that the fractional cost penalty of using the EOQ instead of a more complex, but optimal, order quantity when demand is random is less than 1/8, and is typically much smaller. (Axsäter (1996) has shown that the penalty is less than 11.8%.) Further discussion will be directed to these points later in this chapter and in the subsequent chapters.

1. The demand rate is constant and deterministic.
2. The order quantity need not be an integral number of units, and there are no minimum or maximum restrictions on its size.
3. The unit variable cost does not depend on the replenishment quantity; in particular, there are no discounts in either the unit purchase cost or the unit transportation cost.
4. The cost factors do not change appreciably with time; in particular, inflation is at a low level.
5. The item is treated entirely independently of other items; that is, benefits from joint review or replenishment do not exist or are simply ignored.
6. The replenishment lead time is of zero duration; as we will see in Section 4.9.1, extension to a deterministic, nonzero replenishment lead time creates no problem.
7. No shortages are allowed.
8. The entire order quantity is delivered at the same time.
9. The planning horizon is very long. In other words, we assume that all the parameters will continue at the same values for a long time.

All these assumptions will be relaxed later in this chapter, or elsewhere in the book. Situations where the demand rate changes appreciably with time (i.e., the first assumption is violated) will be treated in Chapters 5 and 8. Problem 4.5 deals with the case in which shortages are allowed. (See also Montgomery et al. (1973) and Moon and Gallego (1994) for a treatment of the case in which there is a mixture of lost sales and backorders. More recently, Bijvank and Vis (2011) review inventory models with lost sales and backordering.) We often refer to “changing the givens,” meaning that the static version of the problem may not be appropriate over time. Rather, some of the assumptions, costs, or other data may, and should, change. For example, in this chapter, we generally assume that the setup cost, A , is given. Setup reduction efforts, however, may significantly reduce this value, yielding enormous benefits. Likewise, supplier portals, electronic data interchange (EDI), or other technological infrastructure may reduce the fixed cost of ordering from suppliers, creating new optimal order quantities and significantly lower costs. Sometimes, it is possible to negotiate with suppliers to reduce the minimum order size required for a quantity discount, or it may be possible to negotiate with transportation firms to use smaller trucks to reduce the minimum order size.*

* See Moinzadeh et al. (1997) for a discussion of the impact on traffic of small lot sizes.

The traditional approach to the EOQ has been to optimize the order quantity without questioning the input data. The “changing the givens” approach recognizes the possibility of changing the parameters. In many cases, the most powerful benefit to the firm is through lowering the setup cost, rather than optimizing the order quantity. We argue, however, that the best approach is to initially optimize the order quantity given the parameters as they are, and then devote management (and labor) effort to changing the givens. A fastener manufacturer that supplies U.S. automotive assemblers had setup times on some machines that took from 12 to 20 hours. One of the authors worked as a consultant to this firm and developed (nearly) optimal order quantities. At the same time, however, factory supervisors and operators were working hard to lower the setup times. The use of the new order quantities would lower costs, but not nearly as much as lower setup times would. Unfortunately, the setup reduction process was very slow, and the firm would have sacrificed savings if management had delayed changing order quantities until this process was finished. As setup times decrease, new order quantities should be calculated. Numerous authors have addressed these issues. See, for instance, Porteus (1985, 1986a,b), Silver (1992), Hong et al. (1996), Spence and Porteus (1987), Van Beek and Van Putten (1987), Freeland et al. (1990), Neves (1992), Gallego and Moon (1995), Bourland et al. (1996), Van der Duyn Schouten et al. (1994), and Freimer et al. (2006). We will discuss these issues further in the context of manufacturing in Section IV.

4.2 Derivation of the EOQ

In determining the appropriate order quantity, we use the criterion of minimization of total relevant costs; relevant in the sense that they are truly affected by the choice of the order quantity. We emphasize that in some situations, there may be certain, perhaps intangible, relevant costs that are not included in the model to be discussed. This is one of the reasons we mention the need for a manual override capability. In addition, cost minimization need not be the only appropriate criterion. For example, Buzacott and Zhang (2004) investigate how inventory decisions are affected by the current financial state of the firm. Singhal (1988) also closely examines the connection to the financial state of the firm and argues that inventories are risky, and thus increase the firm’s opportunity cost of capital. Other work of interest includes Gerchak and Wang (1994b), which investigates the issue of demand changing as a result of inventory decisions. More recently, Buzacott (2013) reflects further on similar issues.

Returning to the consideration of costs, as discussed in Chapter 2, there are five fundamental categories of costs:

1. Basic production or purchase costs
2. Inventory carrying costs
3. Costs of insufficient capacity in the short run
4. Control system costs
5. Costs of changing work force sizes and production rates

The fifth category is not relevant for the item-by-item control we are considering here. Also, given that we are restricting our attention to a particular type of control system (an order quantity system), the control system costs are not influenced by the exact value of the order quantity. Therefore, they need not be taken into account when selecting the value of the order quantity.

Because of the deterministic nature of the demand, the only cost relevant to the third category would be that which is caused by the decision maker deliberately choosing to run short of the

inventory before making a replenishment. For now, we ignore the possibility of allowing planned shortages.

Finally, we are left with only the first two types of costs being relevant to the economic selection of the replenishment quantity.

Before proceeding further, let us introduce some notation:

Q = the replenishment order quantity, in units.

D = the demand rate of the item, in units/unit time.

T = the time between placement (and receipt) of replenishment orders.

A = the fixed cost component (independent of the magnitude of the replenishment quantity) incurred with each replenishment, in dollars. (This cost component has been discussed at length in Chapter 2.)

v = the unit variable cost of the item. This is not the selling price of the item, but rather its value (immediately after the replenishment operation now under consideration) in terms of raw materials and value added through processing and assembly operations. We often think of v as the value of the item on the shelf, including material handling. The dimensions are \$/unit.

r = the carrying charge, the cost of having one dollar of the item tied up in the inventory for a unit time interval (normally 1 year); that is, the dimensions are \$/\$/unit time. (Again, this cost component was discussed at length in Chapter 2.)

$TRC(Q)$ = the total relevant costs per unit time, that is, the sum of those costs per unit time that can be influenced by the order quantity Q . The dimensions are \$/unit time.

Because the parameters involved are assumed not to change with time, it is reasonable (and, indeed, mathematically optimal) to think in terms of using the same order quantity, Q , each time that a replenishment is made. Furthermore, because (1) demand is deterministic, (2) the replenishment lead time is zero, and (3) we have chosen not to allow planned shortages, it is clear that each replenishment will be made when the inventory level is exactly at zero. A graph of the inventory level with time is shown in Figure 4.1.

Note that the time between replenishments is given by $T = Q/D$, the time to deplete Q units at a rate of D units per unit time. (D is the usage over a period of time; normally 12 months is used.) Therefore, the number of replenishments per unit time is D/Q . Say a three-ohm resistor has the following data: $D = 2,400$ units per year and $Q = 800$ units. The time between replenishments is $800/2,400 = 0.33$ year, or 4 months. The number of replenishments per year is $2,400/800 = 3$. Associated with each of these is a replenishment cost given by $A + Qv$, where one of our assumptions ensures that the unit variable cost v does not depend on Q . Therefore, the replenishment costs per unit time (C_r) are given by

$$C_r = (A + Qv)D/Q$$

or

$$C_r = AD/Q + Dv \quad (4.1)$$

We can see that the second component in Equation 4.1 is independent of Q and, thus, can have no effect on the determination of the best Q value. (It represents the constant acquisition

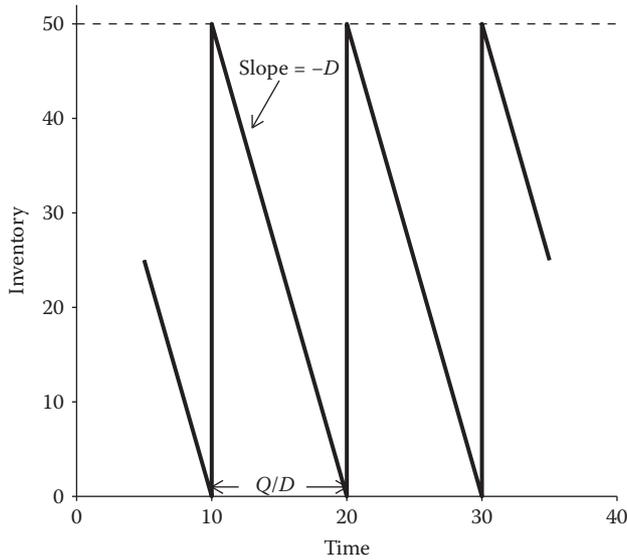


Figure 4.1 Behavior of the inventory level with time.

cost of the item per unit time which cannot be affected by the magnitude of the order quantity.) Therefore, it will be neglected in future discussions.*

As discussed in Chapter 2, the common method of determining the costs of carrying the inventory over a unit time period is through the relation

$$C_c = \bar{I}vr$$

where \bar{I} is the average inventory level, in units. The average height of the sawtooth diagram of Figure 4.1 is $Q/2$. Therefore,

$$C_c = Qvr/2 \quad (4.2)$$

Combining Equations 4.1 and 4.2 and neglecting the Dv term, as discussed above, we have that the total relevant costs per unit time are given by

$$TRC(Q) = AD/Q + Qvr/2 \quad (4.3)$$

The two components of Equation 4.3 and their total are plotted for an illustrative numerical example in Figure 4.2. We can see that the replenishment costs per unit time decrease as Q increases (there are fewer replenishments), whereas the carrying costs increase with Q (a larger Q means a larger average inventory). The sum of the two costs is a u -shaped function with a minimum that can be found in a number of ways. (Adding the neglected Dv term would simply shift every point on the total cost curve up by Dv ; hence, the location of the minimum would not change.) One

* The Dv component will be of crucial interest in Section 4.5 where, in the case of quantity discounts, v depends on Q .

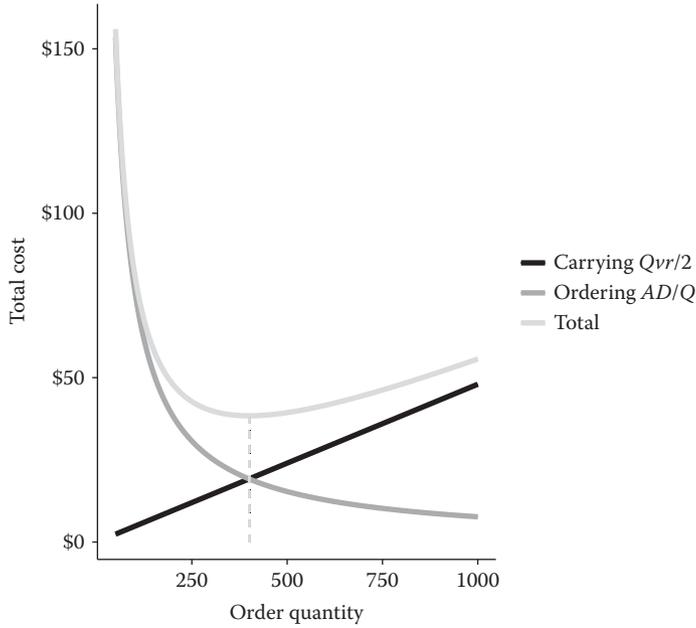


Figure 4.2 Costs as functions of the replenishment quantity.

convenient way to find the minimum is to use the necessary condition that the tangent or slope of the curve is zero at the minimum:

$$\frac{dTRC(Q)}{dQ} = 0$$

that is,

$$\frac{vr}{2} - \frac{AD}{Q^2} = 0$$

$$Q_{opt} \text{ or } EOQ = \sqrt{\frac{2AD}{vr}} \tag{4.4}$$

A sufficient condition for a minimum, namely, that the second derivative is positive, is also satisfied:

$$\frac{d^2 TRC(Q)}{dQ^2} = 2AD/Q^3 > 0 \text{ for any } Q > 0$$

Equation 4.4 defines the EOQ also known as the Wilson lot size. This is one of the earliest and most well-known results of inventory theory (see, e.g., Harris (1913), reprinted as Harris (1990), and Erlenkotter (1990)). Note that Equation 4.4 illustrates, among other points, that the preferred order quantity goes up as the square root of demand, rather than being directly proportional to the D value. Substituting the square root expression for the EOQ in Equation 4.3, we find that the two cost components are equal at the EOQ and we obtain the simple and useful result:

$$TRC(EOQ) = \sqrt{2ADvr} \tag{4.5}$$

It should be emphasized that the equality of the two cost components ($Qvr/2$ and AD/Q) at the point where their sum is minimized is a very special property of the particular cost functions considered here; it certainly does not hold in general.

The EOQ minimizes the total relevant costs under a given set of circumstances. It may in fact be more appropriate to first find ways of driving down the fixed cost A or the carrying charge r . This would lower the entire total cost curve in Figure 4.2. Also note from Equation 4.4 that a lower A value will reduce the EOQ, and thus the average inventory level.

As noted earlier, for any given order quantity Q , the time between replenishments $T = Q/D$. Using this relationship, we can write the total relevant costs for this model as a function of the time supply T rather than the quantity. Substituting $T = Q/D$ in Equation 4.3:

$$TRC(T) = A/T + DvrT/2 \quad (4.6)$$

Using Equation 4.4, we find the expression for the time between orders at the EOQ:

$$T_{\text{EOQ}} = \frac{\text{EOQ}}{D} = \sqrt{\frac{2A}{Dvr}} \quad (4.7)$$

A frequently used measure of inventory performance is the inventory turnover ratio, defined as the annual demand rate divided by the average inventory level. This ratio gives the number of times the average inventory is sold or “turned” over some given time period. For any given quantity Q , we calculate the inventory turnover ratio as

$$TR = \frac{D}{\bar{I}} = \frac{D}{Q/2} = \frac{2D}{Q} \quad (4.8)$$

At the EOQ, this inventory turnover ratio is

$$TR_{\text{EOQ}} = \sqrt{\frac{2Dvr}{A}} \quad (4.9)$$

a result that is again seen to depend on the individual item factors D , v , and A . In particular, all other things remaining constant, we see that the turnover ratio increases in proportion to the square root of the demand rate, D .

4.2.1 Numerical Illustration

Consider a 3-ohm resistor used in the assembly of an automated processor of an x-ray film. The demand for this item has been relatively stable over time at a rate of 2,400 units/year. The unit variable cost of the resistor is \$0.40/unit and the fixed cost per replenishment is estimated to be \$3.20. Suppose further that an r value of 0.24 \$/\$/year is appropriate to use. Then Equation 4.4 gives

$$\begin{aligned} \text{EOQ} &= \sqrt{\frac{2 \times \$3.20 \times 2,400 \text{ units/year}}{\$0.40/\text{unit} \times 0.24\$/\$/\text{year}}} \\ &= 400 \text{ units} \end{aligned}$$

Note that the dimensions appropriately cancel. This would not be the case, for example, if D and r were defined using different time units (e.g., monthly demand with annual carrying charge). Equation 4.5 reveals that the total relevant costs per year for the resistor are

$$\begin{aligned} \text{TRC}(\text{EOQ}) &= \sqrt{2 \times \$3.20 \times 2,400 \text{ units/year} \times \$0.40/\text{unit} \times 0.24\$/\$/\text{year}} \\ &= \$38.40/\text{year} \end{aligned}$$

A check on the computations is provided by noting that substitution of $Q = 400$ units in Equation 4.3 also produces a total cost \$38.40/year with each of the two terms equal at \$19.20/year. If this item has an order quantity of 4 months of time supply, or 800 units, we can evaluate the benefit of changing to the EOQ of 400. Substitution of $Q = 800$ into Equation 4.3 reveals that the total relevant costs for the item are \$48/year or some 25% higher than the costs incurred under the use of the EOQ. Our experience indicates that savings of this magnitude are surprisingly common.

We can use other expressions developed in this section to evaluate the time between orders and inventory turnover at the EOQ. The time between orders at the EOQ is given by Equation 4.7:

$$T_{\text{EOQ}} = \sqrt{\frac{2 \times \$3.20}{2,400 \text{ units/year} \times \$0.40/\text{unit} \times 0.24\$/\$/\text{year}}} = 0.167 \text{ year, or 2 months}$$

The inventory turnover ratio at the EOQ, following Equation 4.9, is

$$TR_{\text{EOQ}} = \sqrt{\frac{2Dvr}{A}} = \sqrt{\frac{2 \times 2,400 \text{ units/year} \times \$0.40/\text{unit} \times 0.24\$/\$/\text{year}}{\$3.20}} = 12 \text{ turns per year}$$

Note that, at the EOQ, we are ordering a 2-month supply, but we are turning the average inventory 12 times per year or once every month. This is because the 2-month supply ($Q = 400$) is the *maximum* inventory level while the average inventory level (used to calculate turns) is half this quantity.

4.3 Sensitivity Analysis

An important property of the EOQ model is that total relevant costs are rather insensitive to deviations from the optimal lot size given by the EOQ. Referring back to Figure 4.2, note that the total cost curve is quite flat in the neighborhood of the EOQ. This indicates that reasonable-sized deviations from the EOQ will have little impact on the total relevant costs incurred. Here, we explicitly examine the increase in the total relevant cost associated with choosing a lot size other than the EOQ. Suppose we use a quantity Q' that deviates from the EOQ according to the following relation:

$$Q' = (1 + p)\text{EOQ}$$

that is, $100p$ is the percentage deviation of Q' from the EOQ. The percentage cost penalty (PCP) for using Q' instead of the EOQ is given by

$$\text{PCP} = \frac{\text{TRC}(Q') - \text{TRC}(\text{EOQ})}{\text{TRC}(\text{EOQ})} \times 100$$

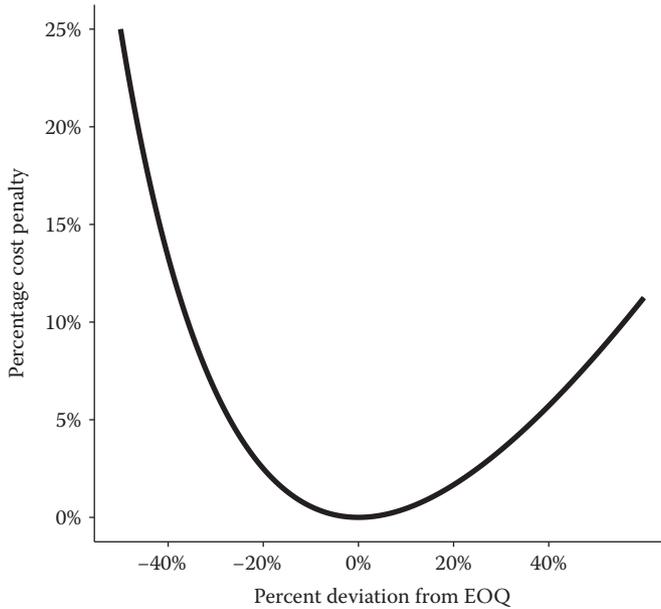


Figure 4.3 Cost penalty for using a $Q' = \text{EOQ}(1 + p)$ value different from the EOQ.

As shown in Section 4A.1 of the Appendix to this chapter,

$$\text{PCP} = \left(\frac{p^2}{2(1+p)} \right) \times 100 \quad (4.10)$$

This expression is plotted in Figure 4.3. It is clear that even for values of p significantly different from zero, the associated cost penalties are quite small.

To illustrate, consider the 3-ohm resistor that was used as a numerical example in Section 4.2. The EOQ was found to be 400 units. Suppose that, instead, a value of 550 units was used; 550 is 1.375×400 ; thus, $p = 0.375$. Equation 4.10 or Figure 4.3 reveals that this 37.5% deviation from optimal lot size results in a percentage increase in the total relevant costs of only 5.1%.

The insensitivity of total costs to the exact value of Q used has two important implications. First, the use of an incorrect value of Q can result from inaccurate estimates of one or more of the parameters D , A , v , and r that are used to calculate the EOQ.* The conclusion is that it is not worth making accurate estimates of these input parameters if considerable effort is involved; in most cases inexpensive, crude estimates should suffice. Second, certain order quantities may have additional appeal over the EOQ (for nonfinancial reasons or because of physical constraints, such as pallet size or truckload size, which are not included in the basic EOQ model). The relative flatness of the total cost curve near the EOQ indicates that such values can be used provided that they are reasonably close to the EOQ. This seemingly trivial result is one of the important cornerstones of

* It is worth noting that the cost penalty can actually be zero even if D , A , v , and r are inaccurately estimated, as long as the resulting ratio AD/vr is correct. For example, if both D and v are overestimated by the same percentage, the obtained EOQ will be the same as that resulting from the use of the correct D and v values.

operational inventory control. Without this flexibility, implementable decision systems would be much more difficult to design. Detailed references on sensitivity analysis include Dobson (1988), Juckler (1970), and Lowe and Schwarz (1983).

4.4 Implementation Aids

In spite of the widespread availability of inexpensive computers, our experience indicates that many firms still manually control inventories. Many other firms use simple spreadsheet models, similar to those described throughout this book. We now discuss a table that can facilitate implementation because it groups multiple items into a few broad categories and can be developed easily on a spreadsheet. While it is not too difficult to develop a spreadsheet model to calculate the EOQ values for a group of items, the approach here simplifies implementation. Furthermore, while the focus of this chapter is lot size decisions for a single item, the approach outlined in this section facilitates the coordination of replenishment across multiple items, a topic that we address in Chapter 10.

The basic idea is that one value of the order quantity, typically expressed as a time supply, is used over a range of one or more of the parameters. This makes it easier for an inventory manager to track and control many items. There is an economic trade-off involved, however. The larger the number of ranges, the smaller the percentage cost penalties resulting from not using the exact EOQ, but of course, the difficulty of compiling and using the tables increases with the number of ranges involved. Further relevant discussion can be found in Chakravarty (1981), Crouch and Oglesby (1978), Donaldson (1981), and Maxwell and Singh (1983).

To illustrate, let us consider a table usable for a set of items having the same values of A and r . Suppose we deal with a group of items with $A = \$3.20$ and $r = 0.24$ \$/\$/year. Furthermore, let us assume that management feels that the replenishment quantity of any item should be restricted to one of nine possible time supplies—namely, $1/4$, $1/2$, $3/4$, 1 , 2 , 3 , 4 , 5 , 6 , and 12 months. Table 4.1 shows the months of supply to use as a function of the annual dollar usage (Dv) of the item under consideration. A numerical illustration of the use of the table will be given shortly. First, let us discuss how such a table is constructed.

Table 4.1 Tabular Aid for Use of EOQ (for $A = 3.20$ and $r = 0.24$ \$/\$/year)

<i>For Annual Dollar Usage (Dv) in This Range</i>	<i>Use This Number of Months of Supply</i>
$30,720 \leq Dv$	$1/4$ (≈ 1 week)
$10,240 \leq Dv < 30,720$	$1/2$ (≈ 2 weeks)
$5,120 \leq Dv < 10,240$	$3/4$ (≈ 3 weeks)
$1,920 \leq Dv < 5,120$	1
$640 \leq Dv < 1,920$	2
$320 \leq Dv < 640$	3
$160 \leq Dv < 320$	4
$53 \leq Dv < 160$	6
$Dv < 53$	12

The use of T months of supply is equivalent to the use of a quantity $Q = DT/12$, provided D is measured in units/year that are the most common units. The use of Equation 4.3 gives

$$\text{TRC}(\text{using } T \text{ months}) = \frac{DTvr}{24} + \frac{12A}{T}$$

Now, consider two adjacent allowable values of the months of supply, call them T_1 and T_2 . Equating the total relevant costs using T_1 months with that using T_2 months, we obtain the value of Dv at which we are indifferent to using T_1 and T_2 :

$$\frac{DT_1vr}{24} + \frac{12A}{T_1} = \frac{DT_2vr}{24} + \frac{12A}{T_2}$$

This reduces to

$$(Dv)_{\text{indifference}} = \frac{288A}{T_1 T_2 r}$$

To illustrate, for the given values of A and r in the example, the indifference point for 1 month and 2 months is

$$(Dv)_{\text{indifference}} = \frac{288 \times 3.20}{1 \times 2 \times 0.24} = \$1,920/\text{year}$$

In a similar fashion, we can develop the rest of Table 4.1. It should be emphasized that more than one value of A (and perhaps r) is likely to exist across a population of items. In such a case, a set of tables would be required, one for each of a number of possible A (and, perhaps, r) values.

4.4.1 Numerical Illustration

Consider a product that is a box of 100 sheets of 8×11 in. stabilization paper. This product has been observed to have a relatively constant demand rate (D) of 200 boxes/year. The unit variable cost (v) is \$16/box. Also assume that it is reasonable to use $A = \$3.20$ and $r = 0.24$ \$/\$/year. We have

$$Dv = \$3,200/\text{year}$$

The table indicates that a 1-month supply should be used. Therefore,

$$Q = \frac{D}{12} = 16.7$$

say, 17 boxes; that is, the paper should be replenished in orders of 17 boxes. (The exact EOQ for this item is 18 boxes and the increase in costs for using 17 instead turns out to be only 0.3%.)

4.5 Quantity Discounts

A number of assumptions were made in Section 4.1 in order to derive the basic EOQ. One of the most severe of these was that the unit variable cost v did not depend on the replenishment quantity. In many practical situations, quantity discounts (on the basic purchase price or transportation

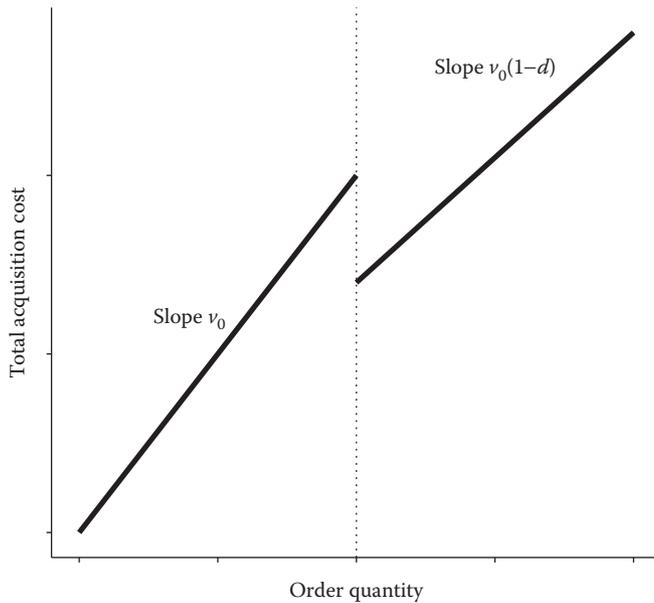


Figure 4.4 “All-units” quantity discount.

costs) exist, and taking advantage of these can result in substantial savings. (See e.g., Katz et al. 1994). We must be able to modify the EOQ to cope with these conditions.

Dolan (1987) provides a number of examples of quantity discounts, including pricing for Sealed Air Corporation’s protection packaging, and prices for advertising in the *Los Angeles Times*. See also Wilson (1993). These authors show that discount structures are quite varied. We restrict our attention to the most common type of discount structure, namely, that of an “all-units” discount. (One less common situation of an incremental discount structure has been treated by Hax and Candea (1984) and Güder et al. (1994). See Benton and Park (1996) and Problem 4.24 also.)*

For the case of a single breakpoint, the unit variable cost in all-units discounts behaves as follows:

$$Dv = \begin{cases} v_0 & 0 \leq Q < Q_b \\ v_0(1-d) & Q_b \leq Q \end{cases} \quad (4.11)$$

where v_0 is the basic unit cost without a discount and d is the discount expressed as a decimal, given on all units when the replenishment quantity is equal to or greater than the breakpoint, Q_b . The total acquisition cost as a function of Q is shown in Figure 4.4. Note the discontinuity at the breakpoint.

Now, it is essential to include the Dv component in a cost expression that is to be used for determining the best replenishment quantity. Proceeding exactly as in Section 4.2, but retaining the Dv component, we end up with two expressions for the total relevant costs, that is, for $0 \leq Q < Q_b$ (and thus no discount)

$$\text{TRC}(Q) = AD/Q + Qv_0r/2 + Dv_0 \quad (4.12)$$

* Das (1988) presents a general discount structure that combines the features of all-units and incremental discounts.

When $Q_b \leq Q$, the discount d is applicable:

$$\text{TRC}_d(Q) = AD/Q + Qv_0(l - d)r/2 + Dv_0(1 - d) \quad (4.13)$$

Although Equations 4.12 and 4.13 are valid for nonoverlapping regions of Q , it is useful, in deriving an algorithm for finding the best Q , to compare the two cost expressions at the same value of Q . A term-by-term comparison of the right-hand sides of the equations reveals that Equation 4.13 gives a lower TRC (Q) than does Equation 4.12 for the same Q . Therefore, if the lowest point on the (4.13) curve is a valid one (i.e., at least as large as Q_b), it must be the optimum, since it is the lowest point on the lower curve.

Note that for nonnegative values of the discount d , the EOQ at the standard price v_0 must be lower than the EOQ calculated with the discount, although the difference in these quantities is typically quite small. This means, if the EOQ with the discount is below Q_b , the EOQ without the discount must also be below Q_b . In the usual situation where both EOQs (calculated with and without the discount) are below the breakpoint, the decision of whether to go for the discount hinges on a trade-off between extra carrying costs versus a reduction in the acquisition costs (primarily the reduction in the unit value, but also a benefit from fewer replenishments per unit time). Where the reduction in acquisition costs is larger than the extra carrying costs, the use of Q_b is preferable, as shown in part a of Figure 4.5. Where the extra carrying costs dominate, the best solution is still the EOQ (with no discount), as shown in part b of Figure 4.5. However, there is one other possibility—namely, that Q_b is relatively low, and it may be attractive to go to the local minimum (low point) on the curve of Equation 4.13. This is the EOQ (with discount) and this case is shown in part c of Figure 4.5.

Taking advantage of the above property of Equation 4.13 being the lower curve, the steps of an efficient algorithm for finding the best value of Q are

Step 1 Compute the EOQ when the discount d is applicable, that is,

$$\text{EOQ}(d) = \sqrt{\frac{2AD}{v_0(1-d)r}}$$

Step 2 Compare $\text{EOQ}(d)$ with Q_b . If $\text{EOQ}(d) \geq Q_b$, then $\text{EOQ}(d)$ is the best order quantity (case c of Figure 4.5). If $\text{EOQ}(d) < Q_b$, go to Step 3.

Step 3 Evaluate

$$\text{TRC}(\text{EOQ}) = \sqrt{2ADv_0r} + Dv_0$$

and $\text{TRC}(Q_b)$, using Equation 4.13.

If $\text{TRC}(\text{EOQ}) < \text{TRC}(Q_b)$, the best order quantity is the EOQ without a discount (case b of Figure 4.5), given by

$$\text{EOQ}(\text{no discount}) = \sqrt{\frac{2AD}{v_0r}}$$

If $\text{TRC}(\text{EOQ}) > \text{TRC}(Q_b)$, the best order quantity is Q_b (case a of Figure 4.5).

This logic can easily be extended to the case of several breakpoints with increasing discounts. The best order quantity is always at a breakpoint or at a feasible EOQ. Kuzdrall and Britney (1982) have used a somewhat different approach that is useful when there is a large number of different

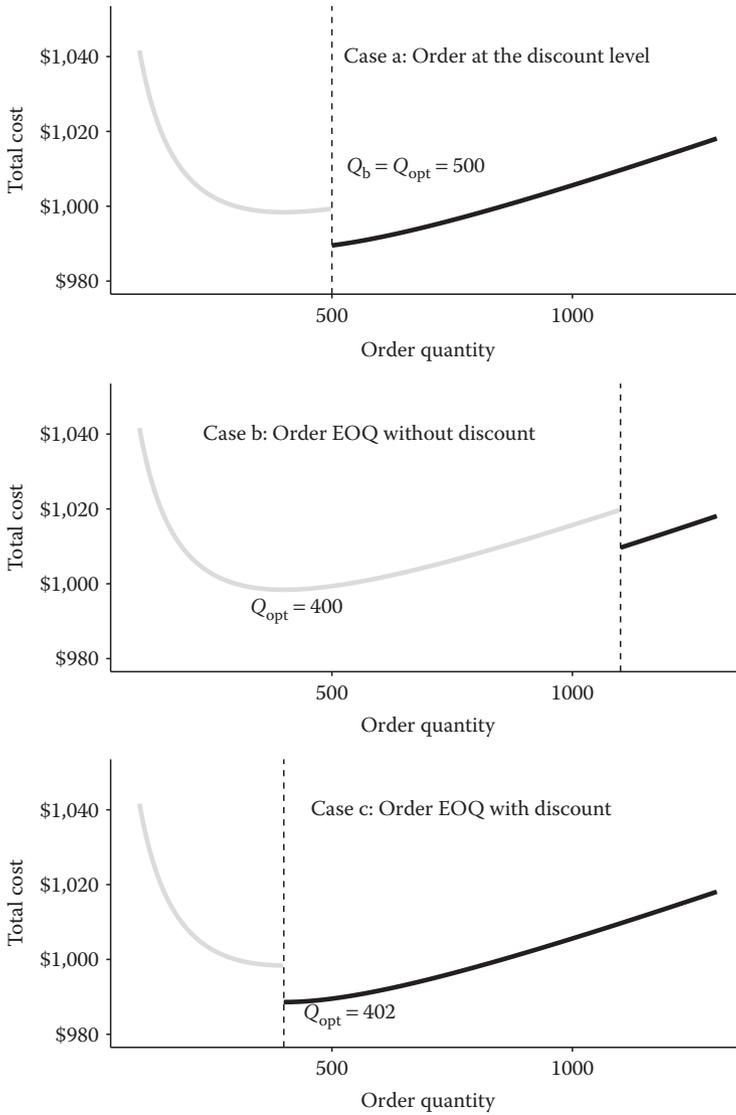


Figure 4.5 Total relevant costs under “all-units” discount.

discounts. Their method estimates the supplier’s fixed and variable costs from a given schedule of discounts and breakpoints. See also Goyal (1995) for a simple procedure.

4.5.1 Numerical Illustrations

Consider three components used in the assembly of the x-ray film processor mentioned in Section 4.2. The supplier offers the same discount structure for each of the items, and discounts are based on replenishment sizes of the *individual* items. The relevant characteristics of the items are given below.

Item	D (Units/Year)	v_0 (\$/Unit)	A (\$)	r (\$/\$/Year)
A	416	14.20	1.50	0.24
B	104	3.10	1.50	0.24
C	4,160	2.40	1.50	0.24

Because of convenience in manufacturing and shipping, the supplier offers a 2% discount on any replenishment of 100 units or higher of a single item.

The computations are as follows:

4.5.2 Item A (An Illustration of Case a of Figure 4.5)

Step 1 EOQ (discount) = 19 units < 100 units.

Step 2 EOQ (discount) < Q_b ; therefore, go to Step 3.

Step 3

$$\begin{aligned} \text{TRC}(\text{EOQ}) &= \sqrt{2 \times 1.50 \times 416 \times 14.20 \times 0.24} + 416 \times 14.20 \\ &= \$5,972.42/\text{year} \\ \text{TRC}(Q_b) &= \text{TRC}(100) = \frac{100 \times 14.20 \times 0.98 \times 0.24}{2} + \frac{1.50 \times 416}{100} \\ &\quad + 416 \times 14.20 \times 0.98 \\ &= \$5,962.29/\text{year} \end{aligned}$$

$\text{TRC}(\text{EOQ}) > \text{TRC}(Q_b)$. Therefore, the best order quantity to use is Q_b , that is, 100 units.

4.5.3 Item B (An Illustration of Case b of Figure 4.5)

Step 1 EOQ (discount) = 21 units < 100 units.

Step 2 EOQ (discount) < Q_b ; therefore, go to Step 3.

Step 3

$$\begin{aligned} \text{TRC}(\text{EOQ}) &= \sqrt{2 \times 1.50 \times 104 \times 3.10 \times 0.24} + 104 \times 3.10 \\ &= \$337.64/\text{year} \\ \text{TRC}(Q_b) &= \text{TRC}(100) = \frac{100 \times 3.10 \times 0.98 \times 0.24}{2} + \frac{1.50 \times 104}{100} \\ &\quad + 104 \times 3.10 \times 0.98 \\ &= \$353.97/\text{year} \end{aligned}$$

$\text{TRC}(\text{EOQ}) < \text{TRC}(Q_b)$. Therefore, use the EOQ without a discount; that is,

$$\text{EOQ} = \sqrt{\frac{2 \times 1.50 \times 104}{3.10 \times 0.24}} \approx 20 \text{ units}$$

4.5.4 Item C (An Illustration of Case c of Figure 4.5)

Step 1

$$\text{EOQ}(\text{discount}) = \sqrt{\frac{2 \times 1.50 \times 4,160}{2.40 \times 0.98 \times 0.24}} = 149 \text{ units} > 100 \text{ units}$$

Step 2 EOQ (discount) is greater than Q_b . Therefore, the Q to use is 149 units (perhaps rounded to 150 for convenience).

We also note that it is very easy to compute Equations 4.12 and 4.13 on a spreadsheet. Using an IF statement to determine which formula to apply based on Q , the optimal value can be found quickly. It is also possible to use Solver in an Excel spreadsheet to minimize the total cost. However, because the TRC functions (illustrated in Figure 4.5) in general have more than one local minimum, use several starting values in Solver to avoid finding just a local minimum.

In an all-units discount situation, a possible strategy is to buy Q_b units to achieve the discount; then, dispose of some of the units, at some unit cost (or revenue), to reduce inventory carrying costs. An application of this philosophy is the case where an entire car of a train is reserved, but not completely filled, by a shipment. Sethi (1984) has analyzed this more general situation.

Much work has been done to further our knowledge of quantity discount situations. Dolan (1987) examines motivations for suppliers to offer discounts and provides several in-depth examples. Monahan (1984), Lal and Staelin (1984), Rosenblatt and Lee (1985), Lee (1986), Dada and Srikanth (1987), Kim and Hwang (1988), Min (1991), Parlar and Wang (1994), and Lu (1995) address the quantity discount problem from the supplier's perspective. For various pricing and discount structures, the supplier's own order quantity and the discounts it offers are determined. Das (1984) provides a simple solution for the problem when a premium is charged for bulk purchases. This work is particularly applicable in developing countries. Grant (1993) describes a simple program for quantity discount decisions used at a small JIT manufacturer who was biased against large quantities, and therefore against taking any quantity discounts. Finally, Abad (1988) examines the all-units discount offered to a retailer whose demand is a function of the price the retailer charges. An algorithm for finding the optimal order quantity and price is given. See also Burwell et al. (1990), Weng and Wong (1993), Weng (1995a,b), Lu and Posner (1994), Kim (1995), and Carlson et al. (1996).

Managers, buyers, and purchasing agents, among others, often think in terms of a group of items purchased from a particular vendor, rather than dealing with single items in isolation. Thus, a potentially important discount can be achieved by having a total order, involving two or more items, exceeding some breakpoint level. This topic will be addressed in Chapter 10, where we will be concerned with coordinated replenishment strategies.

It should be mentioned that, in trying to achieve a quantity discount, one must take into account the implications of a large time supply, a topic to be discussed in Section 4.7.

4.6 Accounting for Inflation

One of the assumptions in the derivation of the EOQ was that the inflation rate was at a negligible level. But, in recent times, many countries have been confronted with fluctuating inflation rates

that often have been far from negligible.* In this section, we investigate the impact of inflation on the choice of replenishment quantities.

There are several options available for modeling the effects of inflation on costs and revenues. We restrict our attention to the case where the fixed replenishment cost A and the unit variable cost v are affected in the same fashion by inflation. A more elaborate model would be necessary if inflation had a different impact on the two types of costs. Moreover, the effects on revenue depend on the choice of pricing policy adopted by the organization (possibly subject to government regulation). In Section 4.6.1, we investigate the case where (selling) price changes are made independent of the replenishment strategy. A special case is where the unit selling price p is assumed to increase continuously at the inflation rate (in the same fashion as A and v). Then, in Section 4.6.2, the result will be shown for the situation where the price is adjusted only once for each replenishment lot. Thus, in this latter case, revenue, as well as costs, now depends on the sizes of the replenishments.

An *exact* analysis in the presence of inflation would be extremely complicated. The reason for this is that costs varying with time, in principle, should lead to the replenishment quantity changing with time. At any point in time, our primary concern is with choosing the value of *only* the very next replenishment. To obtain a tractable analytic result, we assume that all future replenishments will be of the same size as the current replenishment, contrary to what the above discussion would suggest. This assumption is not as serious as it would first appear. In particular, at the time of the next replenishment, we would recompute a *new* value for that replenishment, reflecting any changes in cost (and other) parameters that have taken place in the interim. Second, we employ discounting so that assumptions about future lot sizes should not have an appreciable effect on the current decision. This is an illustration of a heuristic solution procedure, a method, based on analytic and intuitive reasoning, that is not optimal but should result in a reasonable solution. Finally, we do not treat the case where quantity discounts are available. When discounts are available, the best order quantity may be very insensitive to inflation because the breakpoint value often is the best solution over a wide range of parameter values (see Gaither 1981).

4.6.1 Price Established Independent of Ordering Policy

Because the price is established independent of the ordering policy (perhaps pricing is set by market forces external to the company under consideration), we can restrict our attention, as earlier, to the minimization of costs. However, because costs are changing with time, we cannot compare costs over a typical year as in Section 4.2. Instead, we resort to a common tool of investment analysis, namely, the use of the present value (PV) of the stream of future costs.

Suppose that the continuous discount rate is denoted by r ; that is, a cost of c at time t has a PV of ce^{-rt} (e.g., Ross et al. 2008). We denote the continuous inflation rate by i . Thus, if A and v are the cost factors at time zero, then, their values at time t are Ae^{it} and ve^{it} , respectively. With a demand rate of D and an order of size Q , we have time between replenishments of Q/D . At each replenishment, we incur costs of $A + Qv$. If the first replenishment is received at time 0, the PV of

* In some cases, inflation is so high that all cash is quickly used to buy inventories because the increase in the price of the goods exceeds the interest one can earn on invested capital. This section assumes that inflation is at a more reasonable level.

the stream of costs is given by*

$$\begin{aligned}
 PV(Q) &= A + Qv + (A + Qv)e^{iQ/D}e^{-rQ/D} + (A + Qv)e^{2iQ/D}e^{-2rQ/D} + \dots \\
 &= (A + Qv)(1 + e^{-(r-i)Q/D} + e^{-2(r-i)Q/D} + \dots) \\
 &= (A + Qv)\frac{1}{1 - e^{-(r-i)Q/D}}
 \end{aligned}
 \tag{4.14}$$

To find a minimum, we set $dPV(Q)/dQ = 0$. As shown in Section 4A.2 of the Appendix of this chapter, the result is

$$e^{(r-i)Q/D} = 1 + \left(\frac{A}{v} + Q\right)\left(\frac{r-i}{D}\right)$$

For $(r-i)Q/D \ll 1$, which is a reasonable assumption, this equation can be simplified further by approximating the exponential term ($e^x = 1 + x + x^2/2$ for $x \ll 1$). Again, in the Appendix, we obtain the following result:

$$Q_{opt} = \sqrt{\frac{2AD}{v(r-i)}} = EOQ\sqrt{\frac{1}{1-i/r}} \tag{4.15}$$

that is, the EOQ multiplied by a correction factor. Note that, when there is negligible inflation ($i = 0$), this is nothing more than the EOQ. Hence, our justification for using the symbol r for the discount rate. Buzacott (1975) has developed a somewhat more exact decision rule, for which the result of Equation 4.15 is usually a good approximation. See also Jesse et al. (1983), Kanet and Miles (1985), and Mehra et al. (1991) who include a term for storage costs exclusive of the opportunity cost of capital. Sarker and Pan (1994) determine the optimal order quantity and allowable shortages in a finite replenishment inventory system under inflationary conditions. Also, Rachamadugu (1988) incorporates a discounted total cost function.

In an inflationary environment when price is set independent of the ordering policy, the profit per unit increases with time since the last replenishment; thus, intuitively, the replenishment quantity should be larger than under no inflation. Equation 4.15 indicates this type of behavior.

The cost penalty associated with using the EOQ (that ignores inflation) can be found from

$$PCP = \frac{PV(EOQ) - PV(Q_{opt})}{PV(Q_{opt})} \times 100 \tag{4.16}$$

where $PV(Q)$ is given by Equation 4.14. The PCP increases as i approaches r but, as shown in this chapter's Appendix, it reaches a limiting value of $100\sqrt{Ar/2Dv}$, which tends to be quite small, although it should be emphasized that this is a percentage of the *total* cost including purchasing costs, not just ordering and holding costs.

* The last step in this derivation uses the fact that

$$1 + a + a^2 + \dots = \frac{1}{1-a} \text{ for } 0 \leq a < 1$$

Table 4.2 Effects of Inflation on Choice of Order Quantity for 3-Ohm Resistor

<i>i</i>	0	0.02	0.05	0.1	0.2	0.22	0.24
EOQ _{opt}	400	418	450	524	980	1,386	∞
PCP of Using EOQ	0	0.0025	0.021	0.105	0.687	1.26	2.0

4.6.1.1 Numerical Illustration

Let us reconsider the 3-ohm resistor that was used in Section 4.2. Its parameter values were

$D = 2,400$ units/year	$v = \$0.40$ /unit
$A = \$3.20$	$r = 0.24$ \$/\$/year

The EOQ under no inflation was found to be 400 units. The use of Equation 4.15 produces the behavior shown in Table 4.2 for different values of *i*. The percentage cost penalties computed from Equation 4.16 are also presented in Table 4.2 and are seen to be extremely small.

4.6.2 Price Set as a Fixed Fractional Markup on Unit Variable Cost

Let us denote the unit selling price established by the company at time *t* as *p*(*t*). Suppose that a replenishment order is paid for at time *t* at unit variable cost *v*(*t*); the next order will be paid for at time *t* + *T*. Then a fixed fractional markup (*f*) implies

$$p(t + \tau) = v(t)(1 + f) \quad 0 \leq \tau < T \tag{4.17}$$

where *p*(*t* + τ) is the price at time *t* + τ . That is, the same unit price is used throughout the entire time period covered by each replenishment.

v(*t*) depends on the inflation rate as in the previous section; thus, the unit selling price also depends on the inflation rate and on the order quantity *Q*, the latter in that $T = Q/D$ in Equation 4.17. Consequently, both costs and revenues are influenced here by the choice of the order quantity *Q*. An appropriate criterion is to select *Q* in order to maximize the PV of revenues minus costs. Buzacott (1975) uses an analysis similar to that presented* to obtain

$$Q_{opt} = \sqrt{\frac{2AD}{v(r + f)}} = \text{EOQ} \frac{1}{\sqrt{1 + f/r}} \tag{4.18}$$

When the markup for the sale of a whole replenishment lot is set as a fixed fraction of the unit variable cost at the start of a cycle and costs are increasing because of inflation, intuitively the order quantity should be reduced from the EOQ position in order to be able to keep the selling price more closely in line with a markup *f* on the current unit variable cost. Equation 4.18 reflects this

* An added complexity is that the revenue is a continuous function of time that must be discounted.

intuitive behavior. It is seen that, as the product f increases, Q_{opt} becomes smaller relative to the EOQ.

4.7 Limits on Order Sizes

The basic EOQ that we derived earlier considers only the costs of replenishing and carrying the inventory. All the variables used reflect only financial considerations. There are a number of possible physical constraints, not explicitly included in the model, that might prevent the use of the so-called best solution derived from the model. We look at some of these constraints in Sections 4.7.1 through 4.7.3.*

4.7.1 Maximum Time Supply or Capacity Restriction

1. The shelf life (SL) of the commodity may be an important factor. Some items have an SL dictated by perishability, which we investigate more fully in Chapter 9. The SL of other items may be more unusual. For instance, certain computers, generators, and so on have a limited warranty that starts from the date of delivery, and not the date of first use. Previously we derived the EOQ expressed as a time supply as

$$T_{\text{EOQ}} = \frac{\text{EOQ}}{D} = \sqrt{\frac{2A}{Dvr}}$$

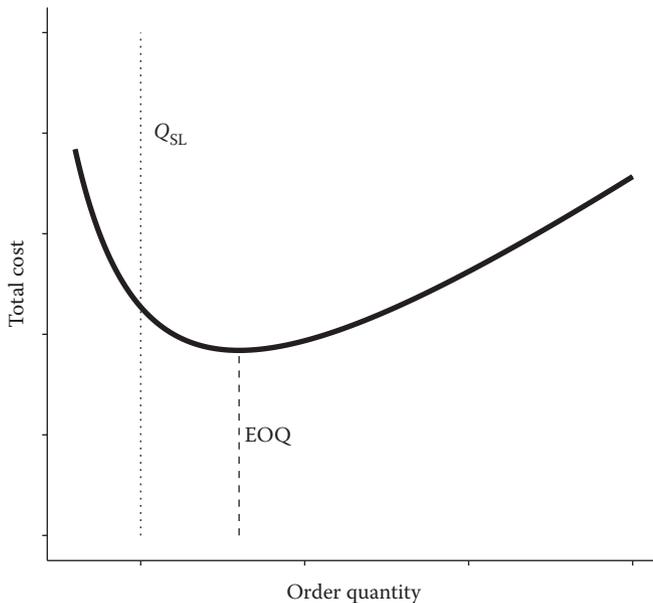


Figure 4.6 Case of an SL constraint.

* Das (1990) provides an algorithm for computing optimal order sizes for several of these cases. See also Moon and Yun (1993) who applies discounted cash flow analysis to the finite horizon case.

As illustrated graphically in Figure 4.6, if this quantity exceeds the allowable shelf life SL , then, the best feasible order quantity to use (i.e., it produces the lowest feasible point on the cost curve) is the SL quantity itself,

$$Q_{SL} = D(SL)$$

2. Even without an SL limitation, an EOQ that represents a very long time supply may be unrealistic for other reasons. A long time supply takes us well out into the future where demand becomes uncertain and obsolescence can become a significant concern. This is an important constraint on class C items, as we will see in Chapter 8. A good example of obsolescence would be an impending engineering change.
3. There may be a storage capacity limitation on the replenishment size for an item (e.g., the entire replenishment must fit in a dedicated storage bin). Equivalently, in production situations (to be discussed in Chapters 10 through 16), there may be a finite production capacity available for the item or a group of related items. A capacity limitation is mathematically equivalent to an SL restriction.

4.7.2 Minimum Order Quantity

The supplier may specify a minimum allowable order quantity. A similar situation exists in an in-house production operation where there is a lower limit on an order quantity that can be realistically considered. If the EOQ is less than this quantity, then, the best allowable order quantity is the supplier or production minimum.

4.7.3 Discrete Units

The EOQ, as given by Equation 4.4, will likely result in a nonintegral number of units. However, it can be shown mathematically (and is obvious from the form of the total cost curve of Figure 4.2) that the best integer value of Q has to be one of the two integers surrounding the best (noninteger) solution given by Equation 4.4. For simplicity, we recommend simply rounding the result of Equation 4.4, that is, the EOQ, to the nearest integer.*

Certain commodities are sold in pack sizes containing more than one unit; for example, many firms sell in pallet-sized loads only, say 48 units (e.g., cases). Therefore, it makes sense to restrict a replenishment quantity to integer multiples of 48 units. In fact, if a new unit, corresponding to 48 of the old unit, is defined, then, we are back to the basic situation of requiring an integer number of (the new) units in a replenishment quantity.

There is another possible restriction on the replenishment quantity that is very similar to that of discrete units. This is the situation where the replenishment must cover an integral number of periods of demand. Again, we simply find the optimal continuous Q expressed as a time supply (see Equation 4.7) and round to the nearest integer value of the time supply.

* Strictly speaking, one should evaluate the total costs, using Equation 4.3, for each of the two surrounding integers and pick the integer having the lower costs. However, rounding produces a trivial percentage cost penalties except possibly when the EOQ is in the range of only one or two units. In addition, the data that are used in Equation 4.3 often have a significantly larger error than the error introduced by rounding to the suboptimal integer.

4.8 Finite Replenishment Rate: The Economic Production Quantity

One of the assumptions inherent in the derivation of the EOQ was that the whole replenishment quantity arrives at the same time. If, instead, we assume that it becomes available at a rate of m per unit time (the production rate of the machinery used to produce the item), then, the sawtoothed diagram of Figure 4.1 is modified to that of Figure 4.7. All that changes from the earlier derivation is the average inventory level that is now $Q(1 - D/m)/2$. The total relevant costs are given by

$$TRC(Q) = \frac{AD}{Q} + \frac{Q(1 - D/m)vr}{2}$$

and the best Q value is now the finite replenishment economic order quantity (FREOQ), also known as the economic production quantity (EPQ) or the economic run quantity.

$$\begin{aligned} \text{FREOQ} &= \sqrt{\frac{2AD}{vr(1 - D/m)}} \\ &= \text{EOQ} \underbrace{\frac{1}{\sqrt{1 - D/m}}}_{\text{correction factor}} \end{aligned} \tag{4.19}$$

As shown in Equation 4.19, the EOQ is multiplied by a correction factor. The magnitude of this correction factor for various values of D/m is shown in Figure 4.8. Also shown are the cost penalties caused by ignoring the correction and simply using the basic EOQ. It is clear that D/m

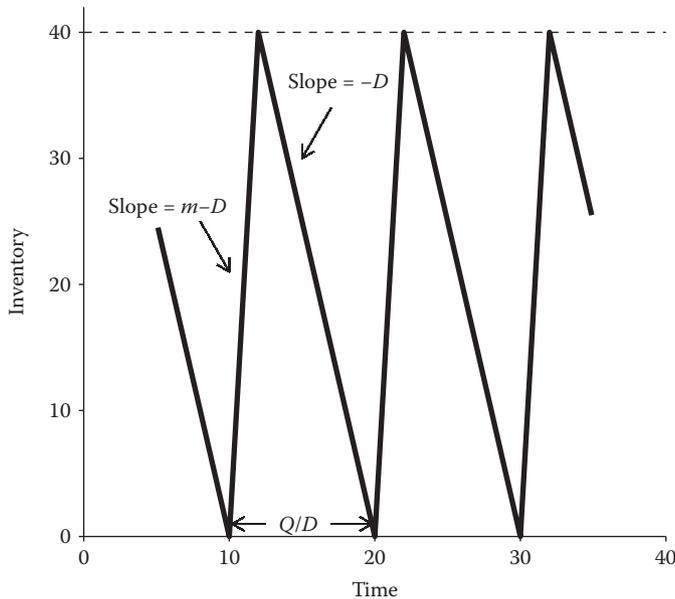


Figure 4.7 Case of a finite replenishment rate.

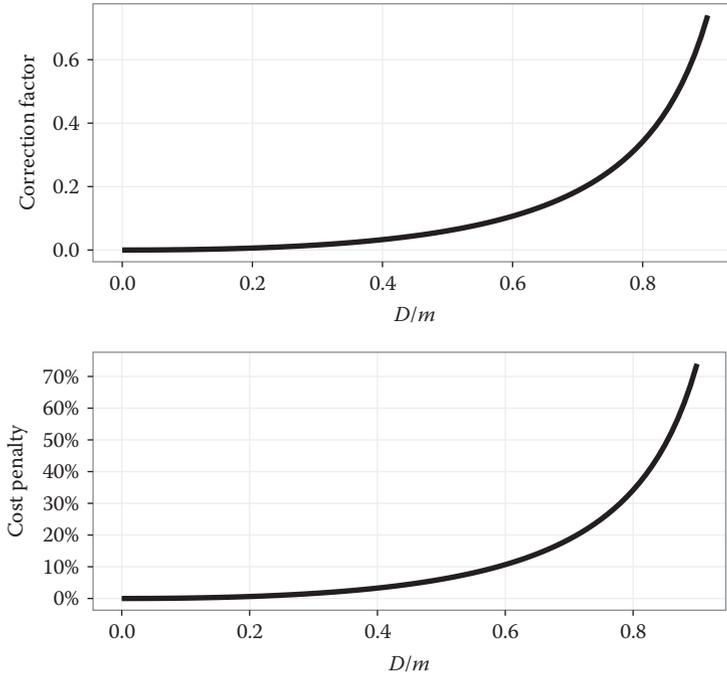


Figure 4.8 Use of EOQ when there is a finite replenishment rate.

has to be fairly large before the cost penalty becomes significant; for example, even when D/m is 0.5 (i.e., the demand rate is one-half of the production rate), the cost penalty for neglecting the finite nature of the production rate is only 6.1%.

Note that as D/m gets very small (equivalent to instantaneous production of the entire lot), the FREOQ reduces to the EOQ as it should. Moreover, as D/m tends to unity, Equation 4.19 suggests that we produce an indefinitely large lot. This makes sense in that demand exactly matches the production rate. For $D/m > 1$, the model breaks down because the capacity of the machine is insufficient to keep up with the demand. However, the model is still useful under these circumstances. To illustrate, suppose that $D = 1.6m$. Here, one would dedicate one machine to the product (with no changeover costs) and produce the remaining $0.6m$ in finite lot sizes on a second machine, where the latter machine's capacity would likely be shared with other items. An example, encountered by one of the authors in a consulting assignment, involved the same fastener manufacturer mentioned earlier in the chapter. One part had enough demand to more than fully use the capacity of one machine. At the time, however, this part was manufactured on several different machines that also made other parts. Managers were working to rearrange the production schedule to dedicate a machine entirely to the part, while producing any excess demand for that part on other shared machines. Further discussion on the sharing of production equipment among several items will be presented in Chapter 10.*

* See Billington (1987), Kim et al. (1992), and Li and Cheng (1994) for “changing the givens” in the EPQ model.

4.9 Incorporation of Other Factors

In this section, we briefly discuss four further modifications of the basic EOQ that result from the relaxation of one or more of the assumptions required in the earlier derivation. The treatment of another important situation—namely, where the quantity supplied does not necessarily exactly match the quantity ordered—can be found in Silver (1976). Other references on this topic can be found in Chapter 7. For the case in which the amount of stock displayed on the shelf influences the demand, see Paul et al. (1996).

4.9.1 Nonzero Constant Lead Time That Is Known with Certainty

As long as demand remains deterministic, the introduction of a known, constant, nonzero replenishment lead time (L) presents no difficulty. The inventory diagram of Figure 4.1 is unchanged. When the inventory level hits DL , an order is placed. It arrives exactly L time units later, just as the inventory hits zero.* If the lead time is longer than the time between replenishments, the order should be placed when the on-hand plus on-order level drops to DL .

The costs are unaltered so that the best order quantity is still given by Equation 4.4. For example, suppose the lead time for the 3-ohm resistor (from Section 4.2) is 3 weeks. The EOQ is 400, and the weekly demand is 48 (based on a 50-week year). When the inventory on hand falls to $48 \times 3 = 144$ units, the order should be placed. Figure 4.9 illustrates this concept. (A nonzero

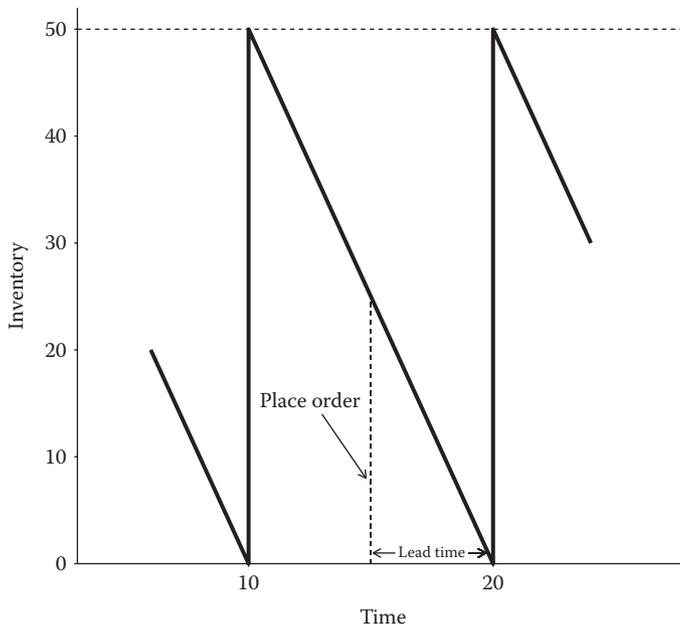


Figure 4.9 Case of a positive lead time.

* If the lead time is longer than the time between replenishments, the order should be placed when the on-hand plus on-order level drops to DL .

lead time considerably complicates matters when demand is probabilistic; and when lead times themselves can be probabilistic, the analysis also gets even more involved. These cases are discussed in Chapter 6.)

4.9.2 Nonzero Payment Period

We have implicitly assumed, by simply looking at average costs per unit time, that payment is immediate upon the receipt of goods. In practice, one can have a period of t years (t , of course, is a fraction) by the end of which the payment must be received. As in Section 4.6, we now must worry about the timing of expenses, so we again use a PV analysis. We assume that the fixed cost, A , is incurred immediately. The PV of repeatedly using an order quantity of size Q is then

$$PV(Q) = A + Qve^{-rt} + PV(Q)e^{-rQ/D}$$

In a manner closely paralleling Section 4.6, we find

$$Q_{\text{opt}} = \sqrt{\frac{2AD}{ve^{-rt}r}} = \text{EOQ}\sqrt{e^{rt}}$$

This result has a simple intuitive interpretation. The unit variable cost v is reduced to ve^{-rt} , because we retain the associated dollars for a time t . Recognizing that rt is likely to be substantially less than unity, we have

$$Q_{\text{opt}} \approx \text{EOQ} \left(1 + \frac{rt}{2} \right)$$

so that the effect of the nonzero payment period is quite small and approximately linear in t . Other treatments of repetitive payment periods are provided by Thompson (1975), Kingsman (1983), and Arcelus and Srinivasan (1993). For the case where the supplier offers a one-time larger payment period ($u > t$), Silver (1999) shows that a simple result holds, namely that the appropriate time supply to use should be $u - t$ years larger than the ongoing best time supply (i.e., Q_{opt}/D). Davis and Gaither (1985) also consider a one-time opportunity to extend the payment.

4.9.3 Different Types of Carrying Charge

In Section 4.2, we assumed that the carrying charge was directly proportional to the average inventory level measured in dollars. We calculated the inventory turnover ratio using similar logic. Modifications to this inventory accounting are possible. Each can be handled with the resulting order quantity expression being somewhat more complex than Equation 4.4.

To illustrate, consider the situation where costs depend on area or volume considerations as well as the value of the inventory. (It is intuitive that it should cost more per unit time to store a dollar of feathers than a dollar of coal.) Suppose that there is a charge of w dollars per unit time per cubic foot of space allocated to an item. Assume that this space must be sufficient to handle the maximum inventory level of the item. One situation where such a change would be appropriate is where items are maintained in separate bins, and a specific bin size (which must house the maximum inventory)

is allocated to each item. The best replenishment quantity under these circumstances is

$$\sqrt{\frac{2AD}{2hw + vr}}$$

where h is the volume (in cubic feet) occupied per unit of the item.

4.9.4 Multiple Setup Costs: Freight Discounts

Consider the situation of an item in which the costs of a replenishment depend on the quantity in the following fashion:

<i>Q</i> Range	Cost
$0 < Q \leq Q_0$	$A + Qv$
$Q_0 < Q \leq 2Q_0$	$2A + Qv$
$2Q_0 < Q \leq 3Q_0$	$3A + Qv$

One interpretation is that the item is produced in some type of container (e.g., a vat or barrel) of fixed capacity Q_0 and there is a fixed cost A associated with the use of each container. Such a fixed cost may arise from a need to clean or repair the vat between uses. Another interpretation relates to transportation costs where Q_0 might be the capacity of a shipping container or rail car. Under the above cost structure, Aucamp (1982) has shown that the best solution is either the standard EOQ or one of the two surrounding integer multiples of Q_0 . Early research incorporating freight cost into EOQ decisions by Baumol and Vinod (1970) highlights the importance of potentially including freight cost in inventory decisions. This paper investigates the trade-off between speed of delivery and cost in an EOQ-type context. Lee (1986) generalizes Aucamp's work by allowing for a cost of A_j for the j th load. Therefore, there is a possible advantage to increasing the number of loads. Knowles and Pantumsinchai (1988) address a situation in which the vendor offers products only in various container sizes. Larger discounts are offered on larger containers. Tersine and Barman (1994) examine incremental and all-units discounts for both purchase price and freight cost. Russell and Krajewski (1991) address the situation in which firms overdeclare the amount of freight shipped in order to reduce the total freight cost. They include a fixed and incremental cost for both purchase and freight. See also Carter and Ferrin (1995). See Gupta (1994) for a treatment of the case in which the ordering cost increases as the replenishment quantity increases. Finally, Van Eijs (1994) examines the problem in a multi-item context with random demands, a problem we will address in Chapter 10. See also Carter et al. (1995a) and Shinn et al. (1996). Thomas and Tyworth (2006), Mendoza and Ventura (2008), and Toptal and Bingöl (2011) investigate the impact of freight costs, specifically with truckload and less-than-truckload (LTL) costs, on the choice of replenishment lot size.

A common theme in much of this work is that the explicit inclusion of freight costs in the EOQ model can dramatically impact the choice of order quantity due to economies of scale in transportation. As an illustration, let us revisit the example from Section 4.2 of the 3-ohm resistor (stable demand $D = 2,400$ units/year, unit variable cost $v = \$0.40$ /unit, fixed cost per replenishment $A = \$3.20$, and holding cost rate $r = 0.24$ \$/\$/year), but now include freight

charges. Furthermore, suppose the transportation provider charges a lower rate per unit for larger orders as follows:

<i>Q</i> Range	<i>Freight Cost per Unit</i>
$0 \geq Q < 250$	\$0.030
$250 \geq Q < 500$	\$0.025
$500 \geq Q < 1,000$	\$0.020
$1,000 \geq Q$	\$0.015

Now that freight costs vary with *Q*, those costs should be included in the total relevant costs per unit time. In a manner quite similar to quantity discounts, these freight cost break points cause discontinuities that must be evaluated as shown in Figure 4.10. When these freight costs are included in this example, the cost minimizing lot size is $Q = 500$; it is worth ordering more than the EOQ of 400 in order to obtain the freight savings.

One nuance worth mentioning is that in some freight settings—LTL carriage for example—carriers* may charge the shipper for more units than are actually shipped if it results in a lower cost.

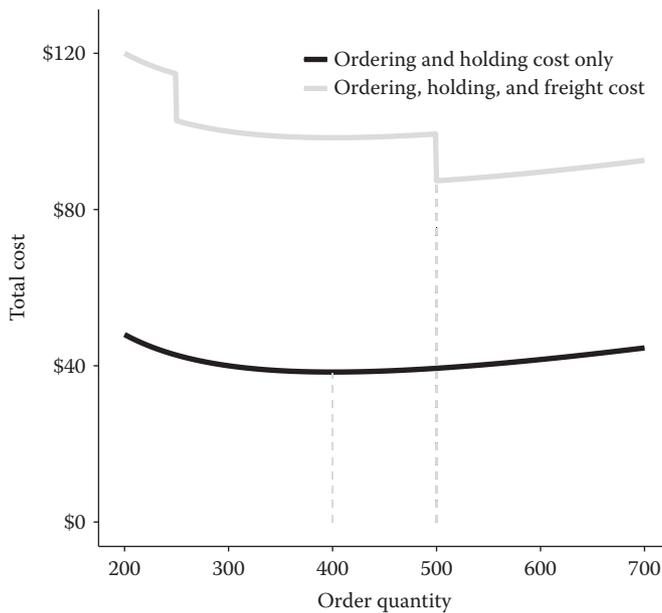


Figure 4.10 Total costs with and without the freight costs included.

* Without getting too distracted with transportation terminology, a *carrier* moves goods while a *shipper* hires a carrier to move goods. Coyle et al. (2015) provide an extensive detail on the modes of transportation including the terminology as well as the pricing structure.

In this case, if the quantity shipped is 225 units, the list price in that range is \$0.03/unit for a cost per shipment of \$6.75 and an annual shipping cost of \$72.00; however, with a quantity of 225, it is better for the shipper to be *charged* as if 250 units were being shipped, giving a cost per shipment of $(250)(\$0.025) = \6.25 and an annual shipping cost of $(\$6.25)(2,400/225) = \66.67 . Our experience suggests that when this situation arises, it is frequently worth increasing the shipping quantity to equal the billed quantity, as is the case in this example.

4.9.5 A Special Opportunity to Procure

An important situation, often faced by both producers and merchants, is a one-time opportunity to procure an item at a reduced unit cost. For example, this may be the last opportunity to replenish before a price rise.

Because of the change in one of the parameters (the unit cost), the current order quantity can differ from future order quantities. (All these future quantities will be identical—namely, the EOQ with the new unit cost.) Thus, the simple sawtooth pattern of Figure 4.1 no longer necessarily applies. Hence, we cannot base our analysis on the average inventory level and the number of replenishments in a typical year (or other units of time). Instead, in comparing the two possible choices for the current value of the order quantity Q , strictly speaking, we should compare costs out to a point in time where the two alternatives leave us in an identical inventory state. To illustrate, suppose that the demand rate is 100 units/month and the new EOQ will be 200 units. Furthermore, suppose that two alternatives under consideration for the current order quantity are 200 and 400 units. As shown in Figure 4.11, an appropriate horizon for cost comparison would be 4 months because, for both alternatives, the inventory level would be zero at that point in time. A comparison of costs out to 1.95 months, for example, would be

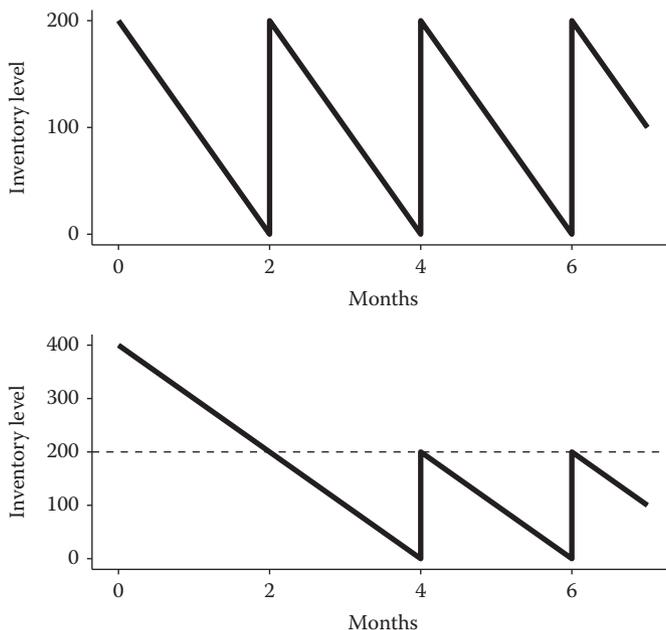


Figure 4.11 Comparison of two alternatives when the order quantity changes with time.

biased in favor of the 200-unit alternative because the imminent (0.05 months later) setup cost under this alternative would not be included. An exact (PV) comparison for all alternatives is rather complex mathematically, particularly when there is a continuum of choices for the order quantity.

Instead, we recommend an approximate approach, suggested by Naddor (1966). Let the decision variable, the current order quantity, be denoted by Q . Let the current unit cost be v_1 and the future unit cost be v_2 ($v_2 > v_1$). The EOQ, after the price rise, is given by Equation 4.4 as

$$\text{EOQ}_2 = \sqrt{\frac{2AD}{v_2 r}} \quad (4.20)$$

Furthermore, from Equation 4.3, the costs per unit time are then

$$\text{TRC}(\text{EOQ}_2) = \sqrt{2ADv_2 r} + Dv_2 \quad (4.21)$$

If the current order quantity is of size Q , then, it will last for Q/D units of time. The average inventory during this period is $Q/2$. Hence, the total costs out to time Q/D are

$$\begin{aligned} \text{TC}(Q) &= A + Qv_1 + \frac{Q}{2}v_1 \frac{Q}{D}r \\ &= A + Qv_1 + \frac{Q^2 v_1 r}{2D} \end{aligned} \quad (4.22)$$

Consider the extreme case where Q was set equal to 0 so that the price increase would occur and we would immediately start ordering the new EOQ_2 . Under this strategy, the total costs out to some time T would be, from Equation 4.21,

$$T\sqrt{2ADv_2 r} + DTv_2$$

It is reasonable to select the value of Q to maximize

$$F(Q) = (Q/D)\sqrt{2ADv_2 r} + Qv_2 - \text{TC}(Q) \quad (4.23)$$

the improvement in the total cost out to time Q/D achieved by ordering Q at the old price instead of not ordering anything at the old price. It can be shown that the use of this criterion need not lead to the exact cost-minimizing solution. Again, we are opting for a heuristic procedure for selecting the value of the decision variable (in this case Q). A convenient approach to maximizing $F(Q)$ is to set $dF(Q)/dQ = 0$. From Equations 4.22 and 4.23

$$\begin{aligned} F(Q) &= \frac{Q}{D}\sqrt{2ADv_2 r} + Qv_2 - A - Qv_1 - \frac{Q^2 v_1 r}{2D} \\ \frac{dF(Q)}{dQ} &= \frac{1}{D}\sqrt{2ADv_2 r} + v_2 - v_1 - \frac{Qv_1 r}{D} = 0 \end{aligned} \quad (4.24)$$

or

$$Q_{\text{opt}} = \frac{\sqrt{2ADv_2 r}}{v_1 r} + \frac{v_2 - v_1}{v_1} \frac{D}{r}$$

that is,

$$Q_{\text{opt}} = \frac{v_2}{v_1} \text{EOQ}_2 + \frac{v_2 - v_1}{v_1} \frac{D}{r} \quad (4.25)$$

The above analysis implicitly assumed that the initial inventory (at the moment that the special opportunity replenishment was to be made) was 0. If instead the inventory was I_0 , then, Lev and Soyster (1979) show that Q_{opt} becomes the result of Equation 4.25 minus I_0 ; that is, Equation 4.25 gives the appropriate order-up-to-level. Taylor and Bradley (1985) generalize to the case of future announced price increases; that is, the firm has time to adjust several order cycles in advance. See also Tersine (1996).

Other analyses of this problem are presented by Brown (1982) and McClain and Thomas (1980). The latter authors, in particular, use a marginal approach that very quickly gives a decision rule. However, the rule, unlike Equation 4.25, is not easily extendible to the case of multiple items, a case to be discussed in a moment. Miltenburg (1987) has shown that the Brown and Naddor decision rules give very similar results on a broad range of problems. Yanasse (1990) arrives at Equation 4.25 through a slightly different approach.*

Lev et al. (1981) extend the analysis to include the possibility of multiple cost or demand parameter changes in two distinct time periods. Ardalan (1994) finds the optimal price and order quantity for a retailer who is offered a temporary price reduction. The retailer's demand increases as its price decreases. Finally, Brill and Chaouch (1995) discuss the EOQ model when a random shock can affect the demand rate. See Goyal et al. (1991) for a review of the literature in this area. See also Aucamp and Kuzdrall (1989), Aull-Hyde (1992, 1996), Tersine and Barman (1995), Arcelus and Srinivasan (1995), Kim (1995), and Moinzadeh (1997).

4.9.5.1 Numerical Illustration

Consider a particular Toronto-based x-ray film dealer supplied by a film manufacturer. Suppose that the manufacturer announces a price increase for a product XMF-082 from \$28.00/box to \$30.00/box. The dealer uses approximately 80 boxes per year and estimates the fixed cost per order to be \$1.50 and the carrying charge as 0.20 \$/\$/year. From Equation 4.20

$$\text{EOQ}_2 = \sqrt{\frac{2 \times 1.50 \times 80}{30.00 \times 0.2}} \approx 6 \text{ boxes}$$

Then, Equation 4.25 gives

$$\begin{aligned} Q_{\text{opt}} &= \frac{30.00}{28.00}(6) + \frac{2.00}{28.00} \frac{80}{0.2} \\ &= 35.3, \text{ say } 35 \text{ boxes} \end{aligned}$$

The increase in unit price from \$28.00/box to \$30.00/box (a 7% increase) causes a one-time procurement quantity of 35 boxes instead of 6 boxes (an increase of almost 500%), a much higher sensitivity than in the basic EOQ itself (see Section 4.3).

* One inventory software package artificially "ages" the inventory by multiplying all reorder points by a factor (say 1.3) when a price increase is announced. Then, orders are placed for all products that have fallen below these new temporary reorder points. If the cost of the orders is too high, the factor is reduced. We will have more to say on reorder points in Chapters 6 and 7.

Substitution of each of $Q = 6$ and $Q = 35$ into Equation 4.24 gives

$$F(6) = \$12.09$$

and

$$F(35) = \$42.23$$

so that the one-time cost savings of using the order quantity of 35 boxes is $(42.23 - 12.09)$ or \$30.14.

Two additional points are worth making:

1. The one-time purchase is likely to represent a large time supply. Obsolescence and other forms of uncertainty may dictate an upper limit on this time supply. Many high-technology firms face rapid obsolescence—the price of their products falls quickly with time. They are therefore very reluctant to carry much stock.
2. The multi-item case, in which a vendor offers a one-time price break on a range of products, is obviously of interest. In such a situation, there is likely to be a constraint on the total amount that can be spent on the one-time purchase or on the total inventory (because of warehouse capacity). For illustrative purposes, let us consider the case of a constraint on the total amount that can be spent. For notation let

- D_i = usage rate of item i , in units/year
- A_i = fixed setup cost of item i , in dollars
- v_{1i} = unit cost of item i in the special opportunity to buy, in \$/unit
- v_{2i} = unit cost of item i in the future, in \$/unit
- Q_i = amount, in units, of item i to be purchased in the special buy
- EOQ $_{2i}$ = economic order quantity of item i in units, under the future unit cost
- n = number of items in the group under consideration (we assume in the sequel that the items are numbered 1, 2, 3, . . . , n)
- W = maximum total amount, in dollars, that can be spent on the special buy

Then, the procedure for selecting the Q'_i 's is as follows:

Step 1 Determine the unconstrained best Q'_i 's by the analog of Equation 4.25, namely,

$$Q_i^* = \frac{v_{2i}}{v_{1i}} \text{EOQ}_{2i} + \frac{v_{2i} - v_{1i}}{v_{1i}} \frac{D_i}{r} \quad i = 1, 2, \dots, n \quad (4.26)$$

Step 2 Calculate the total dollar value of the unconstrained Q'_i 's and compare with the constraining value; thus from Equation 4.26

$$\sum_{i=1}^n Q_i^* v_{1i} = \sum_{i=1}^n [v_{2i} \text{EOQ}_{2i} + (v_{2i} - v_{1i}) D_i / r] \quad (4.27)$$

$\sum_{i=1}^n Q_i^* v_{1i} \leq W$, we use the Q_i^* 's of Step 1. If not, we go to Step 3.

Step 3 Compute the constrained best Q_i values. (The derivation is shown in Section 4A.3 of the Appendix of this chapter):

$$Q_i^* = \frac{v_{2i}}{v_{1i}} \left(\text{EOQ}_{2i} + \frac{D_i}{r} \right) - \frac{D_i}{\sum_{j=1}^n D_j v_{1j}} \times \left[\sum_{j=1}^n v_{2j} \left(\text{EOQ}_{2j} + \frac{D_j}{r} \right) - W \right] \quad i = 1, 2, \dots, n \quad (4.28)$$

If $Q_i^* < 0$, set $Q_i^* = 0$ and solve for all other items.

Note in Equation 4.28 that as the total budget decreases, the term $(D_i W / \sum_{j=1}^n D_j v_{1j})$ decreases linearly. Thus, the item with the highest demand rate, D_i , will have its Q_i^* decrease fastest. This point is also evident from the fact that $(d(Q_i^* v_i) / dW) = -(D_i v_{1i} / \sum_{j=1}^n D_j v_{1j})$. For more discussion on this topic, see Chapter 10, which deals with coordinating the purchase or production of multiple items.

4.10 Selection of the Carrying Charge (r), the Fixed Cost per Replenishment (A), or the Ratio A/r Based on Aggregate Considerations: The Exchange Curve

Often, it is difficult to explicitly determine an appropriate value of the carrying charge r or the fixed cost per replenishment A . An alternate method, which we discussed in Chapter 2, takes into account of the aggregate viewpoint of management. In this and the later sections, we derive the relationships and graphs that we earlier asked you to accept on faith. For a population of inventoried items, the management may impose an aggregate constraint of one of the following forms:

1. The average total inventory cannot exceed a certain dollar value or volume.*
2. The total fixed cost (or total number) of replenishments per unit time (e.g., 1 year) must be less than a certain value.†
3. The maximum allowed backorder delay may not exceed a certain value.
4. The firm should operate at a point where the trade-off (exchange) between the average inventory and the cost (or number) of replenishments per unit time is at some reasonable prescribed value.

We restrict our attention to the case where the fixed cost per replenishment for item i , denoted by A_i , cannot be determined explicitly, but it is reasonable to assume that a common value of A holds (at least approximately) for all items in the portion of the inventory population under consideration. We designate the demand rate, unit variable cost, and order quantity of item i by

* Sometimes, a firm allows its different entities to determine their optimal budgets, and then, it cuts them all by a common factor. Because this “cut across the board” rule is in general not optimal, Rosenblatt and Rothblum (1994) identify the conditions for it to provide globally optimal solutions. See also Pirkul and Aras (1985).

† See Aucamp (1990).

D_i , v_i , and Q_i , respectively. We also let n be the number of items in the population. (The case where the A_i 's are explicitly predetermined can be treated in a similar fashion.)

As shown in Section A4.4 of the Appendix of this chapter, if we use an EOQ for each item, we obtain the total average cycle stock in dollars,

$$\text{TACS} = \sqrt{\frac{A}{r}} \frac{1}{\sqrt{2}} \sum_{i=1}^n \sqrt{D_i v_i} \quad (4.29)$$

and the total number of replenishments per unit time,

$$N = \sqrt{\frac{r}{A}} \frac{1}{\sqrt{2}} \sum_{i=1}^n \sqrt{D_i v_i} \quad (4.30)$$

Both TACS and N depend on the value of the ratio A/r . Multiplication of Equations 4.29 and 4.30 gives

$$(\text{TACS})(N) = \frac{1}{2} \left(\sum_{i=1}^n \sqrt{D_i v_i} \right)^2 \quad (4.31)$$

which is a hyperbola. Moreover, the division of Equation 4.29 by Equation 4.30 gives

$$\frac{\text{TACS}}{N} = \frac{A}{r} \quad (4.32)$$

so that any point on the hyperbolic curve implies a value of A/r . Of course, if either A or r is known explicitly, the implicit value of A/r implies a value of the one remaining unknown parameter.

To summarize, when an EOQ strategy* is used for each item, the management can select a desired point on the tradeoff curve (with associated aggregate conditions), thus implying an appropriate value of r , A , or A/r . The latter parameters can now be thought of as management control variables. In fact, in the late 1980s, a major U.S. automotive firm used a value of 40% for r . When questioned why the value was so high, the reply was, "We want to decrease inventory." In other words, the management artificially increased r so that computerized inventory control routines would force down inventory levels. It should be noted that the use of r as a policy variable is closely related to a technique known as LIMIT (lot-size inventory management interpolation technique) discussed by Eaton (1964). See also Chakravarty (1986).

4.10.1 Exchange Curve Illustration

A small firm stocks just 20 items, all of which are purchased from several suppliers in Germany. Therefore, it is realistic to think in terms of a fixed setup cost A , which does not vary appreciably

* When an aggregate constraint is to be satisfied, it is not obvious that the best strategy for an individual item is to use an EOQ-type order quantity. In fact, the more mathematically oriented reader may wish to verify that, for example, minimization of the total number of replenishments per unit time subject to a specified total average stock (or minimization of total average stock subject to a specified total number of replenishments per unit time) does indeed lead to an EOQ-type formula. If a Lagrange multiplier approach is used, the multiplier turns out to be identical with r/A (or A/r). An introduction to the use of Lagrange multipliers in constrained optimization is provided in Appendix I.

from item to item. Based on a detailed study of costs and policy factors, the management decided to set the value of A at \$2.10. A spreadsheet was used to generate the exchange curve of Figure 4.12. The current operating procedure (estimated by using the most recent replenishment quantity actually used for each of the items in the division) is also shown in Figure 4.12.

It is clear that improvements are possible through the use of EOQs with a suitable value of A/r . (This is because the current logic does not explicitly take the economic factors into account—some inexpensive items are ordered too frequently while some fast movers are not ordered often enough.) At one extreme, by operating at point P (with an implied A/r of 11.07), the same total average cycle stock is achieved as by the current rule but the total number of replenishments per year is reduced from 1700 to 759 (a 55.4% reduction). Point Q (with an implied A/r of 2.21) represents the opposite extreme where the same total number of replenishments is achieved as by the current rule, but the total average cycle stock is cut from \$8,400 to \$3,749 (a 55.4% reduction). Decreases, compared with the current rule, are possible in both aggregate measures at any point between P and Q . Suppose the management has agreed that a total average cycle stock of \$7,000 is a reasonable aggregate operating condition. The associated A/r value (using Equation 4.29) is 7.69. Since A has been explicitly determined, we have that

$$r = A \div A/r = 2.10 \div 7.69 = 0.273\$/\$/\text{year}$$

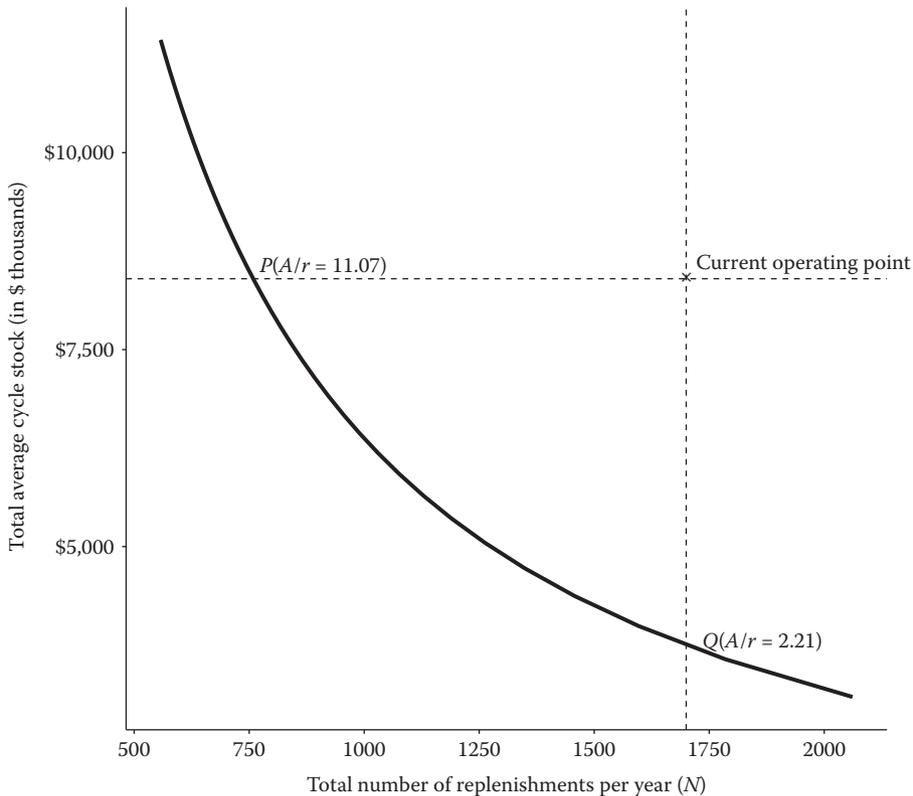


Figure 4.12 Aggregated consequences of different A/r values—an exchange curve.

Rather than using a three-decimal value for r , we choose to round to 0.27 \$/\$/year for convenience in subsequent calculations.

To summarize, exchange curves permit us to show improvements (in terms of aggregate performance measures) over the current operating practices. From a top management perspective, this is far more appealing than the fact that the EOQ minimizes the total costs on an individual item basis (particularly when the individual cost factors are so difficult to estimate). In Section 4A.6 of the Appendix to this chapter, we describe a powerful estimation procedure for exchange curves that is based on the lognormal distribution. (See the Appendix to Chapter 2 as well.)

4.11 Summary

In this chapter, we have specified a procedure for selecting the appropriate size of a replenishment quantity under demand conditions that are essentially static in nature; that is, we assume that the demand rate is at a constant level that does not change with time. Under a rigorous set of additional assumptions, the resulting order quantity is the well-known square root, EOQ.

We argued that the exact values of one or both of the parameters r and A may be very difficult to obtain. In such a case, we showed that a value of r , A , or A/r can be implicitly specified by selecting a desired operating point on an aggregate exchange curve of the total average stock versus the total number or costs of replenishments per year.

Sizable deviations of the order quantity away from the optimum value produce rather small cost errors. This robust feature of the EOQ helps justify its widespread use in practice. Often, the limited savings achievable do not justify the effort required to develop overly sophisticated variations of the EOQ. Nevertheless, as conditions change with time, it becomes necessary to recompute the appropriate value of the order quantity.

Several simple modifications of the basic EOQ have been discussed; these modifications make the results more widely applicable. In the next chapter, we develop a practical procedure for coping with situations where the average or forecast demand rate varies appreciably with time. Three other important extensions will be left to the later chapters, namely,

1. Situations where demand is no longer deterministic (Chapter 6).
2. Situations where savings in replenishment costs are possible through the coordination of two or more items at the same stocking point (Chapter 10).
3. Situations (referred to as multiechelon) where the replenishment quantities of serially related production or storage facilities should not be treated independently (Chapter 11).

In each of these circumstances, we will find that the basic EOQ plays an important role.

Problems

- 4.1 a. For the item used as an illustration in Section 4.2, suppose that A was changed to \$12.80. Now find
 - i. EOQ in units
 - ii. EOQ in dollars
 - iii. EOQ as a months of supply
 Does the effect of the change in A make sense?

- b. For the same item, suppose that A was still \$3.20 but r was increased from 0.24/year to 0.30/year. Find the same quantities requested in part (a). Does the effect of the change in r make sense?
- 4.2 Demand for a product at a depot is at a constant rate of 250 units per week. The product is supplied to the depot from the factory. The factory cost is \$10 per unit, while the total cost of a shipment from the factory to the depot when the shipment has size M is given by

$$\text{Shipment cost} = \$50 + \$2M$$

The inventory carrying cost in \$/week is 0.004. Determine the appropriate shipment size.

- 4.3 A firm can produce and package their own guitar strings, or buy prepacked rolls of “discount” guitar strings. If they load their own strings, there is a production setup cost of \$20. The finished product is valued at \$1.23 a set, and the production rate is 500 sets/day. “Discount” strings cost the firm \$1.26 per set and the fixed ordering charge is \$3/order. In either case, an inventory carrying charge of 0.24 \$/year would be used by the company. Demand for this item is 10,000 sets per year. From the standpoint of replenishment and carrying costs, what should the company do? What other considerations might affect the decision?
- 4.4 A manufacturing firm located in Calgary produces an item in a 3-month time supply. An analyst, attempting to introduce a more logical approach to selecting run quantities, has obtained the following estimates of the characteristics of the item:

$$D = 4,000 \text{ units/year}$$

$$A = \$5$$

$$v = \$4 \text{ per 100 units}$$

$$r = 0.25 \text{ $/year}$$

Note: Assume that the production rate is much larger than D .

- What is the EOQ of the item?
 - What is the time between consecutive replenishments of the item when the EOQ is used?
 - The production manager insists that the $A = \$5$ figure is only a guess. Therefore, he insists on using his simple 3-month supply rule. Indicate how you would find the range of A values for which the EOQ (based on $A = \$5$) would be preferable (in terms of a lower total of replenishment and carrying costs) to the 3-month supply.
- 4.5 Suppose that all the assumptions of the EOQ derivation hold except that we now allow backorders (i.e., we deliberately let the stock level run negative before we order; backordered requests are completely filled out of the next replenishment). Now, there are two decision variables: Q and s (the level below zero at which we place an order).
- Using a sketch and geometrical considerations, find the average on-hand inventory level and the average level of backorders.
 - Suppose that there is a cost B_2v per unit backordered (independent of how long the unit is backordered) where B_2 is a dimensionless factor. Find the best settings of Q and s as a function of A , D , v , r , and B_2 .
 - Repeat the previous part but now with a cost B_3v per unit backordered per unit time. The dimensions of B_3 are equivalent to those for r .

- 4.6 The famous Ernie of “Sesame Street” continually faces replenishment decisions concerning his cookie supply. The Cookie Monster devours the cookies at an average rate of 200 per day. The cookies cost \$0.03 each. Ernie is getting fed up with having to go to the store once a week. His friend, Bert, has offered to do a study to help Ernie with his problem.
- If Ernie is implicitly following an EOQ policy, what can Bert say about the implicit values of the two missing parameters?
 - Suppose that the store offered a special of 10,000 cookies for \$200. Should Ernie take advantage of the offer? Discuss. (Hint: Consult your local TV listing for the timing of and channel selection for “Sesame Street.”)
- 4.7 U. R. Sick Labs manufactures penicillin. Briefly discuss the special considerations required in establishing the run quantity of such an item.
- 4.8 Table 4.1 was developed based on management specifying permitted values of the order quantity expressed as a time supply. Suppose instead that order quantities (in units) were specified.
- Develop an indifference condition between two adjacent permitted values, namely Q_1 and Q_2 .
 - For $A = \$10$ and $r = 0.25$ \$/\$/year, develop a table, similar to Table 4.1 (although the variable involving D and v need not necessarily be Dv), for Q values of 1, 10, 50, 100, 200, 500, and 1,000 units.
 - For an item with $D = 500$ units/year and $v = 0.10$ \$/unit, what is the exact EOQ? What Q does the table suggest using? What is the cost penalty associated with the use of this approximate Q value?
- 4.9 A mining company routinely replaces a specific part on a certain type of equipment. The usage rate is 40 per week and there is no significant seasonality. The supplier of the part offers the following all-units discount structure:

Range of Q	Unit Cost
$0 < Q < 300$	\$10.00
$300 \leq Q$	9.70

The fixed cost of a replenishment is estimated to be \$25 and a carrying charge of 0.26 \$/\$/year is used by the company.

- What replenishment size should be used?
 - If the supplier was interested in having the mining company acquire at least 500 units at a time, what is the largest unit price they could charge for an order of 500 units?
- 4.10 The supplier of a product wants to discourage large quantity purchases. Suppose that all the assumptions of the basic EOQ apply except that a reverse quantity discount is applicable; that is, the unit variable cost is given by

$$v = \begin{cases} v_0 & 0 \leq Q < Q_b \\ v_0(1 + d) & Q_b \leq Q \end{cases} \text{ where } d > 0$$

- a. Write an expression (or expressions) for the total relevant costs per year as a function of the order quantity Q . Introduce (and define) whatever other symbols you feel are necessary.
 - b. Using graphical sketches, indicate the possible positions of the best order quantity (as is done in Figure 4.5 for the regular discount case).
 - c. What is the best order quantity for an item with the following characteristics:
 Demand rate = 50,000 units/year
 Fixed setup cost per replenishment = \$10
 $v = \$1.00/\text{unit}$ $d = 0.005$
 Selling price = \$1.44/unit
 Carrying charge = 0.20 \$/\$/year
 $Q_b = 1,500$ units
- 4.11 A particular product is produced on a fabrication line. Using the finite replenishment EOQ model, the suggested run quantity is 1,000 units. On a particular day of production, this item is being run near the end of a shift and the supervisor, instead of stopping at 1,000 units and having to set up for a new item (which would make her performance look bad in terms of the total output), continues to produce the item to the end of the shift, resulting in a total production quantity of 2,000 units. Discuss how you would analyze whether or not she made a decision that was attractive from a company standpoint. Include comments on the possible intangible factors.
- 4.12 Consider an inventory with only four items with the following characteristics:

Item i	D_i (Units/Year)	v_i (\$/Unit)
1	7,200	2.00
2	4,000	0.90
3	500	5.00
4	100	0.81

The inventory manager vehemently argues that there is no way that he can evaluate A and r ; however, he is prepared to admit that A/r is reasonably constant across the items. He has been following an ordering policy of using a 4-month time supply of each item. He has been under pressure from the controller to reduce the average inventory level by 25%. Therefore, he is about to adopt a 3-month time supply.

- a. Develop an EOQ aggregate exchange curve.
 - b. What are the values of TACS and N for $A/r = 200$?
 - c. What A/r value gives the same TACS as the current rule?
 - d. What A/r value gives the same N as the proposed rule?
 - e. Use the curve to suggest an option or options open to the manager that is (are) preferable to his proposed action.
- 4.13 In a small distribution firm, the controller was advocating a min-max system with specific rules for computing the minimum and maximum levels of each item. Suppose that for each item the values of A , v , D , L , and r could be reasonably accurately ascertained.

- a. Indicate how you would estimate the approximate cost savings of selecting the order quantity of an item through EOQ considerations rather than by the controller's method. (Hint: Assume that the order quantity is approximately equal to the difference between the maximum and minimum levels.)
- b. Illustrate for the following item: $A = \$3.20$; $v = \$10$; $D = 900$ units/year; $L = 10$ weeks; and $r = 0.24$ \$/\$/year.
- 4.14 The Dartmouth Bookstore sells 300 notebooks per term. Each notebook carries a variable cost of \$0.50, while the fixed cost to restock the shelves amounts to \$4.00. The carrying charge of one notebook tied in the inventory is \$0.08 per \$/term.
- Calculate the EOQ.
 - Calculate the total relevant cost associated with the EOQ calculated in the previous part.
 - If the Dartmouth Bookstore instead chooses to order 300 notebooks instead of the EOQ, calculate the PCP of this order size.
 - Calculate the best order quantity now considering a 40% discount on orders greater than 300.
 - Calculate the optimal order quantity if the bookstore decides to incorporate an inflation rate of 2% in the decision.
 - Assume that a particular notebook occupies 0.10 cubic foot of space, for which the bookstore figures that it costs them \$2 per term per cubic foot. Calculate the best replenishment quantity under these circumstances.
 - Now assume that the unit cost of the same notebook increases from \$0.50 to \$0.60. Calculate the new EOQ.
- 4.15 Consider an item whose demand rate is 100 units per year and that costs \$40 per unit. If the fixed cost of ordering is \$50 and the carrying charge is 25%, answer the following:
- Find the EOQ.
 - The purchasing manager is trying to understand the reasons for the high cost of ordering. In this process, she wonders how low A must become for the EOQ to equal 1 unit. What value of A will she find?
- 4.16 An item is produced on a particular machine. The setup time to change over production from some other item to this one is quite large. The machine is run for a single 10-hour shift each day. Changeovers can be done, if desired, in the off-shift time. When a production run of the item is initiated, an obvious decision is how much to produce. How is this decision similar to the one of ordering an item from the supplier where the supplier only sells the item in a rather large pack-size (or container)?
- 4.17 Monopack is a manufacturer of a single product. The company has been concerned about the rising inflation rate in the economy for sometime. Currently, Monopack does not incorporate the effects of inflation into its calculation of its EOQ. The company is interested in finding out if inflation will impact its choice of replenishment sizes. Characteristics of the product were found to be

$$D = 3,000 \text{ units per year}$$

$$A = \$5.45$$

$$v = \$0.60/\text{unit}$$

$$r = 20\%$$

Note: Assume that the product price is established independent of the ordering policy and that inflation affects A and v in the same manner.

- a. Calculate the EOQ in units.
 - b. Calculate the optimum replenishment quantities when the inflation rate is $i = 0.2\%, 4\%, 9\%$, and 15% .
 - c. Calculate the PCP associated with using the EOQ (ignoring inflation) for all values of i .
 - d. What is the limiting value of the PCP?
- 4.18 A firm buys a product from a supplier using the EOQ to determine the order quantity. The supplier recently called and asked the buyer to purchase in larger lots so that the supplier's production process can be made more efficient. Specifically, assume that the buyer will buy k times EOQ units, where $k > 1$. The per unit cost must be reduced to make this attractive to the buyer. What is the fractional discount at which the buyer will be indifferent between the old EOQ at the old price and buying k EOQ at the new price?
- 4.19 A retailer sells about 5,000 units per year of a particular item. The following data has been found for this item:
- $A = \$100$
 $v = \$2$ per unit
 $r = 0.28$ \$/\$/year
- The retailer is now using an order quantity of 500 units. A keen analyst observes that savings could be realized by instead using the EOQ. The retail manager says to forget about mathematics—he has a way of reducing the setup cost to only \$25. He argues for doing this and retaining the current order quantity. Who is right? Discuss.
- 4.20 The Andy Dandy Candy Company has asked you to recommend an EOQ for their Jelly Bean product. They have provided you with the following information:
- a. The only costs involved are the cost of ordering and the cost of holding the inventory that is equal to the average amount of capital times \$0.30. Back orders are not allowed, and there are no other costs.
 - b. The company always has a sale when a shipment of Jelly Beans arrives, until half of the shipment is gone, at which time the company raises its price.
- This implies that there are two known, continuous, constant rates of demand D_2 and D_1 , D_1 being the rate during the sale, with $D_1 > D_2$. As a consultant, you are requested to recommend an EOQ formula for this situation. In your report, you must specifically cover the following points:
- a. A graphical representation of the inventory process. (What is the fraction of time spent at the rate D_1 ?)
 - b. A total cost equation, defining the terms.
 - c. A brief explanation of the method of obtaining the optimal order quantity.
 - d. A numerical example using $A = \$20$; $vr = \$2$; $D_1 = 30$; and $D_2 = 15$.
 - e. How much should be ordered in the above numerical example if $D_1 = D_2 = 30$?
- 4.21 A supplier offers a manufacturer an item at a 5% discount on orders of 500 units or more, and a 15% discount on orders of 1,000 units or more. From past experience, it appears that demand for this item is 3,075 units/year. If the ordering cost is \$50, the basic price (without discount) of the item is \$2, and an r value of 0.20 \$/\$/year is used, how often should the manufacturer order from the supplier? What order quantity should be used?
- 4.22 A supplier offers the following discount structure on purchases of any single item:

$0 < Q < 1,000$	\$5.00 per unit
$1,000 \leq Q < 2,000$	\$4.90 per unit
$2,000 \leq Q$	\$4.75 per unit

The discounts apply to all units. For each of the following items treated separately, what is the appropriate order quantity to use, assuming a common value of $r = 0.30$ \$/\$/year?

Item	D (Units/Year)	A (\$)
1	10,000	25
2	1,000	25
3	4,000	25
4	130,000	25

4.23 A company stocks three items having the following characteristics:

Item i	D_i (Units/Year)	v_i (\$/Unit)	Current Q_i (Units)
1	1,000	5.00	250
2	4,000	0.80	1,000
3	200	4.00	50

- a. Develop an exchange curve of the total average stock (in \$) versus the total number of replenishments per year under an EOQ strategy treating A_i/r as the policy parameter.
 - b. Plot the current operating position on the graph.
 - c. For what values of A_i/r would the EOQ strategy produce an improvement on both aggregate dimensions?
- 4.24 Consider a quantity discount problem in which the form of the discount differs from the all-units case that we have discussed in this chapter. Specifically, the supplier offers an incremental discount such that the lower price applies only to the number of units in the particular discount interval, and not to all units ordered.
- a. Using the notation introduced in Section 4.5, write a total cost expression as a function of the order quantity, Q , for the case of just two discount intervals (the first is the regular price, and the second is the discount price).
 - b. Write a total cost expression for any number, i , discount intervals.
 - c. Define an algorithm for finding the optimal order quantity.
 - d. For the following case, find the optimal order quantity.

$0 < Q < 100$	\$20.00 per unit
$100 \leq Q$	\$19.50 per unit

Also $r = 0.30$ \$/\$/year; $A = \$24$; and $D = 1,200$ units/year.

- 4.25
- Develop a spreadsheet to compute the EOQ and the total relevant cost associated with any order quantity.
 - Repeat part (a) for the case of the all-units quantity discount.
 - Repeat part (a) for the case of the incremental discount described in Problem 4.24.
 - Repeat part (a) for the case of a finite replenishment rate.
 - Repeat part (a) for the EOQ that accounts for an inflation rate.
- 4.26 The May-Bee Company distributes fire-fighting equipment. Approximately 50% of the company's SKUs are imported and the rest are acquired from domestic suppliers. May-Bee sales have been increasing at a steady rate and future trends appear to be very optimistic. The company is wholly owned by a Mr Brown who relies heavily on his bank for working capital. His current credit limit is set at \$150,000 and he is experiencing some resistance from his banker in getting more cash to increase his investment in inventories to meet an anticipated increase in sales. In the past, Brown has used only simple decision rules for ordering from his suppliers. He orders a 3-week supply of SKUs that cost more than \$10 per unit and a 6-week supply of items that cost less than \$10 per unit. Last year, May-Bee's average inventory was around \$1,000,000. He paid \$19,500 interest on his bank loan, \$80,000 rent for his warehouse, and \$15,000 to insurance companies. Brown believes that it would be profitable to increase his bank loan because he feels that he could probably earn 14–15% on the borrowed money. The following six items represent 30% of May-Bee's annual dollar sales volume and according to Brown are his bread and butter:

SKU Number	Value/Unit v (\$)	Annual Demand (Units)
1	15.78	6,400
2 ^a	40.50	2,200
3 ^a	18.30	4,500
4	8.40	3,500
5 ^a	9.70	6,000
6	25.00	4,800

^a Items imported from the same supplier.

Brown gets directly involved with the ordering of imported items because for these items he has to go to the bank to get a letter of credit and on the average spends about 60 minutes per order for doing so. An order for imported items had in the past included from 1 to 10 SKUs. The table below gives a summary of the costs involved in placing an order. These estimates were supplied by the persons involved, who guessed at the amount of time involved in each activity.

Hourly Person Involved	Rate (\$)	Time per Order (Minutes)	
		Domestic ^a	Imported
Typist	3.00	20	24
Inventory clerk	3.20	20	20
Receiver	4.50	30	30
Bookkeeper	4.20	20	20
Mr Brown	20.00	20	60

^a Domestic orders usually included 1–5 SKUs.

Apart from the above, there was a charge of \$10 per order that represented the bank charge for the letter of credit for imported items whereas the bank charges were only \$0.25 per order for items acquired from domestic suppliers.

- The above cost description was compiled by the bookkeeper. In your opinion are there any costs that have been left out? Comment.
- Compute May-Bee's carrying cost (r).
- Compute the cost of placing an order for a single domestic and for a single imported SKU.
- Compute the EOQ for all six items.
- Compute the total cycle stock for the six items under the present ordering rules.
- Compute the total relevant costs for the six items using EOQs.
- Compute the difference in the total relevant costs under 4.26e and 4.26f. Estimate the change in the total relevant costs for all the 100 items.
- Compute the total number of replenishments for the sample of six items under the present system and under the EOQ policy.
- Draw an exchange curve of the total relevant costs versus N for the imported items. From the graph, calculate the increase or decrease in the total relevant costs if the combined total number of orders for the three imported items is changed to 36 per year.
- Compute the total relevant costs (ordering costs plus holding costs) under the present system and under the EOQ policy. What are the estimated savings (percent) if the EOQ policy is applied throughout the whole population?

Appendix 4A: Derivations

4A.1 Penalty for Using an Erroneous Value of the Replenishment Quantity

The percentage cost penalty is

$$\text{PCP} = \frac{\text{TRC}(Q') - \text{TRC}(\text{EOQ})}{\text{TRC}(\text{EOQ})} \times 100 \quad (4A.1)$$

where

$$Q' = (1 + p)\text{EOQ} = (1 + p)\sqrt{\frac{2AD}{vr}}$$

Substituting this Q' expression into the expression for TRC from Equation 4.3, we obtain

$$\begin{aligned}\text{TRC}(Q') &= (1 + p)\sqrt{\frac{ADvr}{2}} + \frac{1}{1 + p}\sqrt{\frac{ADvr}{2}} \\ &= \sqrt{2ADvr}\frac{1}{2}\left(1 + p + \frac{1}{1 + p}\right)\end{aligned}\quad (4A.2)$$

Also from Equation 4.5

$$\text{TRC}(\text{EOQ}) = \sqrt{2ADvr} \quad (4A.3)$$

Substituting the results of Equations 4A.2 and 4A.3 into Equation 4A.1 gives

$$\text{PCP} = \left[\frac{1}{2}\left(1 + p + \frac{1}{1 + p}\right) - 1\right] \times 100$$

or

$$\text{PCP} = \left(\frac{1}{1 + p} - 1 + p\right) = \frac{p^2}{2(1 + p)} \times 100$$

4A.2 Order Quantity under Inflation

From Equation 4.14:

$$\begin{aligned}PV(Q) &= (A + Qv)\frac{1}{1 - e^{-(r-i)Q/D}} \\ \frac{dPV(Q)}{dQ} &= \frac{(1 - e^{-(r-i)Q/D})v - (A + Qv)\left(\frac{r-i}{D}\right)e^{-(r-i)Q/D}}{[1 - e^{-(r-i)Q/D}]^2} = 0\end{aligned}$$

The numerator must be zero, which (dividing through by v) implies

$$1 - e^{-(r-i)Q/D} - \left(\frac{A}{v} + Q\right)\left(\frac{r-i}{D}\right)e^{-(r-i)Q/D} = 0$$

at the optimal order quantity Q . This simplifies to

$$e^{(r-i)Q/D} = 1 + \left(\frac{A}{v} + Q\right)\left(\frac{r-i}{D}\right)$$

Using the second-order Taylor's approximation for e^x at $x = 0$, $e^x \approx 1 + x + x^2/2$, gives

$$1 + \frac{(r-i)Q}{D} + \frac{(r-i)^2 Q^2}{2D^2} \approx 1 + \left(\frac{A}{v}\right)\left(\frac{r-i}{D}\right) + \frac{Q(r-i)}{D}$$

Solving for the optimal quantity (given the approximation to e^x) gives

$$Q_{\text{opt}} = \sqrt{\frac{2AD}{v(r-i)}}$$

From Equations 4.14 and 4.16, the PCP for using the EOQ instead of the Q_{opt} is

$$\text{PCP} = \left[\frac{A + \text{EOQ}v}{A + Q_{\text{opt}}v} \frac{1 - e^{-(r-i)Q_{\text{opt}}/D}}{1 - e^{-(r-i)\text{EOQ}/D}} - 1 \right] \times 100$$

Substituting from Equation 4.15 gives

$$\text{PCP} = \left[\frac{A + \sqrt{\frac{2ADv}{r}}}{A + \sqrt{\frac{2ADv}{(r-i)}}} \left(\frac{1 - e^{-(r-i)\sqrt{2A/(r-i)Dv}}}{1 - e^{-(r-i)\sqrt{2A/rDv}}} \right) - 1 \right] \times 100 \quad (4A.4)$$

As $x \rightarrow 0$, we know that $e^{-x} \cong 1 - x$. Using this result in Equation 4A.4 leads, after some simplification, to

$$\begin{aligned} \text{PCP} &= \left[\frac{\sqrt{\frac{Ar}{2Dv}} + 1}{\sqrt{\frac{Ar}{2Dv}} + \sqrt{\frac{r}{r-i}}} \frac{\sqrt{r}}{\sqrt{r-1}} - 1 \right] \times 100 \\ &= \left[\frac{\sqrt{\frac{Ar}{2Dv}} + 1}{\sqrt{\frac{A(r-i)}{2Dv}} + 1} - 1 \right] \times 100 \end{aligned}$$

As $i \rightarrow r$, we obtain

$$\text{PCP} = \sqrt{Ar/2Dv} \times 100$$

4A.3 Cost Penalty Incurred by Using EOQ When There Is a Finite Replenishment Rate

Recall from Equation 4.19 that the optimal order quantity for the case when the replenishment rate is finite can be expressed as the EOQ times a correction factor of $1/\sqrt{1 - D/m}$. Figure 4.8 plots this correction factor as well as the PCP incurred when using the EOQ. To make this PCP calculation, we first find the percentage deviation in quantity p as follows:

$$\begin{aligned} \text{EOQ} &= (1 + p)\text{FREOQ} \\ \text{EOQ} &= (1 + p)\text{EOQ} \frac{1}{\sqrt{1 - D/m}} \\ \sqrt{1 - D/m} &= (1 + p) \\ p &= \sqrt{1 - D/m} - 1 \end{aligned}$$

Using this value for p , we can apply the percentage deviation in quantity from the PCP formula in Equation 4.10.

4A.4 Multi-Item Special Opportunity to Procure

As in Section 4.9.5, we would like to pick each Q_i so as to maximize

$$\frac{Q_i}{D_i} \sqrt{2A_i D_i v_{2i} r} + Q_i v_{2i} - A_i - Q_i v_{1i} - \frac{Q_i^2 v_{1i} r}{2D_i}$$

However, there is the constraint

$$\sum_{i=1}^n Q_i v_{1i} = W \tag{4A.5}$$

which, given that we have reached Step 3, must be binding.

We use a standard constrained optimization procedure, the Lagrange multiplier approach (see Appendix I at the end of the book). We maximize

$$L(Q_1, \dots, Q_n, M) = \sum_{i=1}^n \left[\frac{Q_i}{D_i} \sqrt{2A_i D_i v_{2i} r} + Q_i v_{2i} - A_i - Q_i v_{1i} - \frac{Q_i^2 v_{1i} r}{2D_i} \right] - M \left(\sum_{i=1}^n Q_i v_{1i} - W \right)$$

where M is a Lagrange multiplier. Taking partial derivatives and setting them to zero

$$\frac{\partial L}{\partial M} = 0$$

gives Equation 4A.5, and

$$\frac{\partial L}{\partial Q_i} = 0 \quad (j = 1, 2, \dots, n)$$

leads to

$$Q_j^* = \frac{v_{2j}}{v_{1j}} \text{EOQ}_{2j} + \frac{v_{2j} - (M + 1)v_{1j}}{v_{1j}} \frac{D_j}{r} \quad j = 1, 2, \dots, n \tag{4A.6}$$

where Q_j^* is the best constrained value of Q_j . Therefore,

$$Q_j^* v_{1j} = v_{2j} \text{EOQ}_{2j} + [v_{2j} - (M + 1)v_{1j}] \frac{D_j}{r} \tag{4A.7}$$

Summing Equation 4A.7 over j and using Equation 4A.5 gives

$$\sum_{j=1}^n v_{2j} \text{EOQ}_{2j} + \sum_{j=1}^n [v_{2j} - (M + 1)v_{1j}] \frac{D_j}{r} = W \tag{4A.8}$$

This is solved for $M + 1$ and the result is substituted into Equation 4A.6, producing Equation 4.28 of the main text.

4A.5 Exchange Curve Equations When A/r Is Constant across All Items

Suppose that we have n items in the population and we designate item i 's characteristics by v_i and D_i . The total average cycle stock in dollars

$$\text{TACS} = \sum_{i=1}^n \frac{Q_i v_i}{2}$$

and the total number of replenishments per unit time

$$N = \sum_{i=1}^n D_i / Q_i$$

If we use the EOQ for each item, we have, from Equation 4.4

$$Q_i = \sqrt{\frac{2AD_i}{v_i r}}$$

Therefore,

$$\text{TACS} = \sum_{i=1}^n \sqrt{\frac{AD_i v_i}{2r}} = \sqrt{\frac{A}{r}} \frac{1}{\sqrt{2}} \sum_{i=1}^n \sqrt{D_i v_i}$$

which is Equation 4.29. Also,

$$N = \sum_{i=1}^n \sqrt{\frac{D_i v_i r}{2A}} = \sqrt{\frac{r}{A}} \frac{1}{\sqrt{2}} \sum_{i=1}^n \sqrt{D_i v_i}$$

which is Equation 4.30.

4A.6 The EOQ Exchange Curve Using the Lognormal Distribution

From the Appendix to Chapter 2, we know that if for a population of n items (numbered $i = 1, 2, \dots, n$) a variable x has a lognormal distribution with parameters m and b , then

$$\sum_{i=1}^n x_i^j = n m^j \exp [b^2 j(j-1)/2] \quad (4A.9)$$

Suppose a quantity of interest (e.g., the cycle stock in dollars) for item i , call it y_i , can be expressed in the form

$$y_i = \sum_{b=1}^H c_b (D_i v_i)^{j_b} \quad (4A.10)$$

where the c_b 's and j_b 's are coefficients and H is a constant, that is, a linear combination of the annual dollar usage raised to various powers. Let

$$Y = \sum_{i=1}^n y_i \quad (4A.11)$$

be the aggregate across the population of the quantity of interest (e.g., the total cycle stock in dollars). Then, because the annual dollar usage is approximately lognormally distributed (with parameters m and b), we have from Equations 4A.9 through 4A.11

$$Y = n \sum_{b=1}^H c_b m^{j_b} \exp[b^2 j_b (j_b - 1)/2] \quad (4A.12)$$

The cycle stock of item i in dollars is

$$\begin{aligned} \frac{Q_i v_i}{2} &= \sqrt{\frac{A_i D_i v_i}{2r}} \\ &= \sqrt{\frac{A}{2r}} (D_i v_i)^{1/2} \end{aligned}$$

assuming the same A for all the items.

This is a special case of Equation 4A.9 with $H = 1$, $c_1 = \sqrt{A/2r}$, and $j_1 = 1/2$. Thus, from Equation 4A.12 that the total cycle stock in dollars is given by

$$\text{TACS} = n\sqrt{A/2r} m^{1/2} \exp(-b^2/8)$$

Suppose for a hypothetical firm we have that $n = 849$, $m = 2,048$, and $b = 1.56$. Thus

$$\begin{aligned} \text{TACS} &= (849)\sqrt{A/2r}(2,048)^{1/2} \exp[-(1.56)^2/8] \\ &= 20,042\sqrt{A/r} \end{aligned} \quad (4A.13)$$

The number of replenishments per year of item i is

$$\begin{aligned} D_i/Q_i &= D_i v_i / Q_i v_i \\ &= D_i v_i / \sqrt{2AD_i v_i / r} \\ &= \sqrt{\frac{r}{2A}} (D_i v_i)^{1/2} \end{aligned}$$

As above, we have that the total number of replenishments per year N is given by

$$N = n\sqrt{r/2A} m^{1/2} \exp(-b^2/8)$$

and for the hypothetical firm

$$N = 20,042\sqrt{r/A} \quad (4A.14)$$

Multiplication of Equations 4A.13 and 4A.14 gives

$$(\text{TACS})N \approx 401,700,000 \quad (4A.15)$$

which is the lognormal approximation of the exchange curve of TACS versus N . The exact exchange curve is

$$(\text{TACS})N = 406,500,000$$

The latter was developed by a complete item-by-item computation. The lognormal approximation of Equation 4A.15 falls a negligible amount above the exact curve—such a small amount that one would have difficulty in discerning any difference between the curves.

References

- Abad, P. L. 1988. Determining optimal selling price and lot size when the supplier offers all-unit quantity discounts. *Decision Sciences* 19, 622–634.
- Arcelus, F. J. and G. Srinivasan. 1993. Delay of payments for extraordinary purchases. *Journal of the Operational Research Society* 44, 785–795.
- Arcelus, F. J. and G. Srinivasan. 1995. Discount strategies for one-time-only sales. *IIE Transactions* 27, 618–624.
- Ardalan, A. 1994. Optimal prices and order quantities when temporary price discounts result in increase in demand. *European Journal of Operational Research* 72, 52–61.
- Aucamp, D. C. 1982. Nonlinear freight costs in the EOQ problem. *European Journal of Operational Research* 9, 61–63.
- Aucamp, D. C. 1990. Aggregate backorder exchange curves. *IIE Transactions* 22(3), 281–287.
- Aucamp, D. C. and P. J. Kuzdrall. 1989. Order quantities with temporary price reductions. *Journal of the Operational Research Society* 40(10), 937–940.
- Aull-Hyde, R. 1992. Evaluation of supplier-restricted purchasing options under temporary price discounts. *IIE Transactions* 24(2), 184–186.
- Aull-Hyde, R. L. 1996. A backlog inventory model during restricted sale periods. *Journal of the Operational Research Society* 47, 1192–1200.
- Axsäter, S. 1996. Using the deterministic EOQ formula in stochastic inventory control. *Management Science* 42(6), 830–834.
- Baumol, W. J. and H. D. Vinod. 1970. An inventory theoretic model of freight transport demand. *Management Science* 16(7), 413–421.
- Benton, W. C. and S. Park. 1996. A classification of literature on determining the lot size under quantity discounts. *European Journal of Operational Research* 92, 219–238.
- Bijvank, M. and I. F. Vis. 2011. Lost-sales inventory theory: A review. *European Journal of Operational Research* 215(1), 1–13.
- Billington, P. J. 1987. The classic economic production quantity model with setup cost as a function of capital expenditure. *Decision Sciences* 18(1), 25–42.
- Bourland, K., S. Powell, and D. Pyke. 1996. Exploiting timely demand information to reduce inventories. *European Journal of Operational Research* 92(2), 239–253.
- Brill, P. H. and B. A. Chaouch. 1995. An EOQ model with random variations in demand. *Management Science* 41(5), 927–936.
- Brown, R. G. 1982. *Advanced Service Parts Inventory Control*. Norwich, VT: Materials Management Systems, Inc.
- Burwell, T. H., D. S. Dave, K. E. Fitzpatrick, and M. R. Roy. 1990. A note on determining optimal selling price and lot size under all-unit quantity discounts. *Decision Sciences* 21, 471–474.
- Buzacott, J. A. 1975. Economic order quantities with inflation. *Operational Research Quarterly* 26(3), 553–558.

- Buzacott, J. A. 2013. Then and now—50 years of production research. *International Journal of Production Research* 51(23–24), 6756–6768.
- Buzacott, J. A. and R. Q. Zhang. 2004. Inventory management with asset-based financing. *Management Science* 50(9), 1274–1292.
- Carlson, M. L., G. J. Miltenburg, and J. J. Rousseau. 1996. Economic order quantity and quantity discounts under date-terms supplier credit: A discounted cash flow approach. *Journal of the Operational Research Society* 47, 384–394.
- Carter, J. R. and B. G. Ferrin. 1995. The impact of transportation costs on supply chain management. *Journal of Business Logistics* 16(1), 189–212.
- Carter, J. R., B. G. Ferrin, and C. R. Carter. 1995. On extending Russell and Krajewski's algorithm for economic purchase order quantities. *Decision Sciences* 26(6), 819–829.
- Chakravarty, A. K. 1981. Multi-item inventory aggregation into groups. *Journal of the Operational Research Society* 32(1), 19–26.
- Chakravarty, A. K. 1986. Quantity discounted inventory replenishments with limited storage space. *INFOR* 24(1), 12–25.
- Crouch, R. and S. Oglesby. 1978. Optimization of a few lot sizes to cover a range of requirements. *Journal of the Operational Research Society* 29(9), 897–904.
- Coyle, J. J., R. A. Novack, and B. Gibson. 2015. *Transportation: A Global Supply Chain Perspective*. Australia: Cengage Learning.
- Dada, M. and K. N. Srikanth. 1987. Pricing policies for quantity discounts. *Management Science* 33, 1247–1252.
- Das, C. 1984. A graphical approach to price-break analysis-3. *Journal of the Operational Research Society* 35(11), 995–1001.
- Das, C. 1988. Some tips for discount buyers. *Production and Inventory Management Journal* 29(2), 23–26.
- Das, C. 1990. An algorithm for selecting quantity discounts from realistic schedules. *Journal of the Operational Research Society* 41(2), 165–172.
- Davis, R. and N. Gaither. 1985. Optimal ordering policies under conditions of extended payment privileges. *Management Science* 31(4), 499–509.
- Dobson, G. 1988. Sensitivity of the EOQ model to parameter estimates. *Operations Research* 36(4), 570–574.
- Dolan, R. J. 1987. Quantity discounts: Managerial issues and research opportunities. *Marketing Science* 6(1), 1–22.
- Donaldson, W. A. 1981. Grouping of inventory items by review period. *Journal of the Operational Research Society* 32(12), 1075–1076.
- Eaton, J. A. 1964. New—The LIMIT technique. *Modern Materials Handling* 19(2), 38–43.
- Erlenkotter, D. 1990. Ford Whitman Harris and the economic order quantity model. *Operations Research* 38(6), 937–946.
- Freeland, J. R., J. P. Leschke, and E. N. Weiss. 1990. Guidelines for setup reduction programs to achieve zero inventory. *Journal of Operations Management* 9(1), 85–100.
- Freimer, M., D. Thomas, and J. Tyworth. 2006. The value of setup cost reduction and process improvement for the economic production quantity model with defects. *European Journal of Operational Research* 173(1), 241–251.
- Gaither, N. 1981. The effects of inflation and fuel scarcity upon inventory policies. *Production and Inventory Management Journal* 22(2), 37–48.
- Gallego, G. and I. Moon. 1995. Strategic investment to reduce setup times in the economic lot scheduling problem. *Naval Research Logistics* 42(5), 773–790.
- Gerchak, Y. and Y. Wang. 1994. Periodic review inventory models with inventory-level-dependent demand. *Naval Research Logistics* 41(1), 99–116.
- Goyal, S. K. 1995. A one-vendor multi-buyer integrated inventory model: A comment. *European Journal of Operational Research* 82, 209–210.
- Goyal, S. K., G. Srinivasan, and F. J. Arcelus. 1991. One time only incentives and inventory policies. *European Journal of Operational Research* 54, 1–6.
- Grant, M. 1993. EOQ and price break analysis in a JIT environment. *Production and Inventory Management Journal* 34(3), 64–69.

- Güder, F., J. Zydiak, and S. Chaudhry. 1994. Capacitated multiple item ordering with incremental quantity discounts. *Journal of the Operational Research Society* 45(10), 1197–1205.
- Gupta, O. K. 1994. An inventory model with lot-size dependent ordering cost. *Production Planning and Control* 5(6), 585–587.
- Harris, F. W. 1913. How many parts to make at once. *Factory, The Magazine of Management* 10, 135–136, 152.
- Harris, F. W. 1990. How many parts to make at once. *Operations Research* 38(6), 947–950.
- Hax, A. C. and D. Candea. 1984. *Production and Inventory Management*. Englewood Cliffs, NJ: Prentice-Hall.
- Hong, J., S. Kim, and J. C. Hayya. 1996. Dynamic setup reduction in production lot sizing with nonconstant deterministic demand. *European Journal of Operational Research* 90, 182–196.
- Jesse Jr., R., A. Mitra, and J. Cox. 1983. EOQ formula: Is it valid under inflationary conditions? *Decision Sciences* 14(3), 370–374.
- Juckler, F. 1970. *Modeles de Gestion des Stocks et Coûts Marginaux*. Louvain, Belgium: Vandeur.
- Kanet, J. J. and J. A. Miles. 1985. Economic order quantities and inflation. *International Journal of Production Research* 23(3), 597–608.
- Katz, P., A. Sadrian, and P. Tendick. 1994. Telephone companies analyze price quotations with Bellcore's PDSS software. *Interfaces* 24(1), 50–63.
- Kim, D. H. 1995. A heuristic for replenishment of deteriorating items with a linear trend in demand. *International Journal of Production Economics* 39, 265–270.
- Kim, K. H. and H. Hwang. 1988. An incremental discount pricing schedule with multiple customers and single price break. *European Journal of Operational Research* 35, 71–79.
- Kim, S. L., J. C. Hayya, and J.-D. Hong. 1992. Setup reduction in the economic production quantity model. *Decision Sciences* 32(2), 500–508.
- Kingsman, B. G. 1983. The effect of payment rules on ordering and stockholding in purchasing. *Journal of the Operational Research Society* 34(11), 1085–1098.
- Knowles, T. W. and P. Pantumsinchai. 1988. All-units discounts for standard container sizes. *Decision Sciences* 19, 848–857.
- Kuzdrall, P. and R. Britney. 1982. Total setup lot sizing with quantity discounts. *Decision Sciences* 13(1), 101–112.
- Lal, R. and R. Staelin. 1984. An approach for developing an optimal discount pricing policy. *Management Science* 30(12), 1524–1539.
- Lee, C. Y. 1986. The economic order quantity for freight discount costs. *IIE Transactions* 18(3), 318–320.
- Lev, B. and A. L. Soyster. 1979. An inventory model with finite horizon and price changes. *Journal of the Operational Research Society* 30(1), 43–53.
- Lev, B., H. Weiss, and A. Soyster. 1981. Optimal ordering policies when anticipating parameter changes in EOQ systems. *Naval Research Logistics Quarterly* 28(2), 267–279.
- Li, C. and T. C. E. Cheng. 1994. An economic production quantity model with learning and forgetting considerations. *Production and Operations Management* 3(2), 118–131.
- Lowe, T. J. and L. B. Schwarz. 1983. Parameter estimation for the EOQ lot size model: Minimax and expected value choices. *Naval Research Logistics* 30, 367–376.
- Lu, L. 1995. A one-vendor multi-buyer integrated inventory model. *European Journal of Operational Research* 81, 312–323.
- Lu, L. and M. Posner. 1994. Approximation procedures for the one-warehouse multi-retailer system. *Management Science* 40(10), 1305–1316.
- Maxwell, W. L. and H. Singh. 1983. The effect of restricting cycle times in the economic lot scheduling problem. *IIE Transactions* 15(3), 235–241.
- McClain, J. O. and L. J. Thomas. 1980. *Operations Management: Production of Goods and Services*. Englewood Cliffs, NJ: Prentice-Hall.
- Mehra, S., S. P. Agrawal, and M. Rajagopalan. 1991. Some comments on the validity of EOQ formula under inflationary conditions. *Decision Sciences* 22, 206–212.
- Mendoza, A. and J. A. Ventura. 2008. Incorporating quantity discounts to the EOQ model with transportation costs. *International Journal of Production Economics* 113(2), 754–765.

- Miltenburg, G. J. 1987. Co-ordinated control of a family of discount-related items. *INFOR* 25(2), 97–116.
- Min, K. J. 1991. Inventory and quantity discount pricing policies under maximization. *Operations Research Letters* 11, 187–193.
- Moinzadeh, K. 1997. Replenishment and stocking policies for inventory systems with random deal offerings. *Management Science* 43(3), 334–342.
- Moinzadeh, K., T. D. Klastorin, and E. Berk. 1997. The impact of small lot ordering on traffic congestion in a physical distribution system. *IIE Transactions* 29(8), 671–680.
- Monahan, J. 1984. A quantity discount pricing model to increase vendor profits. *Management Science* 30(6), 720–726.
- Montgomery, D., M. Bazaraa, and A. Keswani. 1973. Inventory models with a mixture of backorders and lost sales. *Naval Research Logistics Quarterly* 20(2), 255–263.
- Moon, I. and W. Yun. 1993. An economic order quantity model with a random planning horizon. *The Engineering Economist* 39(1), 77–86.
- Moon, I. and G. Gallego. 1994. Distribution free procedures for some inventory models. *Journal of Operational Research Society* 45(6), 651–658.
- Naddor, E. 1966. *Inventory Systems*. New York: John Wiley & Sons.
- Neves, J. S. 1992. Average setup cost inventory model: Performance and implementation issues. *International Journal of Production Research* 30(3), 455–468.
- Parlar, M. and Q. Wang. 1994. Discounting decisions in a supplier–buyer relationship with a linear buyer’s demand. *IIE Transactions* 26(2), 34–41.
- Paul, K., T. K. Datta, K. S. Chaudhuri, and A. K. Pal. 1996. An inventory model with two-component demand rate and shortages. *Journal of the Operational Research Society* 47, 1029–1036.
- Pirkul, H. and O. Aras. 1985. Capacitated multiple item ordering problem with quantity discounts. *IIE Transactions* 17(3), 206–211.
- Porteus, E. L. 1985. Investing in reducing setups in the EOQ model. *Management Science* 31(8), 998–1010.
- Porteus, E. L. 1986a. Investing in new parameter values in the discounted EOQ model. *Naval Research Logistics Quarterly* 33(1), 39–48.
- Porteus, E. L. 1986b. Optimal lot sizing, process quality improvement and setup cost reduction. *Operations Research* 34(1), 137–144.
- Rachamadugu, R. 1988. Error bounds for EOQ. *Naval Research Logistics* 35, 419–425.
- Rosenblatt, M. J. and H. L. Lee. 1985. Improving profitability with quantity discounts under fixed demand. *IIE Transactions* 17(4), 388–395.
- Rosenblatt, M. J. and U. G. Rothblum. 1994. The optimality of the “cut across the board” rule applied to an inventory model. *IIE Transactions* 26(2), 102–108.
- Ross, S. A., R. Westerfield, and B. D. Jordan. 2008. *Fundamentals of Corporate Finance*. New York: Tata McGraw-Hill Education.
- Russell, R. M. and L. J. Krajewski. 1991. Optimal purchase and transportation cost lot sizing for a single item. *Decision Sciences* 22(4), 940–952.
- Sarker, B. and H. Pan. 1994. Effects of inflation and the time value of money on order quantity and allowable shortage. *International Journal of Production Economics* 34, 65–72.
- Sethi, S. P. 1984. A quantity discount model with disposals. *International Journal of Production Research* 22(1), 31–39.
- Shinn, S. W., H. Hwang, and S. S. Park. 1996. Joint price and lot size determination under conditions of permissible delay in payments and quantity discounts for freight cost. *European Journal of Operational Research* 91, 528–542.
- Silver, E. A. 1976. Establishing the order quantity when the amount received is uncertain. *INFOR* 14(1), 32–39.
- Silver, E. A. 1992. Changing the givens in modelling inventory problems: The example of just-in-time systems. *International Journal of Production Economics* 26, 347–351.
- Silver, E. A. 1999. Simple replenishment rules when payment periods (dating deals) are offered. *Production and Inventory Management Journal* 40(4), 1.
- Singhal, V. R. 1988. Inventories, risk, and the value of the firm. *Journal of Manufacturing and Operations Management* 1, 4–43.

- Spence, A. M. and E. L. Porteus. 1987. Setup reduction and increased effective capacity. *Management Science* 33(10), 1291–1301.
- Taylor, S. and C. Bradley. 1985. Optimal ordering strategies for announced price increases. *Operations Research* 33(2), 312–325.
- Tersine, R. and S. Barman. 1994. Optimal lot sizes for unit and shipping discount situations. *IIE Transactions* 26(2), 97–101.
- Tersine, R. J. 1996. Economic replenishment strategies for announced price increases. *European Journal of Operational Research* 92, 266–280.
- Tersine, R. J. and S. Barman. 1995. Economic purchasing strategies for temporary price discounts. *European Journal of Operational Research* 80, 328–343.
- Thomas, D. J. and J. E. Tyworth. 2006. Pooling lead-time risk by order splitting: A critical review. *Transportation Research Part E: Logistics and Transportation Review* 42(4), 245–257.
- Thompson, H. 1975. Inventory management and capital budgeting: A pedagogical note. *Decision Sciences* 6, 383–398.
- Toptal, A. and S. O. Bingöl. 2011. Transportation pricing of a truckload carrier. *European Journal of Operational Research* 214(3), 559–567.
- Van Beek, P. and C. Van Putten. 1987. OR contributions to flexibility improvement in production/inventory systems. *European Journal of Operational Research* 31, 52–60.
- Van der Duyn Schouten, F. A., M. J. G. van Eijs, and R. Heuts. 1994. The value of supplier information to improve management of a retailer's inventory. *Decision Sciences* 25(1), 1–14.
- Van Eijs, M. J. G. 1994. Multi-item inventory systems with joint ordering and transportation decisions. *International Journal of Production Economics* 35, 285–292.
- Weng, Z. K. and R. T. Wong. 1993. General models for the supplier's all-unit quantity discount policy. *Naval Research Logistics (NRL)* 40(7), 971–991.
- Weng, Z. K. 1995a. Channel coordination and quantity discounts. *Management Science* 41(9), 1509–1522.
- Weng, Z. K. 1995b. Modeling quantity discounts under general price-sensitive demand functions: Optimal policies and relationships. *European Journal of Operational Research* 86(2), 300–314.
- Wilson, R. 1993. *Nonlinear Pricing*. New York: Oxford University Press.
- Yanasse, H. H. 1990. EOQ systems: The case of an increase in purchase cost. *Journal of the Operational Research Society* 41(7), 633–637.
- Zheng, Y. S. 1992. On properties of stochastic inventory systems. *Management Science* 38(1), 87–103.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Chapter 5

Lot Sizing for Individual Items with Time-Varying Demand

In Chapter 4, we developed the economic order quantity (EOQ) (and various modifications of it) by essentially assuming a *level, deterministic* demand rate. In this chapter, we relax this assumption and allow the average demand rate to vary with time. In a manner similar to Chapter 4, here, we focus on settings where replenishment quantities are not constrained by capacity. We analyze the implications of finite capacity extensively in Chapters 13 through 15. As compared to the EOQ setting, the time-varying demand situations addressed in this chapter encompass a broader range of practical situations, including:

1. Multiechelon assembly operations where a firm schedule of finished products, exploded back through the various assembly stages, leads to production requirements at these earlier levels. These requirements are relatively deterministic but almost always vary appreciably with time. (This is a topic to which we return in Chapter 15 under the heading of Material Requirements Planning, or Material Resources Planning [MRP].)
2. Production to contract, where the contract requires that certain quantities have to be delivered to the customer on specified dates.
3. Items having a seasonal demand pattern. (Artificial seasonality can be induced by pricing or promotion actions.)
4. Replacement parts for an item that is being phased out of operation. Here, the demand rate drops off with time. (In some cases, particularly toward the end of its life, such an item is better treated by the Class C procedures to be discussed in Chapter 8.)
5. More generally, items with known patterns in demand that are expected to continue.
6. Parts for preventive maintenance where the maintenance schedule is accurately known.

This topic has been termed *lot sizing* in the academic and practitioner literatures.

In Section 5.1, we point out how much more complex the analysis becomes when we allow the demand rate to vary with time. Section 5.2 introduces three different approaches, namely: (1) straightforward use of the EOQ (even though one of the key assumptions on which it is based is violated), (2) an exact optimal procedure, and (3) an approximate heuristic method. In Section 5.3, the assumptions common to all three approaches are laid out, as well as a numerical

example to be used for illustrative purposes throughout Sections 5.4 through 5.6. Further details on each of the three approaches are presented in Sections 5.4, 5.5, and 5.6, including a discussion of when to use each method. In Section 5.7, we describe a procedure for handling quantity discounts when demand varies with time. Then, in Section 5.8, we again focus our attention on aggregate considerations by means of the exchange curve concept.

5.1 The Complexity of Time-Varying Demand

When the demand rate varies with time, we can no longer assume that the best strategy is always to use the same replenishment quantity; in fact, this will seldom be the case. (A similar difficulty arose in Chapter 4 when economic factors changed with time, such as the conditions of inflation or a special opportunity to buy.) An exact analysis becomes very complicated because the diagram of the inventory level versus time, even for a constant replenishment quantity, is no longer the simple repeating sawtooth pattern that we saw in Figure 4.1 of Chapter 4. This prevents us from using simple average costs over a typical unit period as was possible in the EOQ derivation. Instead, we now have to use the demand information over a finite period, extending from the present, when determining the appropriate value of the current replenishment quantity. This period is known as the planning *horizon* and its length can have a substantial effect on the total relevant costs of the selected strategy. Moreover, all other factors being equal, we would prefer to have the planning horizon as short as possible; the farther into the future we look for demand information, the less accurate it is likely to be.

It should be kept in mind, however, that in actual applications, a rolling schedule procedure is almost always used. Specifically, although replenishment quantities may be computed over the entire planning horizon, *only the imminent decision is implemented*. Then, at the time of the next decision, new demand information is appended to maintain a constant length of horizon. This topic is discussed in depth in Chapter 15 when we present one of the primary applications of lot sizing, or materials resources planning (MRP).

As indicated by the examples of time-varying demand discussed earlier, the demand can be either continuous with time or can occur only at discrete equispaced points in time. The former represents a stream of small-sized demands where the arrival rate varies with time, whereas the latter corresponds to something such as weekly shipments where a large demand quantity is satisfied at one time. The decision system that we propose is not affected by whether the demand is continuous or discrete with respect to time; all that will be needed is the total demand in each basic period. A common case is one in which the demand rate stays constant throughout a period, only changing from one period to another. An illustration is shown in Figure 5.1. As discussed in Chapter 3, demand forecasts are usually developed in this form, for example, a rate in January, another rate in February, and so on.

Another element of the problem that is important in selecting the appropriate replenishment quantities is whether replenishments must be scheduled at specified discrete points in time (e.g., replenishments can only be scheduled at intervals that are integer multiples of a week), or whether they can be scheduled at any point in continuous time. Considering a single item in isolation, the second case is probably more appropriate if the demand pattern is continuous with time. However, to ease scheduling coordination problems in multi-item situations, it usually makes sense to limit the opportunity times for the replenishment of each item. Furthermore, if the demand pattern is such that all the requirements for each period must be available at the beginning of that period, it is appropriate to restrict the replenishment opportunities to the beginnings

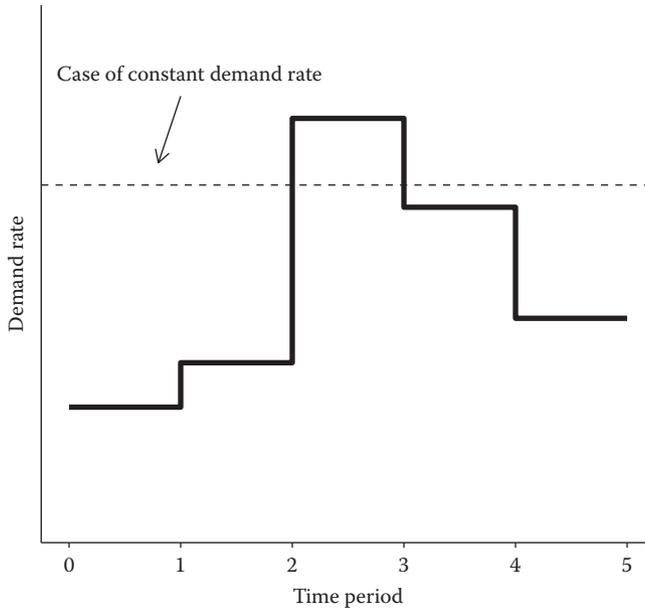


Figure 5.1 Demand pattern when the rate stays constant through each period.

of periods. This latter assumption will be made in developing the logic of the decision system that we propose. (The case where replenishments can take place at any time has been treated by Silver and Meal 1969.)

Still another factor that can materially influence the logic in selecting the replenishment quantities is the duration of the demand pattern. A pattern with a clearly specified end resulting from a production contract is very different from the demand situation where no well-defined end is in sight. In the former case, it is important to plan to reach the end of the pattern with a low (if not zero) level of the inventory. In contrast, for the continuing demand situation, if we use a finite planning horizon, there is no need to force the solution to include a low inventory level at the end of the horizon. The remaining inventory can be used in the interval beyond the horizon, where demand will continue (albeit, at a possibly different rate).

5.2 The Choice of Approaches

There are essentially three approaches to dealing with the case of a deterministic, time-varying demand pattern:

1. *Use of the basic EOQ.* This is a very simple approach. Use a fixed EOQ, based on the average demand rate out to the horizon, anytime a replenishment is required. As would be expected, this approach makes sense when the variability of the demand pattern is low; that is, the constant demand rate assumption of the fixed EOQ is not significantly violated.
2. *Use of the exact best solution to a particular mathematical model of the situation.* As we will see, under a specific set of assumptions, this approach, known as the Wagner–Whitin algorithm,

minimizes the total of certain costs. (For reasons that will become evident later in the chapter, we purposely avoid the use of the words “total relevant costs” here.)

3. *Use of an approximate or heuristic method.* The idea here is to use an approach that captures the essence of the time-varying complexity but at the same time remains relatively simple for the practitioner to understand, and does not require lengthy computations.

Many MRP systems offer a variety of approaches, although the most common is the EOQ, or some other fixed order quantity. These systems do not typically recommend an approach. Rather, the user must specify which technique to use. The discussion in this chapter, as well as in Chapter 15, should help inform these decisions.

5.3 General Assumptions and a Numerical Example

5.3.1 The Assumptions

We now specify a number of assumptions that will be made (at least implicitly) in all three of these approaches. Certain of these assumptions can be relaxed by somewhat minor modifications of the decision methodology; such modifications will be demonstrated later.

1. The demand rate is given in the form of $D(j)$ to be satisfied in period j ($j = 1, 2, \dots, N$), where the planning horizon is at the end of period N . Of course, the demand rate may vary from one period to the next, but it is assumed as known. The case of probabilistic, time-varying demand will be addressed in Chapter 7.
2. The entire requirements of each period must be available at the beginning of that period. Therefore, a replenishment arriving partway through a period cannot be used to satisfy that period's requirements; it is cheaper, in terms of the reduced carrying costs, to delay its arrival until the start of the next period. Thus, replenishments are constrained to arrive at the beginnings of periods.
3. The unit variable cost does not depend on the replenishment quantity; in particular, there are no discounts in either the unit purchase cost or the unit transportation cost.
4. The cost factors do not change appreciably with time; in particular, inflation is at a negligibly low level.
5. The item is treated entirely independently of other items; that is, the benefits from joint review or replenishment do not exist or are ignored.
6. The replenishment lead time is known with certainty (a special case being zero duration) so that delivery can be timed to occur right at the beginning of a period.
7. No shortages are allowed.
8. The entire order quantity is delivered at the same time.
9. For simplicity, it is assumed that the carrying cost is only applicable to the inventory that is carried over from one period to the next. It should be emphasized that all the three approaches can easily handle the situation where carrying charges are included on the material during the period in which it is used to satisfy the demand requirements but, for practical purposes, this is an unnecessary complication.

These assumptions, except for 1, 2, and 9, are identical with those required in the derivation of the basic EOQ (as seen in Section 4.1 of Chapter 4). There are problems associated with these

assumptions, as discussed in Chapter 4. Nevertheless, the methods described below provide a good starting point for most of the useful techniques.

Because of the assumption of deterministic demand, it is clear, from assumptions 2, 6, and 7, that an appropriate method of selecting replenishment quantities should lead to the arrival of replenishments only at the beginning of periods when the inventory level is exactly zero.

5.3.2 A Numerical Example

A plant in Germany uses the following simple decision rule for ascertaining production run quantities: “Each time a production run is made, a quantity sufficient to satisfy the total demand in the next 3 months is produced.” A Canadian subsidiary of the German firm is the only customer of the German plant for a seasonal product PSF-007, a 25-by-30-cm lithographic film. The basic unit of the product is a box of 50 sheets of film. Canadian requirements by month (shifted to take into account of the shipping time) in the upcoming year are given in Table 5.1.*

We can see that the demand pattern has two peaks, one in late spring, and the other in autumn.

Table 5.1 Monthly Requirements

<i>Month</i>	<i>Sequential Number</i>	<i>Requirements (Boxes)</i>
January	1	10
February	2	62
March	3	12
April	4	130
May	5	154
June	6	129
July	7	88
August	8	52
September	9	124
October	10	160
November	11	238
December	12	41
		Total = 1,200

* These requirements are often termed “net requirements” because any inventory on hand and on order is “netted out” before determining the requirements. We shall assume that there is no inventory on hand or on order. Equivalently, we could assume that any initial inventory has been “netted out.” For example, an initial inventory level of 50 and January requirements of 60 would produce the same net requirements shown in Table 5.1.

Table 5.2 Results of Using the Company's "Three Month" Rule on the Numerical Example

Month	1	2	3	4	5	6	7	8	9	10	11	12	Total
Starting inventory	0	74	12	0	283	129	0	176	124	0	279	41	–
Replenishment	84	–	–	413	–	–	264	–	–	439	–	–	1,200
Requirements	10	62	12	130	154	129	88	52	124	160	238	41	1,200
Ending inventory	74	12	0	283	129	0	176	124	0	279	41	0	1,118
Total replenishment costs = $4 \times \$54 = \216.00													
Total carrying costs = $1,118 \text{ box-months} \times \$20/\text{box} \times 0.02 \text{ \$/\$/month} = \447.20													
Total replenishment plus carrying costs = $\$663.20$													

The German plant estimates the fixed setup cost (A) per replenishment to be approximately \$54, and the carrying charge (r) has been set by the management at 0.02 \$/\$/month. The unit variable cost (v) of the film is \$20/box.*

The production-planning department of the German plant would like to establish the size of the first production run (that was needed by January 1), and also estimate the timing and sizes of future production quantities. The word "estimate" is used because the Canadian subsidiary could conceivably revise its requirements, particularly those late in the year, but still give adequate time to permit adjustment of the later portions of the production schedule.

The use of the company's "3 month" decision rule leads to the replenishment schedule and the associated costs shown in Table 5.2. There are four replenishments covering the 12 months out to the horizon. The total relevant costs are \$663.20. The average month-ending inventory is 1,118/12 or 93.17 boxes. Therefore, the turnover ratio is 1,200/93.17 or 12.9.

5.4 Use of a Fixed EOQ

When the demand rate is approximately constant, we have advocated, in Chapter 4, the use of the basic EOQ. One possible approach to the case of a time-varying rate is to simply ignore the time variability, thus continuing to use the EOQ. (An alternative, closely related, approach is to use a fixed time supply equal to the EOQ time supply, T_{EOQ} . This approach will be discussed further in Section 5.6.2.) To be more precise, the average demand rate (\bar{D}) out to the horizon (N periods), or to whatever point our forecast information extends, is evaluated and the EOQ

$$EOQ = \sqrt{\frac{2A\bar{D}}{vr}}$$

* See Rachamadugu and Schriber (1995) for an analysis of "changing the givens"—specifically, finding the best run size when setup times decrease because of learning. See also Diaby (1995) and Tzur (1996).

is used anytime a replenishment is needed. Actually \bar{D} can simply be based on a relatively infrequent estimate of the average demand per period and need not necessarily be reevaluated at the time of each replenishment decision. *To account for the discrete opportunities to replenish, at the time of a replenishment the EOQ should be adjusted to exactly satisfy the requirements of an integer number of periods.* A simple way to do this is to keep accumulating periods of requirements until the closest total to the EOQ is found. As discussed in Section 5.1, and as we will see in more detail in Section 5.5, it is desirable to plan to receive a replenishment order when the inventory is low (or zero) as this keeps inventory holding costs down. Alternatively, the quantity can be fixed, ensuring that the factory floor always sees the same production quantity. While this will lead to an increase in inventory holding costs, there may be efficiencies in production, packaging, shipping, or similar activities associated with always producing the same quantity.

To illustrate, for our numerical example,

$$\bar{D} = \frac{\text{Total requirements}}{12 \text{ months}} = 100 \text{ boxes/month}$$

Therefore,

$$\text{EOQ} = \sqrt{\frac{2 \times \$54 \times 100 \text{ boxes/month}}{\$20/\text{box} \times 0.02\$/\$/\text{month}}} \approx 164 \text{ boxes}$$

Consider the selection of the replenishment quantity at the beginning of January. The following table is helpful:

	January	February	March	April
Requirements	10	62	12	130
Cumulative requirements to the end of the month	10	72	84	214

The EOQ of 164 boxes lies between 84 and 214, and 214 is closer to 164 than is 84. Therefore, the first replenishment quantity is 214 boxes, lasting through the end of April. The detailed results of applying the fixed EOQ approach to the numerical example are as shown in Table 5.3. We can see from Tables 5.2 and 5.3 that the fixed EOQ approach, compared to the company’s “three month” rule, reduces the total of replenishment and carrying costs from \$663.20 to \$643.20 or about 3%, a rather small saving. The turnover ratio has been increased from 12.9 to $1,200 \times 12/528$ or 27.3, but the replenishment costs have also increased.

5.5 The Wagner–Whitin Method: An “Optimal” Solution under an Additional Assumption

Wagner and Whitin (1958), in a classic article, developed an algorithm that guarantees an optimal (in terms of minimizing the total costs of replenishment and carrying inventory) selection of

Table 5.3 Results of Using the Fixed EOQ Approach on the Numerical Example

Month	1	2	3	4	5	6	7	8	9	10	11	12	Total
Starting inventory	0	204	142	130	0	0	0	52	0	0	0	0	–
Replenishment	214	–	–	–	154	129	140	–	124	160	238	41	1,200
Requirements	10	62	12	130	154	129	88	52	124	160	238	41	1,200
Ending inventory	204	142	130	0	0	0	52	0	0	0	0	0	528
Total replenishment costs = $8 \times \$54 = \432.00													
Total carrying costs = $528 \text{ box-months} \times \$20/\text{box} \times 0.02 \text{ \$/\$/month} = \211.20													
Total replenishment plus carrying costs = $\$643.20$													

replenishment quantities under the set of assumptions that were listed in Section 5.3.1.* They also identified one additional assumption that may be needed—namely, that either the demand pattern terminates at the horizon or else the ending inventory must be prespecified.†

5.5.1 The Algorithm

The algorithm is an application of dynamic programming, a mathematical procedure for solving sequential decision problems.‡

The computational effort, often prohibitive in dynamic programming formulations, is significantly reduced because of the use of two key properties (derived by Wagner and Whitin) that the optimal solution must satisfy:

Property 1 A replenishment only takes place when the inventory level is zero (previously discussed in Section 5.3).

Property 2 There is an upper limit to how far before a period j we would include its requirements, $D(j)$, in a replenishment quantity. Eventually, the carrying costs become so high that it is less expensive to have a replenishment arrive at the start of period j than to include its requirements in a replenishment from many periods earlier.

Suppose we define $F(t)$ as the total costs of the best replenishment strategy that satisfies the demand requirements in periods $1, 2, \dots, t$. To illustrate the procedure for finding $F(t)$ and the

* In this chapter, we focus on the case where demand is deterministic. See for example, Sox et al. (1997), Vargas (2009), and Kang and Lee (2013) for extensions to the stochastic demand setting.

† The Appendix to this chapter contains additional details of the Wagner–Whitin model, including an integer-programming formulation as well as the dynamic programming model.

‡ The problem here is a sequential decision problem because the outcome of the replenishment quantity decision at one point has effects on the possible replenishment actions that can be taken at later decision times; for example, whether or not we should replenish at the beginning of March depends very much on the size of the replenishment quantity at the beginning of February. Vidal (1970) provides an equivalent linear-programming formulation and an associated efficient solution procedure. See also Federgruen and Tzur (1991, 1994, 1995), Heady and Zhu (1994), and Wagelmans et al. (1992). Pochet and Wolsey (2006) discuss the techniques for improving the efficiency for numerous production-planning models.

associated replenishments, we again use the example of Table 5.1. In the following table, we replicate the requirements given in Table 5.1.

Month	Sequential Number	Requirements (Boxes)
January	1	10
February	2	62
March	3	12
April	4	130
May	5	154
June	6	129

Month	Sequential Number	Requirements (Boxes)
July	7	88
August	8	52
September	9	124
October	10	160
November	11	238
December	12	41
		Total = 1,200

$F(1)$ is the total costs of a replenishment of size 10 at the start of January, which is simply the setup cost A or \$54. To determine $F(2)$, we have two possible options to consider:

Option 1 Replenish 10 boxes at the start of January and 62 boxes at the start of February.

Option 2 Replenish enough (72 boxes) at the start of January to cover the requirements of both January and February.

Costs of Option 1 = (Costs of best plan from the start to the end of January) + (Costs of a replenishment at the start of February to meet February's requirements)

$$\begin{aligned}
 &= F(1) + A \\
 &= \$54 + \$54 \\
 &= \$108
 \end{aligned}$$

Costs of Option 2 = (Setup cost for January replenishment) + (Carrying costs for February's requirements)

$$= \$54 + 62 \text{ boxes} \times \$0.40/\text{box}/\text{month} \times 1 \text{ month}$$

$$= \$54 + \$24.80$$

$$= \$78.80$$

The cost of the second option is less than that of the first. Therefore, the best choice to satisfy the requirements of January and February is Option 2 and $F(2) = \$78.80$.

To satisfy requirements through to the end of March, there are three options, namely, where we position the last replenishment:

Option 1 Cover to the end of February in the best possible fashion and replenish 12 boxes at the start of March.

Option 2 Cover to the end of January in the best possible fashion and replenish 74 boxes at the start of February.

Option 3 Have a single replenishment of 84 boxes at the start of January.

Costs of Option 1 $= F(2) + A$

$$= \$78.80 + \$54$$

$$= \$132.80$$

Costs of Option 2 $= F(1) + A + (\text{Carrying costs for March's requirements})$

$$= \$54 + \$54 + (12 \times 0.40 \times 1)$$

$$= \$112.80$$

Costs of Option 3 $= A + (\text{Carrying costs for February's requirements}) + (\text{Carrying costs for March's requirements})$

$$= \$54 + (62 \times 0.40 \times 1) + (12 \times 0.40 \times 2)$$

$$= \$88.40$$

Therefore, Option 3, a single replenishment at the start of January, is best in terms of meeting requirements through to the end of March.

We continue forward in this fashion until we complete period N (here $N = 12$). For any specific month t , there are t possible options to evaluate. It is important to note that the method, as just described, requires an ending point where it is known that the inventory level is to be at zero or some other specified value.

To illustrate the effect of property 2 discussed at the beginning of this section, consider the options open to us in terms of meeting requirements through to the end of July (month 7). The discussion in the previous paragraph would indicate that we must evaluate seven options. However, consider the requirements of 88 boxes in the month of July. The carrying costs associated with these requirements, if they were included in a replenishment at the beginning of June, would be

$$88 \text{ boxes} \times \$0.40/\text{box}/\text{month} \times 1 \text{ month or } \$35.20$$

Similarly, if these requirements were instead included in a replenishment at the beginning of May, the carrying costs associated with the 88 units would be

$$88 \text{ boxes} \times \$0.40/\text{box}/\text{month} \times 2 \text{ months or } \$70.40$$

which is larger than \$54, the fixed cost of a replenishment. Therefore, the best solution could never have the requirements of July included in a replenishment at the beginning of May or earlier. It would be less expensive to have a replenishment arrive at the beginning of July. Thus, we need to consider only two options—namely, having the last replenishment at the start of either June or July.

If we were interested in computing the size of only the first replenishment quantity, then, it may not be necessary to go all the way out to the horizon (month N). The use of property 2 shows that, if for a period j , the requirements $D(j)$ are so large that the cost of carrying the inventory needed to meet $D(j)$ (for one period) exceeds the cost of a replenishment,

$$D(j)vr > A$$

or equivalently,

$$D(j) > A/vr$$

then the optimal solution will have a replenishment at the beginning of period j ; that is, the inventory must go to zero at the start of period j . Therefore, the earliest j where this happens can be used as a horizon for the calculation of the first replenishment. In the numerical example,

$$A/vr = \frac{\$54}{\$20/\text{boxes} \times 0.02\$/\$/\text{month}} = 135 \text{ box} - \text{months}$$

We can see from the demand pattern that the earliest month with $D(j) > 135$ is May. Therefore, the end of April can be considered as a horizon in computing the replenishment quantity needed at the beginning of January. It should be emphasized that it is possible for a given problem instance that all per-period requirements are below this threshold.

Table 5.4 illustrates a simple spreadsheet used to calculate the optimal solution. Notice that, because of the ease of computation on the spreadsheet, we have included all possible production levels for each month, even though we could have eliminated certain values based on the insights from the previous paragraph. Notice also how the minimum value in the column for month 12 is highlighted. The underlined values represent the number of months of production. For instance, the underline from month 12 to month 11 (or \$501.20 to \$484.80) indicates that production in month 11 will meet the requirements in months 11 and 12. The next step is to look for the minimum value in month 10 (which again is highlighted), and underline values to the left, if possible. Continuing in this manner, we develop the results shown in Table 5.5.

There are seven replenishments and the total costs amount to \$501.20. Comparison with Tables 5.2 and 5.3 shows that the Wagner–Whitin algorithm produces the total costs that are some 24.4% lower than those of the company’s three-month rule and some 22.1% lower than those of a fixed EOQ strategy. Furthermore, the turnover ratio has been increased to $1,200 \times 12/308$ or 46.8.

5.5.2 Potential Drawbacks of the Algorithm

As mentioned earlier, the Wagner–Whitin algorithm is guaranteed to provide a set of replenishment quantities that minimize the sum of replenishment plus carrying costs out to a specified horizon. (In fact, a generalized form permits including backordering as well as replenishment costs that depend on the period in which the replenishment occurs.) As we have seen, it is possible to develop

Table 5.5 Results of Using the Wagner–Whitin Algorithm or the Silver–Meal Heuristic on the Numerical Example

Month	1	2	3	4	5	6	7	8	9	10	11	12	Total
Starting inventory	0	74	12	0	0	129	0	52	0	0	0	41	
Replenishment	84	0	0	130	283	0	140	0	124	160	279	0	1,200
Requirements	10	62	12	130	154	129	88	52	124	160	238	41	1,200
Ending inventory	74	12	0	0	129	0	52	0	0	0	41	0	308
Total replenishment costs = $7 \times \$54 = \378.00													
Total carrying costs = $308 \text{ box-months} \times \$20/\text{box} \times 0.02 \text{ \$/month} = \$123.20$													
Total replenishment plus carrying costs = $\$501.20$													

spreadsheet solutions to the dynamic programming model; and with linear programming add-ins to spreadsheets, it is possible to use that formulation and find an optimal solution. Modified versions of the algorithm have been successfully implemented. See, for example, Potamianos and Orman (1996). In general, however, the algorithm has received extremely limited acceptance in practice. The primary reasons for this lack of acceptance are:

1. The relatively complex nature of the algorithm that makes it more difficult for the practitioner to understand than other approaches.
2. The possible need for a well-defined ending point for the demand pattern. (This would be artificial for the typical inventoried item where termination of demand is not expected in the near future.) As shown above, such an ending point is not needed when there is at least one period whose requirements exceed A/vr . Moreover, all information out to the end point may be needed for computing even the initial replenishment quantity. In this connection, considerable research effort (see e.g., Eppen et al. 1969; Kunreuther and Morton 1973; Blackburn and Kunreuther 1974; Lundin and Morton 1975) has been devoted to ascertaining the minimum required length of the planning horizon to ensure the selection of the optimal value of the initial replenishment.
3. A related issue is the fact that the algorithm is often used in conjunction with MRP software. Because MRP typically operates on a rolling schedule, the replenishment quantities chosen should not change when new information about future demands becomes available. Unfortunately, the Wagner–Whitin approach does not necessarily have this property (which Baker (1989) defines as the *insulation* property). This is potentially a critical flaw in this approach.*
4. The necessary assumption that replenishments can be made only at discrete intervals (namely, at the beginning of each of the periods). This assumption can be relaxed by subdividing the periods; however, the computational requirements of the algorithm go up rapidly with the number of periods considered. (In contrast, the other two methods are easily modified to account for continuous opportunities to replenish.)

* Sahin et al. (2013) review the literature on production planning with rolling horizons.

When taking into account of the total relevant costs, particularly including system control costs, it is not at all surprising that the algorithm has received such limited acceptability. We do not advocate its use for B items. (Let us not lose sight of the fact that, in this chapter, we are still talking about this class of items as opposed to Class A items. For the latter, considerably more control expense is justified because of the higher potential savings per item.) Instead, it is appropriate to resort to simpler heuristic methods that result in reduced control costs that more than offset any extra replenishment or carrying costs that their use may incur.

5.6 Heuristic Approaches for a Significantly Variable Demand Pattern

As indicated by the results in the previous two sections, the Wagner–Whitin approach does substantially better than a fixed EOQ, at least for the illustrative numerical example. However, as mentioned in Section 5.5.2, the Wagner–Whitin approach has drawbacks from the practitioner’s standpoint. Therefore, the natural question to ask is: “Is there a simpler approach that will capture most of the potential savings in the total of replenishment and carrying costs?” A number of researchers (see, e.g., Diegel 1966; DeMatteis 1968; Gorham 1968; Mendoza 1968; Silver and Meal 1973; Donaldson 1977; Groff 1979; Karni 1981; Wemmerlöv 1981; Brosseau 1982; Free-land and Colley 1982; Boe and Yilmaz 1983; Mitra et al. 1983; Bookbinder and Tan 1984; Baker 1989; Coleman and McKnew 1990; Triantaphyllou 1992; Teng 1994; Hariga 1995a; Goyal et al. 1996; Zhou 1996; Schulz 2011) have suggested various decision rules, some of which have been widely used in practice.* In this section, we review several heuristics that, as we will find, accomplish exactly what we desire. Moreover, in numerous test examples, some of these heuristics have performed extremely well. Because they are so frequently seen in practice, we also discuss some commonly used heuristics that perform less well.

As mentioned earlier, the fixed EOQ approach should perform suitably when the demand pattern varies very little with time. For this reason, we advocate the use of the somewhat more complicated heuristics only when the pattern is significantly variable. An operational definition of “significantly variable” will be given in Section 5.6.7.

We begin with the Silver–Meal (or least period cost) heuristic (see Silver and Meal 1973), using it to illustrate the basic concepts. Then we briefly describe other heuristics.

5.6.1 *The Silver–Meal, or Least Period Cost, Heuristic*

The Silver–Meal heuristic selects the replenishment quantity in order to replicate a property that the basic EOQ possesses when the demand rate is constant with time, namely, the *total relevant costs per unit time for the duration of the replenishment quantity are minimized*. If a replenishment arrives at the beginning of the first period and it covers requirements through to the end of the T th period, then, the criterion function can be written as follows:

$$\frac{(\text{Setup cost}) + (\text{Total carrying costs to end of period } T)}{T}$$

* Wee (1995) develops an exact solution to the problem of finding a replenishment policy for an item that faces declining demand, and that deteriorates over time. See Hariga (1994), Ting and Chung (1994), Hariga (1995b), Bose et al. (1995), and Chapter 8 as well. See Hill (1995) for the case of increasing demand, and Hariga (1996) for the case of inflationary conditions.

This is a reasonable criterion and it has the desirable feature of not including, in the present replenishment, a large requirement well in the future (inclusion of such a requirement would make the “costs per unit time” measure too high). It is not difficult to develop numerical examples in which the use of the criterion does not lead to the overall optimal solution, particularly for the case where the demand pattern has a well-defined ending point. Fortunately, this is not a major drawback because our primary concern is with demand patterns that do not have a clearly defined ending point in the near future. (This follows from our basic definition of Class B items.)

Because we are constrained to replenishing at the beginnings of periods, the best strategy must involve replenishment quantities that last for an integer number of periods. Consequently, we can think of the decision variable for a particular replenishment as being the time T that the replenishment will last, with T constrained to integer values. The replenishment quantity Q , associated with a particular value of T , is

$$Q = \sum_{j=1}^T D(j) \tag{5.1}$$

provided we set the time origin so that the replenishment is needed and arrives at the beginning of period 1. According to the chosen criterion, we wish to pick the T value that minimizes the total relevant costs per unit time of replenishment and carrying inventory over the time period T .

Let the total relevant costs associated with a replenishment that lasts for T periods be denoted by $\text{TRC}(T)$. These costs are composed of the fixed replenishment cost A and the inventory carrying costs. We wish to select T to minimize the total relevant costs per unit time, $\text{TRCUT}(T)$, where

$$\text{TRCUT}(T) = \frac{\text{TRC}(T)}{T} = \frac{A + \text{carrying costs}}{T} \tag{5.2}$$

If $T = 1$, there are no carrying costs (we only replenish enough to cover the requirements of period 1), that is,

$$\text{TRCUT}(T) = \frac{A}{1} = A$$

If the setup cost A is large, this may be unattractive when compared with including the second period’s requirements in the replenishment, that is, using $T = 2$.

With $T = 2$, the carrying costs are $D(2)vr$, which are the cost of carrying the requirements $D(2)$ for one period. Therefore,

$$\text{TRCUT}(2) = \frac{A + D(2)vr}{2}$$

Now, the setup cost is apportioned across two periods, but a carrying cost is incurred. With $T = 3$ we still carry $D(2)$ for one period, but now we also carry $D(3)$ for two periods. Thus,

$$\text{TRCUT}(3) = \frac{A + D(2)vr + 2D(3)vr}{3}$$

In this case, the setup charge is apportioned across three periods, but this may not be attractive because of the added carrying costs.

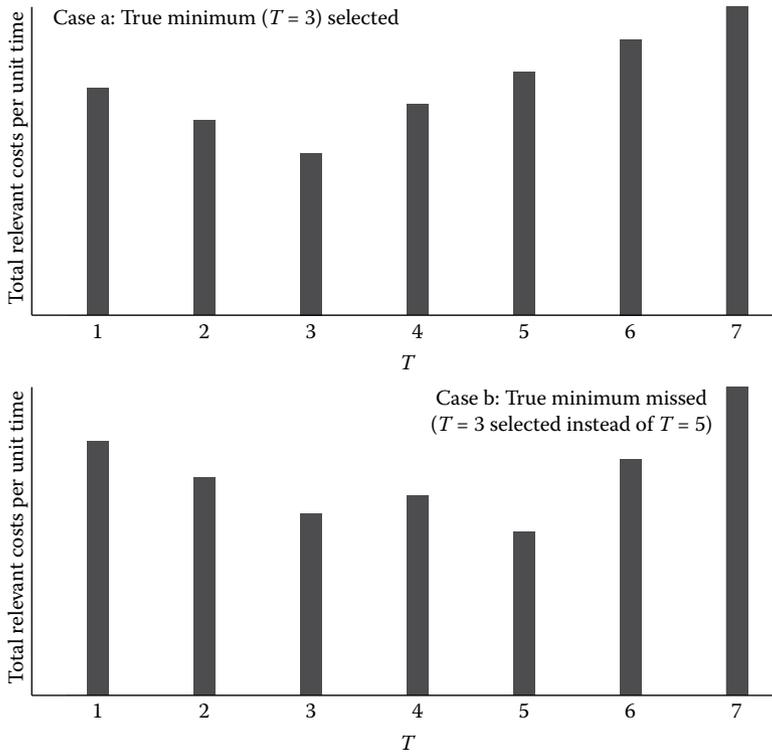


Figure 5.2 Graph of the selection of T in the Silver–Meal heuristic.

The basic idea of the heuristic is to evaluate $\text{TRCUT}(T)$ for increasing values* of T until, for the first time,

$$\text{TRCUT}(T + 1) > \text{TRCUT}(T)$$

that is, the total relevant costs per unit time start increasing. When this happens, the associated T is selected as the number of periods that the replenishment should cover. The corresponding replenishment quantity Q is given by Equation 5.1.

As evidenced by Figure 5.2, this method guarantees only a local minimum in the total relevant costs per unit time, *for the current replenishment*. It is possible that still larger values of T would yield still lower costs per unit time since we stop testing with the first increase in costs per unit time. We could protect against this eventuality by computing the ratio for a few more values of T , but the likelihood of improvement in most real cases is small. Furthermore, let us not lose sight of the fact that “costs per unit time for the first replenishment” is only a surrogate criterion.

* When some of the D 's are zeros, we proceed as follows: Suppose $D(j - 1) > 0, D(j) > 0, D(j + 1) = 0$, and $D(j + 2) > 0$. We evaluate $\text{TRCUT}(j - 1)$, and then jump to $\text{TRCUT}(j + 1)$, because a replenishment that covers period j will also, trivially, cover the zero requirement in period $j + 1$.

It is conceivable that the $TRCUT(T)$ may continue to decrease all the way out to $T = N$. This says that it is appropriate to cover all the requirements out to the horizon with the current replenishment. In such a case, it may be appropriate to cover an even longer period, but the heuristic is unable to provide the answer without forecast information beyond period N .

Except for the unusual situation just described, the performance of the heuristic, in contrast with the Wagner–Whitin algorithm, does not crucially depend on the N value chosen. It tends to use demand data of only the first few periods, an attractive property when one recognizes that the deterministic demand assumption becomes less reasonable as one projects further into the future.

To illustrate the application of the heuristic, let us again use the numerical example of Section 5.3. As previously introduced, the values of A , v , and r and the first part of the requirements pattern are as follows:

$$A = \$54 \quad v = \$20/\text{box} \quad r = 0.02 \text{ \$/\$/month}$$

Month j	1	2	3	4	5	6...
Requirements $D(j)$	10	62	12	130	154	129...

The calculations for the first replenishment quantity (assuming that the inventory is zero at the beginning of month 1) are shown in Table 5.6. The heuristic selects a T value of 3 with an associated Q , using Equation 5.1, of

$$\begin{aligned} Q &= D(1) + D(2) + D(3) \\ &= 10 + 62 + 12 \\ &= 84 \text{ boxes} \end{aligned}$$

The table illustrates that the computations are simply multiplications, additions, divisions, and comparisons of numbers.

It turns out for this numerical example (and for a substantial portion of all others tested) that this simple heuristic gives the same solution as the Wagner–Whitin algorithm. Thus, the solution has already been shown in Table 5.5.

Table 5.6 Computations for the First Replenishment Quantity Using the Silver–Meal Heuristic

T	Requirements	Months in Inventory	Carrying Cost	Order Cost	Total Cost	Total Cost $\div T$
1	10	0	\$0.00	\$54.00	\$54.00	\$54.00/month
2	62	1	\$24.80 ^a	\$54.00	\$78.80	\$39.40/month
3	12	2	\$34.40 ^b	\$54.00	\$88.40	\$29.47/month
4	130	3	\$190.40	\$54.00	\$244.40	\$61.10/month

^a $\$24.80 = D(2)vr = 62(20)(0.02)$.

^b $\$34.40 = D(2)vr + 2D(3)vr = 62(20)(0.02) + 2(12)(20)(0.02)$.

5.6.2 The EOQ Expressed as a Time Supply (POQ)

One approach described earlier (in Section 5.4) was to use a fixed order quantity, based on using the average demand rate in the EOQ equation. Empirically (see, e.g., Brown 1977), where there is significant variability in the demand pattern, better cost performance has been obtained by proceeding slightly differently. The EOQ is expressed as a time supply using, namely,

$$T_{\text{EOQ}} = \frac{\text{EOQ}}{\bar{D}} = \sqrt{\frac{2A}{\bar{D}vr}} \quad (5.3)$$

rounded to the nearest integer greater than zero. Then, any replenishment of the item is made large enough to cover exactly the requirements of this integer number of periods. Another name for this approach is the periodic order quantity (POQ). This approach has the advantage that arrivals of orders to the factory floor are regular even though the order quantities will be different from order to order. Note that this is in contrast to the fixed order quantity approach that has irregular timing, but constant quantities.

For the numerical example of Section 5.3.2, Equation 5.3 gives

$$T_{\text{EOQ}} = \sqrt{\frac{2(54)}{(100)(20)(0.02)}} = 1.64$$

which rounds to 2. The repeated use of a 2-period time supply in the example gives the total costs of \$553.60 as compared to \$501.20 using the Silver–Meal heuristic.

5.6.3 Lot-for-Lot

Perhaps, the easiest approach is Lot-for-Lot (L4L), which simply orders the exact amount needed for each time period. Thus, inventory holding costs are zero with L4L. Clearly, this approach is not cost effective when fixed replenishment costs are significant. For our example, the total costs are $12 \times \$54$ or \$648. We mention L4L here only because it will reappear in the context of MRP where it is commonly used. (See Chapter 15.)

5.6.4 Least Unit Cost

The Least Unit Cost (LUC) heuristic is similar to the Silver–Meal heuristic except that it accumulates requirements until the cost per unit (rather than the cost per period) increases. Calculations for the first order are shown in Table 5.7.

Table 5.8 contains the results for the entire 12-month time. The total cost of \$558.80 is 11.5% higher than the optimal solution. Research has shown that this is not unusual performance for the LUC heuristic.

5.6.5 Part-Period Balancing

The basic criterion used here is to select the number of periods covered by the replenishment such that the total carrying costs are made as close as possible to the setup cost, A . (Exact equality is usually not possible because of the discrete nature of the decision variable, T .) To illustrate, for the

Table 5.7 Computations for the First Replenishment Quantity Using the LUC Heuristic

<i>T</i>	Requirements	Months in Inventory	Carrying Cost	Order Cost	Total Cost	Lot Size	TC per Unit
1	10	0	\$0.00	\$54.00	\$54.00	10	\$5.40
2	62	1	\$24.80	\$54.00	\$78.80	72	\$1.09
3	12	2	\$34.40	\$54.00	\$88.40	84	\$1.05
4	130	3	\$190.40	\$54.00	\$244.40	214	\$1.14
5	154	4	\$436.80	\$54.00	\$490.80	368	\$1.33

numerical example, we have the calculations for the first replenishment quantity (assuming that the inventory is zero at the beginning of month 1):

<i>T</i>	Carrying Costs
1	0
2	$D(2)vr = \$24.80 < \54
3	$24.80 + 2D(3)vr = \$34.40 < \54
4	$\$34.40 + 3D(4)vr = \$194.40 > \$54$

\$34.40 is closer to \$54 (the *A* value) than is \$190.40. Therefore, a *T* value of 3 is selected for the first replenishment. Repeated use of this procedure gives replenishments listed in Table 5.9, with the total costs of \$600—again well above the \$501.20 found by the Silver–Meal heuristic (Table 5.5).

Table 5.8 Results of Using LUC on the Numerical Example

Month	1	2	3	4	5	6	7	8	9	10	11	12	Total
Starting inventory	0	74	12	0	154	0	88	0	124	0	0	0	
Replenishment	84	0	0	284	0	217	0	176	0	160	238	41	1,200
Requirements	10	62	12	130	154	129	88	52	124	160	238	41	1,200
Ending inventory	74	12	0	154	0	88	0	124	0	0	0	0	452
Total replenishment costs = $7 \times \$54 = \378.00													
Total carrying costs = $452 \text{ box-months} \times \$20/\text{box} \times 0.02 \text{ \$/month} = \$180.80$													
Total replenishment plus carrying costs = \$558.80													

Table 5.9 Results of Using PPB on the Numerical Example

Month	1	2	3	4	5	6	7	8	9	10	11	12	Total
Starting inventory	0	74	12	0	154	0	88	0	124	0	238	0	
Replenishment	84	0	0	284	0	217	0	176	0	398	0	41	1,200
Requirements	10	62	12	130	154	129	88	52	124	160	238	41	1,200
Ending inventory	74	12	0	154	0	88	0	124	0	238	0	0	690
Total replenishment costs = $6 \times \$54 =$													\$378.00
Total carrying costs = $690 \text{ box-months} \times \$20/\text{box} \times 0.02 \text{ \$/month} =$													\$276.00
Total replenishment plus carrying costs =													\$600.00

Refinements of the part-period balancing (PPB) method, requiring more computational effort, have been developed (see e.g., DeMatteis 1968; Mendoza 1968). One refinement is called the Look-Ahead/Look-Back technique. The Look-Ahead technique evaluates the cost of moving an order later in time. For instance, because the first order of 84 units cannot be moved to period 2 without incurring a stockout, we first examine the order for 284 units in period 4. If that order were moved to period 5, the 130 units required in period 4 must be produced in period 1 and held for 3 periods—at a cost of 130 units (or parts) \times 3 periods = 390 part periods. The 154 units held from period 4 to period 5 are saved—or 154 parts \times 1 period = 154 part periods. Therefore, no adjustment is made to that order. We come to a similar conclusion with the order for 217 units in period 6: save 88 units for 1 period (or 88 part periods) at a cost of 129 units held for 2 periods (or 258 part periods).

It is attractive to make an adjustment to the 176 units ordered in period 8. Moving that order to period 9 requires shifting the 52 units needed in period 8 to the order in period 6—at a cost of $52 \times 2 = 104$ part periods. The cost of holding 124 units from period 8 to period 9 is saved— $124 \times 1 = 124$ part periods. So, the adjustment would be made; the order in period 6 would now be $217 + 52 = 269$ units, and the next order would be 124 units in period 9.

Unfortunately, Wemmerlöv (1983) has shown that the Look-Ahead/Look-Back techniques do not necessarily improve the performance of the PPB heuristic. He also notes that the PPB heuristic generally performs poorly in dynamic environments.

5.6.6 Performance of the Heuristics

Many of these heuristics have been tested against the Wagner–Whitin algorithm, the basic EOQ, and other heuristics on a wide range of examples. Baker (1989) summarizes this literature and reconciles the differences among different papers. Van den Heuvel and Wagelmans (2010) analyze the worst-case behavior of several lot-sizing heuristics. One lesson from all the literature is that most heuristics outperform the fixed EOQ approach. Another lesson, illustrated in Baker's paper, is that the Silver–Meal heuristic incurs an average cost penalty for using the heuristic instead of the “optimal” Wagner–Whitin algorithm of less than 1%; in many cases there is no penalty whatsoever. (See Table 5.10.) Finally, tests in a rolling-horizon environment (see, e.g., Blackburn and Millen 1980) have revealed that frequently the Silver–Meal heuristic actually outperforms the dynamic programming algorithm (because the latter now gives the optimal solution to the wrong problem).

Table 5.10 Performance of Selected Heuristics on 20 Test Problems

Heuristic	Number Optimal	Average Percent Error (%)
POQ	7	10.004
LUC	8	9.303
PPB	16	1.339
Silver–Meal	14	0.943

Source: Adapted from Baker, K. R. 1989. *Journal of Manufacturing and Operations Management* 2(3), 199–221.

5.6.7 When to Use Heuristics

These heuristics, although quite simple, are still more involved than the determination of the basic EOQ. We know that the latter is the best replenishment quantity to use when there is no variability in the demand rate. In fact, the variability of the demand pattern should exceed some threshold value before it makes sense to change to a heuristic.

A useful measure of the variability of a demand pattern is the squared coefficient of variation. This statistic, denoted by SCV, is given by

$$SCV = \frac{\text{Variance of demand per period}}{\text{Square of average demand per period}}$$

which can be computed easily on a spreadsheet. Tests have shown that a threshold value of SCV appears to be in the neighborhood of 0.2, that is, if $SCV < 0.2$, use a simple EOQ involving \bar{D} as the demand estimate. If $SCV \geq 0.2$, use a heuristic.

For the numerical example discussed in this chapter, we get $SCV = 0.426$, which is greater than the threshold value indicating that a heuristic, rather than the EOQ, should be used. (Recall from Tables 5.3 and 5.5 that the Silver–Meal heuristic produced total relevant costs of 22.1% below those resulting from the use of the EOQ.)

In addition, consider the absolute importance of the item, in terms of the potential savings in replenishment and carrying costs. A useful surrogate for this quantity is the total of these costs per unit time under the use of the EOQ in the EOQ cost model (we know that the EOQ model is, strictly speaking, only an approximation for the case of time-varying demand, but it is simple to use), namely,

$$\sqrt{2A\bar{D}vr}$$

If this quantity was very small, departing from the simple EOQ decision rule would not be justified. However, in this chapter, we are concerned with B items, defined to be those with intermediate values of $\bar{D}v$. Thus, by definition, for such items, the above test quantity will be large enough to ensure that we should not blindly remain with the basic EOQ when the demand rate varies enough.

A word of caution is in order. There are two situations where the use of the heuristic can lead to significant cost penalties:

1. When the demand pattern drops rapidly with time over several periods.
2. When there are a large number of periods having no demand.

Silver and Miltenburg (1984) have suggested modifications of the heuristic to cope with these circumstances. Baker (1989) also notes other heuristics that perform slightly better at little computational cost. When unit production costs vary over time, the heuristic can also perform quite poorly. ? develop an approximation that performs well in this case.

5.6.8 Sensitivity to Errors in Parameters

In Section 4.3 of Chapter 4, it was illustrated that, for the case of the basic EOQ, the total cost curve is quite shallow in the neighborhood of the best order quantity. Thus, even substantial deviations of the order quantity away from its best value tend to produce small percentage cost penalties. One case of such deviations is errors in the parameter (cost or demand) values. We concluded that the costs were relatively insensitive to errors in the input parameters for the case of the EOQ formulation.

Fortunately, tests have revealed that a similar phenomenon exists for the Silver–Meal heuristic. The results of one such test are shown in Figure 5.3. The basic data used (taken from Kaimann 1969) were

$$A/vr = 150$$

Period	1	2	3	4	5	6	7	8	9	10	11	12
Requirements	10	10	15	20	70	180	250	270	230	40	0	10

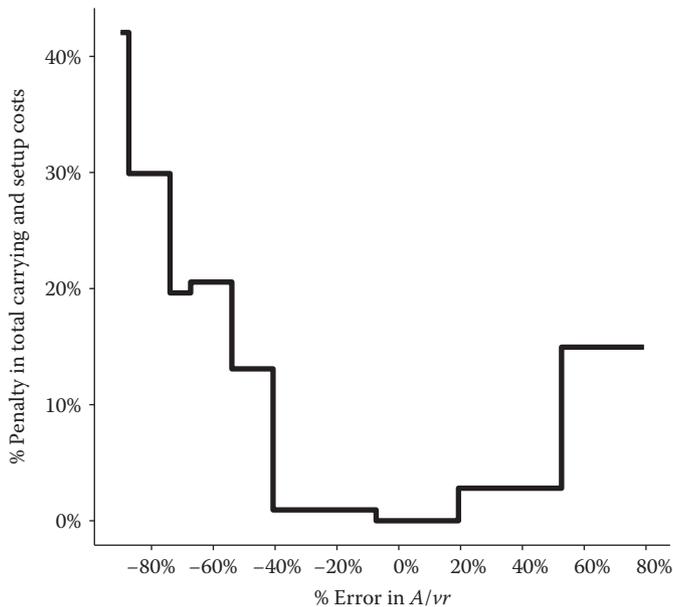


Figure 5.3 Illustration of the insensitivity of heuristic results to errors in cost parameters.

A/vr was deliberately changed from its correct value. For this incorrect value of A/vr , the Silver–Meal heuristic was used to compute the replenishment quantities. The true value of A/vr was then employed to compute the costs associated with this sequence of replenishments. These costs were compared with those of the replenishment pattern resulting from using the correct value of A/vr . This was repeated for a number of different percentage errors in A/vr . To illustrate, in Figure 5.3, we can see that if A/vr is erroneously set at a value of 40% that is too high (i.e., at a value of 210 instead of the correct level of 150), then, the cost penalty is less than 2%. The plot is not a smooth curve (as was the case in Figure 4.3 of Chapter 4) because here the replenishment opportunities are discrete in nature. Nonetheless, in all cases tested, the percentage cost penalties, even for reasonable-sized errors in A/vr , were quite small. Pan (1994) reaches the same conclusion.

5.6.9 Reducing System Nervousness

In a rolling-horizon environment, as new demand information becomes available, the timing and sizes of future replenishments may be altered. Frequent alterations of this type may be unattractive from an operational standpoint. Carlson et al. (1979) and Kropp et al. (1983) have introduced the idea of a penalty cost for changing a previously established schedule; they have shown how to incorporate this idea into the Silver–Meal heuristic and the Wagner–Whitin algorithm. As a by-product, this modification permits the handling of the situation where the setup cost varies from period to period. See also Blackburn et al. (1986), Sridharan and Berry (1990), Sridharan et al. (1987), Kropp and Carlson (1984), DeMatta and Guignard (1995), and Toklu and Wilson (1995). We will have more to say on this topic in Chapter 15.

5.7 Handling of Quantity Discounts

An important extension of the heuristic is to permit the inclusion of quantity discounts. As in Chapter 4, we illustrate for the case of an all-units fractional discount d if $Q \geq Q_b$. The presence of a discount removes the validity of one of the key properties of a solution first mentioned in Section 5.5—namely, that replenishments are received only when the inventory level is 0. Sometimes, it may indeed be appropriate to order exactly Q_b , which need not cover precisely an integer number of periods of requirements.

A revised heuristic, developed by Bregman and Silver (1993), involves a three-step procedure. First, determine the value of T , the replenishment quantity expressed as a time supply, which gives the lowest cost per period ignoring the discount possibility. Denote this lowest cost time supply, T' . If $T' \geq T_b$, where T_b is the time supply covered by a replenishment of size Q_b , then T' is used. The second step is needed if $T' < T_b$ and it involves checking if the use of T_b (i.e., taking advantage of the discount) is attractive. If so, the third step is to check if it is attractive to cover the integer T value above T_b . More formally,

- Step 1* Find the first minimum ignoring the discount possibility.
 Find the lowest integer T that gives a local minimum of

$$\text{TRCUT}(T) = \begin{cases} \frac{A + vr \sum_{j=1}^T (j-1)D(j)}{T} & T < T_b \\ \frac{A + v(1-d)r \sum_{j=1}^T (j-1)D(j)}{T} & T \geq T_b \end{cases} \quad (5.4)$$

If $\text{TRCUT}(T + 1)$ is less than or equal to $\text{TRCUT}(T)$, continue trying $T + 2$, and so on. If there is a $D(T + 1) = 0$, then, immediately compare $\text{TRCUT}(T + 1)$ to $\text{TRCUT}(T - 1)$ because $\text{TRCUT}(T + 1)$ will always be less than $\text{TRCUT}(T)$ in this situation. Denote the minimizing T value as T' . If $T' \geq T_b$, then use $T = T'$ with

$$Q = \sum_{j=1}^{T'} D(j) \tag{5.5}$$

If $T' < T_b$, then go to Step 2.

Step 2 Check if the discount is attractive when $T' < T_b$.

Evaluate the modified total relevant costs per unit time (modified to take into account of the differences in the unit costs)

$$\text{MTRCUT}(T') = \frac{A + vr \sum_{j=1}^{T'} (j - 1)D(j) + vd \sum_{j=1}^{T'} D(j)}{T'} \tag{5.6}$$

$$\text{MTRCUT}(T_b) = \frac{A}{[T_b]} + \frac{v(1 - d)r \sum_{j=1}^{[T_b]} (j - 1)D(j) + v(1 - d)r[T_b] \left(Q_b - \sum_{j=1}^{[T_b]} D(j) \right)}{T_b} \tag{5.7}$$

where $[T_b]$ is the integer portion of T_b (e.g., if $T_b = 3.2$, then $[T_b] = 3$). If $\text{MTRCUT}(T_b) > \text{MTRCUT}(T')$, then use T' . If $\text{MTRCUT}(T_b) \leq \text{MTRCUT}(T')$, then go to Step 3.

The $vd \sum D(j)$ term in Equation 5.6 represents the discount dollar savings passed up in using T' . The A term in Equation 5.7 is divided by $[T_b]$ because a new order will be required $[T_b]$ periods later even though some stock will still be on hand. The vr terms in the numerators of Equations 5.6 and 5.7 represent the relevant carrying costs.

Step 3 Check if it is attractive to cover $[T_b] + 1$ periods.

To determine if the added carrying costs for an additional period are less than the savings from avoiding a fixed replenishment cost for the additional period, compute $\text{TRCUT}([T_b] + 1)$ using Equation 5.4 for $T > T_b$. If $\text{MTRCUT}(T_b) < \text{TRCUT}([T_b] + 1)$, use T_b with $Q = Q_b$. If $\text{MTRCUT}(T_b) \geq \text{TRCUT}([T_b] + 1)$, use

$$Q = \sum_{j=1}^{[T_b]+1} D(j) \tag{5.8}$$

Bregman and Silver (1993) have shown that this heuristic performs extremely well, and that its computation time is quite small compared to other heuristics.

Other work on this topic includes Callarman and Whybark (1981), Benton (1983, 1985, 1991), La Forge (1985), Chung et al. (1987), Christoph and La Forge (1989), Federgruen and Lee (1990), Bregman (1991), and Hariga (1995c). Hu et al. (2004) investigate heuristics for the incremental quantity discount setting. They propose a modified version of the Silver–Meal heuristic that works quite well, particularly when the horizon is not too short, on a large set of test problems.

5.8 Aggregate Exchange Curves

As has been discussed earlier in the book, certain parameter values, such as r or A/r , may be selected implicitly by the top management; they specify a reasonable operating situation (such as an inventory budget) for all items in the inventory. In Chapter 4, for the case of the basic EOQ, it was relatively easy to develop an exchange curve because of the analytic relationships between (1) the total average cycle stock and A/r (see Equation 4.29), and (2) the total number of replenishments per unit time and A/r (see Equation 4.30).

For the case of time-varying demand, no such simple results exist because of the discrete nature of the decision variable T , the number of periods of requirements to be included in a replenishment. Therefore, in this case, in order to develop an exchange curve, we must proceed as follows.

A representative sample of items is selected from the population of items under consideration. The exact sample size to select depends on the number and characteristics of items in the population that are to be controlled by the replenishment batch size heuristic or algorithm (i.e., have a squared coefficient of variation exceeding 0.2).

A value of A/r is selected. Then, for each item in the sample, the heuristic, say Silver–Meal, is used to determine the replenishment quantities over a reasonable time period such as 12 months. This implies a certain number of replenishments and an average inventory level (in dollars) for the item. Summing these quantities across the items of the sample produces the total number of replenishments and the total average inventory of the sample. These figures must be appropriately scaled up to correspond with the total group of items to be controlled by the heuristic. This must be repeated for several values of A/r to develop the tradeoff curve.

An additional point is worth mentioning here. It is often desirable to have an exchange curve for all the items in the population. In such a case, the results for the Silver–Meal heuristic items must be added to those for the items to be replenished according to the basic EOQ (Section 4.10 of Chapter 4). This produces one composite curve as a function of A/r .

5.9 Summary

In this chapter, we have provided a decision system for coping with the relaxation of another of the key assumptions inherent in the use of the basic EOQ. The system permits the practitioner to deal, in a realistic way, with time-varying demand patterns. The Silver–Meal heuristic, as well as some other approaches, has been found to be robust. A guideline has been provided to aid the decision maker in choosing when to use the simple EOQ approach.

Time-varying demand patterns will play a crucial role in the Material Requirements Planning framework to be discussed in Chapter 15. In addition, an extension of the Silver–Meal heuristic to permit multiple items within a capacitated production context will be discussed in Chapter 10.

In the next chapter, we return to the context of an essentially level demand pattern but, for the first time, incorporate uncertainty in the demand rate. In Chapter 7, a brief mention will be made of the treatment of the combination of time-varying and uncertain demand.

Problems

- 5.1 A sale of polarizing filters is held twice annually by a photographic products distributor. The demand pattern for a particular size of filter for the past year is as follows:

<i>Jan.</i>	<i>Feb.</i>	<i>Mar.</i>	<i>Apr.</i>	<i>May</i>	<i>June</i>
21	29	24	86	31	38
<i>July</i>	<i>Aug.</i>	<i>Sept.</i>	<i>Oct.</i>	<i>Nov.</i>	<i>Dec.</i>
45	39	31	78	29	32

It is anticipated that demand for the next year will follow this pattern; hence, these figures are being used as the “best estimates” of forthcoming sales. Demand will also continue in future years. The cost of these filters is \$8.65, ordering costs are approximately \$35, and the carrying cost is 0.24 \$/\$/year.

Calculate the squared coefficient of variation and select the appropriate order quantities.

5.2 The demand pattern for another type of filter is

<i>Jan.</i>	<i>Feb.</i>	<i>Mar.</i>	<i>Apr.</i>	<i>May</i>	<i>June</i>
18	31	23	95	29	37
<i>July</i>	<i>Aug.</i>	<i>Sept.</i>	<i>Oct.</i>	<i>Nov.</i>	<i>Dec.</i>
50	39	30	88	22	36

These filters cost the company \$4.75 each; ordering and carrying costs are as in Problem 5.1. The squared coefficient of variation equals 0.33. Use the Silver–Meal heuristic to determine the sizes and timing of replenishments of stock.

5.3 a. For the following item having zero inventory at the beginning of period 1, develop the magnitude of the *first* replenishment *only* using the LUC method.

Item characteristics:

$$A = \$50 \quad v = \$2/\text{unit} \quad r = 0.05 \text{ } \$/\$/\text{period}$$

<i>Period j</i>	1	2	3	4	5	6	7
<i>D(j)</i>	200	300	500	500	400	400	300

b. A marked difference between this method and that of Silver–Meal exists in the dependence on $D(1)$. Briefly discuss. In particular, suppose $D(1)$ was much larger than any of the other $D(j)$'s. What effect would this have on the best T for

i. Silver–Meal?

ii. LUC?

5.4 Consider an item with the following properties:

$$A = \$20 \quad v = \$2/\text{unit} \quad r = 0.24 \text{ \$/\$/year}$$

At time 0, the inventory has dropped to zero and a replenishment (with negligible lead time) must be made. The demand pattern for the next 12 months is

Month j	1	2	3	4	5	6
Demand (units) $D(j)$	50	70	100	120	110	100
Month j	7	8	9	10	11	12
Demand (units) $D(j)$	100	80	120	70	60	40

All the requirements of each month must be available at the beginning of the month. Replenishments are restricted to the beginnings of the months. No shortages are allowed. Using each of the following methods, develop the pattern of replenishments to cover the 12 months and the associated total costs of each pattern (do not bother to count the costs of carrying $D(j)$ during its period of consumption, namely, period j). In each case, the size of the last replenishment should be selected to end month 12 with no inventory.

- Fixed EOQ (rounded to the nearest integer number of months of supply; i.e., each time the EOQ, based on the average demand through the entire 12 months, is adjusted so that it will last for exactly an integer number of months).
- A fixed time supply (an integer number of periods) based on the EOQ expressed as a time supply, using the average demand rate for the 12 months.
- On each replenishment, the selection of Q (or, equivalently, the integer T), which minimizes the costs per unit of quantity ordered to cover demand through T .
- The Silver–Meal heuristic.
- One replenishment at the start of month 1 to cover all the requirements to the end of month 12.
- A replenishment at the beginning of every month.

Hint: For each case, it would be helpful to develop a table with at least the following columns: (1) Month, (2) Replenishment Quantity, (3) Starting Inventory, (4) Demand, and (5) Ending Inventory.

5.5 Consider an item with the following deterministic, time-varying demand pattern:

Week	1	2	3	4	5	6
Demand	50	80	180	80	0	0
Week	7	8	9	10	11	12
Demand	180	150	10	100	180	130

Suppose that the pattern terminates at week 12. Let other relevant characteristics of the item be:

Inventory carrying cost per week (incurred only on units carried over from 1 week to the next) is \$0.20/unit.

Fixed cost per replenishment is \$30.

Initial inventory and replenishment lead time are both zero.

Perform an analysis similar to that of Problem 5.4 (of course, substituting “week” for “month”).

- 5.6 Consider an item with the following declining demand pattern:

<i>Period</i>	1	2	3	4	5	6	7	8	9	10
Demand	600	420	294	206	145	101	71	50	35	25
<i>Period</i>	11	12	13	14	15	16	17	18	19	20
Demand	17	12	9	6	5	3	2	2	1	1

There is no further demand after period 20.

$$A = \$50 \quad v = \$2.50/\text{unit} \quad r = 0.02 \text{ \$/\$/period}$$

Perform an analysis similar to that of Problem 5.4.

- 5.7 A Calgary-based company (Piedmont Pipelines) has been awarded a contract to lay a long stretch of pipeline. The planned schedule for the laying of pipe can be considered to be accurately known for the duration of the project (18 months). Because of weather conditions and the availability of labor, the rate of placement of the pipe is not constant throughout the project. The supplier of the pipe offers a discount structure for purchases. Outline an analysis that would assist Piedmont in deciding on the purchase schedule for the piping.
- 5.8 Consider a population of items having essentially deterministic but time-varying demand patterns. Suppose that one was interested in developing an exchange curve of total average cycle stock (in dollars) versus the total number of replenishments per year as a function of A/r (the latter being unknown, but constant for all the items). An analyst has suggested using the approach of Section 4.10 with

$$D_i = \text{annual demand for item } i$$

Discuss whether or not this is a reasonable suggestion.

- 5.9 Suppose you were a supplier for a single customer having a deterministic, but time-varying demand pattern. Interestingly enough, the time-varying nature is actually of benefit to you!
- Making use of the Silver–Meal heuristic, indicate how you would decide on the maximum discount that you would be willing to give the customer in return for him/her maintaining his/her current pattern instead of providing a perfectly level demand pattern. Assume that replenishments can take place only at the beginning of periods.
 - Illustrate for the numerical example of Problem 6.5. Suppose that the current selling price is \$60/unit.

5.10 Consider a company facing a demand pattern and costs as follows:

Month	Sequential Number	Requirements (Units)
January	1	20
February	2	40
March	3	110
April	4	120
May	5	60
June	6	30
July	7	20
August	8	30
September	9	80
October	10	120
November	11	130
December	12	40
	Total	800

A	\$25.00
r (per month)	0.05\$/\\$ (carrying costs are very high in this industry)
v	\$4.00

Using a “3-month” decision rule, the replenishment schedule and associated costs are as follows:

Month	1	2	3	4	5	6	7	8	9	10	11	12	Total
Starting inventory	0	150	110	0	90	30	0	110	80	0	170	40	
Replenishment	170	0	0	210	0	0	130	0	0	290	0	0	800
Requirements	20	40	110	120	60	30	20	30	80	120	130	40	800
Ending inventory	150	110	0	90	30	0	110	80	0	170	40	0	780

Total replenishment costs	\$100.00
Total carrying costs	\$156.00
Total replenishment + carrying	\$256.00

- a. Construct a replenishment schedule and calculate the associated costs using the fixed EOQ method.
 - b. Repeat using the Wagner–Whitin algorithm.
 - c. Repeat using the Silver–Meal heuristic.
 - d. Repeat using the LUC method.
 - e. Repeat using the PPB method.
 - f. Repeat using the POQ method.
- 5.11 Consider a company facing a demand pattern and costs as follows:

<i>Month</i>	<i>Sequential Number</i>	<i>Requirements (Units)</i>
January	1	200
February	2	300
March	3	180
April	4	70
May	5	50
June	6	20
July	7	10
August	8	30
September	9	90
October	10	120
November	11	150
December	12	190
	Total	1,410

<i>A</i>	\$160.00
<i>r</i> (per month)	0.10\$/
<i>v</i>	\$5.00

- a. Construct a replenishment schedule and calculate the associated costs using the fixed EOQ method.
 - b. Repeat using the Wagner–Whitin algorithm.
 - c. Repeat using the Silver–Meal heuristic.
 - d. Repeat using the LUC method.
 - e. Repeat using the PPB method.
 - f. Repeat using the POQ method.
- 5.12 What are we trying to achieve by using the Silver–Meal or any of the other heuristics in this chapter? Note that we are not asking for the criterion used in the Silver–Meal heuristic.
- 5.13 Consider an item with the following demand pattern:

Period j	1	2	3	4	5
$D(j)$ (units)	50	80	65	100	30

The initial inventory at the start of period 1 is 0 and a replenishment, to cover an integer number (T) of period, must be placed. Suppose that the use of the Silver–Meal heuristic leads to the choice of $T = 2$. What can you say about the value of the ratio A/vr where r is on a per-period basis? Hint: Use the normalized total relevant costs per unit time by dividing through by vr .

- 5.14 Develop a spreadsheet for the Wagner–Whitin algorithm as shown in Table 5.4.
- a. Simply perform the calculations on the spreadsheet, finding the optimal solution manually.
 - b. Now use the spreadsheet to find the optimal solution.
 - c. Repeat parts 5.14a and 5.14b for the Silver–Meal heuristic.
 - d. Repeat parts 5.14a and 5.14b for the LUC method.
 - e. Repeat parts 5.14a and 5.14b for the PPB method.
- 5.15 In this chapter, we consider *stocking* problems in which stock is ordered in anticipation of depletion by demand processes. A different kind of problem arises in transportation and distribution systems such as freight railroads, express and delivery services, freight forwarders, etc., in which the inventory accumulates according to supply processes and is depleted by dispatches. Consider such a *dispatching* problem, in which an inventory of dissimilar items accumulates according to a time-varying process. A periodic review is assumed, whereby at the end of each time period, there is an opportunity to dispatch, that is, to reduce the inventory to zero. A cost trade-off similar to the stocking problem is at issue, that is, there are inventory holding costs, and a significant, fixed dispatching cost.

We suppose the system starts (at the beginning of period 1) with an initially empty inventory (i.e., a dispatch cleared the system at the end of period 0). In the t th period, $t = 1, 2, \dots$, we let W_t denote the set of items supplied to the inventory whose first opportunity for dispatch is at the end of period t , and $HC(W_t)$ shall denote the cost of holding items W_t in the inventory for one period. (E.g., if the set of items W_t are kept in the inventory for 3 periods before dispatching, the associated holding costs are $3 HC(W_t)$). We let A denote the fixed cost per dispatching event. A dispatching policy is desired, that is, a determination in each time period of whether to dispatch or not dispatch in that period.

- a. Develop a decision procedure, such as the Silver–Meal heuristic, based on minimizing the relevant costs per unit time between dispatch moments.
- b. Illustrate your procedure (for just the first two dispatches) with the following example involving the dispatching of freight trains from a switchyard terminal:

At a terminal, freight cars with a particular destination W are switched to departure tracks where they accumulate until a train to destination W is dispatched. The holding costs per hour vary from car to car according to ownership, age, car type, etc., but we suppose that an information system at the terminal allows the calculation of holding costs for a given group of cars. Let W_t denote the set of cars destined to W switched in hour t , and let $HC(W_t)$ denote the hourly holding costs for W_t . We consider the case where a train was dispatched to W just before hour 1, and the switching results of the next several hours are as follows:

Hour, t	1	2	3	4	5	6	7
Holding cost, $HC(W_t)$	220	85	145	210	310	175	300

The fixed cost of dispatching a train to W is $A = 800$.

- 5.16 Consider an item where the assumptions of Section 5.3.1 are reasonable (including $L = 0$), but with one important exception, namely that there is a significant chance of the perishability of material kept too long in the inventory. Specifically, consider an order that arrives at time 0 (the beginning of period 1). The units of this order can become unusable at the end of each period that they are kept in stock. In particular, it is reasonable to assume that, for any period, there is a fixed chance p that any good inventory at the start of that period, which is still left over at the end of the period, will not be usable for satisfying the demand in future periods. Note that we are assuming that at the end of a period, one of the two things happens: (1) all of the remaining inventory is still usable, or (2) all of the remaining inventory has become worthless (with 0 salvage value). In the latter case, a new replenishment must be immediately placed (remember $L = 0$) to avoid a shortage in the next period. Assume that, if spoilage occurs right at the end of a period, then carrying costs are still incurred for the (now useless) inventory carried over into the next period but that this inventory is discarded before the end of that next period.
- a. What should tend to happen to the best value of T as p increases?
 - b. Develop a heuristic decision rule (in the spirit of the Silver–Meal heuristic) for choosing an appropriate order quantity at time 0. Use the notation consistent with the text. Make sure that your heuristic gives reasonable results for the limiting cases of $p = 0$ and $p = 1$.
 - c. Illustrate your rule with the following numerical example (but only choose the first-order quantity, the one at time 0). Also, determine the order quantity that would be found by the Silver–Meal heuristic (i.e., assuming $p = 0$) and comment on any difference from the value obtained using the heuristic with $p > 0$.

$$A = \$50, \quad v = \$10/\text{unit}, \quad r = 0.02\$/\$/\text{period}, \quad p = 0.01$$

j	1	2	3	4	5	6...
$D(j)$ units	200	100	80	50	30	100...

5.17 Consider a company facing a demand pattern and costs as follows:

Month	Sequential Number	Requirements (Units)
January	1	350
February	2	200
March	3	0
April	4	150
May	5	500
June	6	600
July	7	450
August	8	350
September	9	200
October	10	0
November	11	150
December	12	200
	Total	3,150

A	\$50.00
r (per month)	0.02\$/\$/month
v	\$65.00

- a. Construct a replenishment schedule and calculate the associated costs using the fixed EOQ method.
 - b. Repeat using the Wagner–Whitin algorithm.
 - c. Repeat using the Silver–Meal heuristic.
 - d. Repeat using the LUC method.
 - e. Repeat using the PPB method.
 - f. Repeat using the POQ method.
- 5.18 Suppose that there is a finite replenishment rate m (as discussed in Section 4.8).
- a. Modify the Silver–Meal heuristic to take account of this fact.

Note: For simplicity, assume that m is large enough so that, if a replenishment is initiated at the start of period j , then, the demand for period j , $D(j)$, can be met from

that replenishment. In addition, assume that it is now necessary to include in the analysis the carrying costs of $D(j)$ during its period of consumption, j . To do this, assume a uniform demand rate during any single period (without the extra cost included, the finite replenishment rate would have no effect on the decision).

- b. For the item of Problem 5.4 suppose that $m = 150$ units/month. Now, use the modified heuristic to find the timing and sizes of the replenishments.
- c. For this example, what is the percentage cost penalty over the 12-month horizon associated with ignoring the finite m value, that is, using the unmodified heuristic?

Appendix 5A: Dynamic Programming and Linear Programming Formulations

Define the following notation:

$D(j)$ = requirements in period j .

I_j = ending inventory in period j .

$F(j)$ = total costs of the best replenishment strategy that satisfies the demand requirements in periods $1, \dots, j$.

Q_j = replenishment quantity in period j .

The problem can be formulated as an integer program, and solved quite easily with a spreadsheet (e.g., the Solver in Excel).

$$\text{Minimize TRC} = \sum_{j=1}^T [A\delta(A_j) + I_jvr]$$

subject to

$$\begin{aligned} I_j &= I_{j-1} + Q_j - D(j) & j = 1, \dots, T \\ Q_j &\geq 0 & j = 1, \dots, T \\ I_j &\geq 0 & j = 1, \dots, T \end{aligned}$$

where $\delta(Q_j) = \begin{cases} 0 & \text{if } Q_j = 0 \\ 1 & \text{if } Q_j > 0 \end{cases}$

The equivalent dynamic programming formulation is

$$F(j) = \min \left\{ \min_{1 \leq t < j} \left[a + \sum_{b=t}^{j-1} \sum_{k=b+1}^j rvD(k) + F(t-1) \right], A + F(j-1) \right\}$$

where $F(1) = A$, $F(0) = 0$, and $\sum_{b=t}^{j-1} \sum_{k=b+1}^j rvD(k)$ represents the inventory carrying cost of periods $t + 1$ through j .

$A + F(j - 1)$ represents the case in which a setup is performed in period j , increased by the best solution through period $j - 1$.

Ferreira and Vidal (1984) provide an equivalent linear-programming model.

Let $c_{ij} = \sum_{k=i}^{j-1} rvI_k$ for all $i < j \leq T$.

Define

$$\begin{aligned} z_{ij} &= Q_{ij}/D(j) \text{ for all } i \leq j \\ C_{ij} &= c_{ij}D(j) \text{ for all } i \leq j \\ Y_i &= \delta \left(\sum_{j=1}^T Q_{ij} \right) = \delta \left(\sum_{j=1}^T z_{ij} \right) \text{ for all } i \end{aligned}$$

c_{ij} is the cost of producing in period i for the subsequent use in period j

z_{ij} is the fraction of units to be used in period j that should be produced in period i

Now, relax the requirement that Y_i should be 0 or 1 (and write y_i instead of Y_i). The linear-programming model with decision variables y_i and z_{ij} is

$$\begin{aligned} \min & \left(\sum_{j=1}^T \sum_{i=1}^j C_{ij}z_{ij} + \sum_{i=1}^T Ay_i \right) \\ & \sum_{i=1}^j z_{ij} = 1 \text{ for all } j \\ & y_i - z_{ij} \geq 0 \text{ for all } i \leq j \\ & z_{ij}, y_i \geq 0 \text{ for all } i, j \end{aligned}$$

References

- Baker, K. R. 1989. Lot-sizing procedures and a standard data set: A reconciliation of the literature. *Journal of Manufacturing and Operations Management* 2(3), 199–221.
- Benton, W. C. 1983. Purchase quantity discount procedures and MRP. *Journal of Purchasing and Materials Management* 19(1), 30–34.
- Benton, W. C. 1985. Multiple price breaks and alternative purchase lot-sizing procedures in material requirements planning systems. *International Journal of Production Research* 23(5), 1025–1047.
- Benton, W. C. 1991. Safety stock and service levels in periodic review inventory systems. *Journal of the Operational Research Society* 42(12), 1087–1095.
- Blackburn, J., D. Kropp, and R. Millen. 1986. Comparison of strategies to dampen nervousness in MRP systems. *Management Science* 32(4), 413–429.
- Blackburn, J. and H. Kunreuther. 1974. Planning and forecast horizons for the dynamic lot size model with backlogging. *Management Science* 21(3), 215–255.
- Blackburn, J. and R. Millen. 1980. Heuristic lot-sizing performance in a rolling-schedule environment. *Decision Sciences* 11, 691–701.
- Boe, W. J. and C. Yilmaz. 1983. The incremental order quantity. *Production and Inventory Management 2nd Quarter*, 24, 94–100.
- Bookbinder, J. H. and J. Y. Tan. 1984. Two lot-sizing heuristics for the case of deterministic time-varying demands. *International Journal of Operations and Production Management* 5(4), 30–42.
- Bose, S., A. Goswami, and K. S. Chaudhuri. 1995. An EOQ model for deteriorating items with linear time-dependent demand rate and shortages under inflation and time discounting. *Journal of the Operational Research Society* 46, 771–782.

- Bregman, R. L. 1991. An experimental comparison of MRP purchase discount methods. *Journal of the Operational Research Society* 42(3), 235–245.
- Bregman, R. L. and E. A. Silver. 1993. A modification of the Silver–Meal heuristic to handle MRP purchase discount situations. *Journal of the Operational Research Society* 44(7), 717–723.
- Brousseau, L. J. A. 1982. An inventory replenishment policy for the case of a linear decreasing trend in demand. *INFOR* 20(3), 252–257.
- Brown, R. G. 1977. *Materials Management Systems*. New York: Wiley-Interscience.
- Callarman, T. and D. C. Whybark. 1981. Determining purchase quantities for MRP requirements. *Journal of Purchasing and Materials Management* 17(3), 25–30.
- Carlson, R., J. Jucker, and D. Kropp. 1979. Less nervous MRP systems: A dynamic economic lot-sizing approach. *Management Science* 25(8), 754–761.
- Christoph, O. B. and R. L. La Forge. 1989. The performance of MRP purchase lot-size procedures under actual multiple purchase discount conditions. *Decision Sciences* 20, 348–358.
- Chung, C. S., D. T. Chiang, and C. Y. Lu. 1987. An optimal algorithm for the quantity discount problem. *Journal of Operations Management* 7, 165–177.
- Coleman, B. J. and M. A. McKnew. 1990. A technique for order placement and sizing. *Journal of Purchasing and Materials Management* 26(2), 32–40.
- DeMatta, R. and M. Guignard. 1995. The performance of rolling production schedules in a process industry. *IIE Transactions* 27, 564–573.
- DeMatteis, J. 1968. An economic lot-sizing technique I: The part-period algorithm. *IBM Systems Journal* 7(1), 30–38.
- Diaby, M. 1995. Optimal setup time reduction for a single product with dynamic demands. *European Journal of Operational Research* 85, 532–540.
- Diegel, A. 1966. A linear approach to the dynamic inventory problem. *Management Science* 12(7), 530–540.
- Donaldson, W. A. 1977. Inventory replenishment policy for a linear trend in demand—An analytic solution. *Operational Research Quarterly* 28(3ii), 663–670.
- Eppen, G. D., F. J. Gould, and B. P. Pashigian. 1969. Extensions of the planning horizon theorem in the dynamic lot size model. *Management Science* 15(5), 268–277.
- Federgruen, A. and C.-Y. Lee. 1990. The dynamic lot size model with quantity discount. *Naval Research Logistics* 37, 707–713.
- Federgruen, A. and M. Tzur. 1991. A simple forward algorithm to solve general dynamic lot sizing models with n periods in $O(n \log n)$ or $O(n)$ time. *Management Science* 37(8), 909–925.
- Federgruen, A. and M. Tzur. 1994. Minimal forecast horizons and a new planning procedure for the general dynamic lot sizing model: Nervousness revisited. *Operations Research* 42, 456–469.
- Federgruen, A. and M. Tzur. 1995. Fast solution and detection of minimal forecast horizons in dynamic programs with a single indicator of the future: Applications to dynamic lot-sizing models. *Management Science* 41(5), 874–893.
- Ferreira, J. A. S. and R. V. V. Vidal. 1984. Lot sizing algorithms with applications to engineering and economics. *International Journal of Production Research* 22(4), 575–595.
- Freeland, J. R. and J. L. J. Colley. 1982. A simple heuristic method for lot-sizing in a time-phased reorder system. *Production and Inventory Management* 23(1), 15–22.
- Gorham, T. 1968. Dynamic order quantities. *Production and Inventory Management Journal* 9(1), 75–79.
- Goyal, S. K., M. A. Hariga, and A. Alyan. 1996. The trended inventory lot sizing problem with shortages under a new replenishment policy. *Journal of the Operational Research Society* 27, 1286–1295.
- Groff, G. 1979. A lot sizing rule for time-phased component demand. *Production and Inventory Management Journal* 20(1), 47–53.
- Hariga, M. 1995a. An EOQ model for deteriorating items with shortages and time-varying demand. *Journal of the Operational Research Society* 46, 398–404.
- Hariga, M. 1995b. Comparison of heuristic procedures for the inventory replenishment problem with a linear trend in demand. *Computers and Industrial Engineering* 28(2), 245–258.
- Hariga, M. 1995c. Effects of inflation and time-value of money on an inventory model with time-dependent demand rate and shortages. *European Journal of Operational Research* 81, 512–520.

- Hariga, M. A. 1994. Economic analysis of dynamic inventory models with non-stationary costs and demand. *International Journal of Production Economics* 36, 255–266.
- Hariga, M. A. 1996. Optimal EOQ models for deteriorating items with time varying demand. *Journal of the Operational Research Society* 47, 1228–1246.
- Heady, R. B. and H. Zhu. 1994. An improved implementation of the Wagner–Whitin algorithm. *Production and Operations Management* 3(1), 55–63.
- Hill, R. M. 1995. Inventory models for increasing demand followed by level demand. *Journal of the Operational Research Society* 46, 1250–1259.
- Hu, J., C. Munson, and E. A. Silver. 2004. A modified Silver–Meal heuristic for dynamic lot sizing under incremental quantity discounts. *Journal of the Operational Research Society* 55(6), 671–673.
- Kaimann, R. A. 1969. E.O.Q. vs. dynamic programming—Which one to use for inventory ordering. *Production and Inventory Management Journal* 10(4), 66–74.
- Kang, H.-Y. and A. H. Lee. 2013. A stochastic lot-sizing model with multi-supplier and quantity discounts. *International Journal of Production Research* 51(1), 245–263.
- Karni, R. 1981. Maximum part-period gain (MPG)—A lot sizing procedure for unconstrained and constrained requirements planning systems. *Production and Inventory Management Journal* 22(2), 91–98.
- Kropp, D., R. Carlson, and J. Jucker. 1983. Heuristic lot-sizing approaches for dealing with MRP system nervousness. *Decision Sciences* 14(2), 156–169.
- Kropp, D. and R. C. Carlson. 1984. A lot-sizing algorithm for reducing nervousness in MRP systems. *Management Science* 20(2), 240–244.
- Kunreuther, H. and T. Morton. 1973. Planning horizons for production smoothing with deterministic demands: I. *Management Science* 20(1), 110–125.
- La Forge, R. L. 1985. MRP lot sizing with multiple purchase discounts. *Computer and Operations Research* 12(6), 579–587.
- Lundin, R. and T. Morton. 1975. Planning horizons for the dynamic lot size model: Zabel vs protective procedures and computational results. *Operations Research* 23(4), 711–734.
- Mendoza, A. 1968. An economic lot-sizing technique II: Mathematical analysis of the part-period algorithm. *IBM Systems Journal* 7(1), 39–46.
- Mitra, A., J. Cox, J. Blackstone, and R. Jesse. 1983. A re-examination of lot sizing procedures for requirements planning systems: Some modified rules. *International Journal of Production Research* 21, 471–478.
- Pan, C.-H. 1994. Sensitivity analysis of dynamic lot-sizing heuristics. *OMEGA* 22(3), 251–261.
- Pochet, Y. and L. A. Wolsey. 2006. *Production Planning by Mixed Integer Programming*. New York: Springer.
- Potamianos, J. and A. J. Orman. 1996. An interactive dynamic inventory-production control system. *Journal of the Operational Research Society* 47, 1017–1028.
- Rachamadugu, R. and T. J. Schriber. 1995. Optimal and heuristic policies for lot sizing with learning in set-ups. *Journal of Operations Management*, 13, 229–245.
- Sahin, F., A. Narayanan, and E. P. Robinson. 2013. Rolling horizon planning in supply chains: Review, implications and directions for future research. *International Journal of Production Research* 51(18), 5413–5436.
- Schulz, T. 2011. A new Silver–Meal based heuristic for the single-item dynamic lot sizing problem with returns and remanufacturing. *International Journal of Production Research* 49(9), 2519–2533.
- Silver, E. and J. Miltenburg. 1984. Two modifications of the Silver–Meal lot sizing heuristic. *INFOR* 22(1), 56–68.
- Silver, E. A. and H. Meal. 1969. A simple modification of the EOQ for the case of a varying demand rate. *Production and Inventory Management Journal* 10(4, 4th Qtr), 52–65.
- Silver, E. A. and H. C. Meal. 1973. A heuristic for selecting lot size quantities for the case of a deterministic time-varying demand rate and discrete opportunities for replenishment. *Production and Inventory Management Journal 2nd Quarter*, 14, 64–74.
- Sox, C., L. J. Thomas, and J. McClain. 1997. Coordinating production and inventory to improve service. *Management Science* 43(9), 1189–1197.

- Sridharan, V. and L. Berry. 1990. Freezing the master production schedule under demand uncertainty. *Decision Sciences* 21(1), 97–120.
- Sridharan, V., W. L. Berry, and V. Udayabhanu. 1987. Freezing the master production schedule under rolling planning horizons. *Management Science* 33(9), 1137–1149.
- Teng, J. 1994. A note on inventory replenishment policy for increasing demand. *Journal of the Operational Research Society* 45(11), 1335–1337.
- Ting, P. and K. Chung. 1994. Inventory replenishment policy for deteriorating items with a linear trend in demand considering shortages. *International Journal of Operations and Production Management* 14(8), 102–110.
- Toklu, B. and J. M. Wilson. 1995. An analysis of multi-level lot-sizing problems with a bottleneck under a rolling schedule environment. *International Journal of Production Research* 33(7), 1835–1847.
- Triantaphyllou, E. 1992. A sensitivity analysis of a (ti, Si) inventory policy with increasing demand. *Operations Research Letters* 11(3), 167–172.
- Tzur, M. 1996. Learning in setups: Analysis, minimal forecast horizons, and algorithms. *Management Science* 42(12), 1732–1743.
- Van den Heuvel, W. and A. P. Wagelmans. 2010. Worst-case analysis for a general class of online lot-sizing heuristics. *Operations Research* 58(1), 59–67.
- Vargas, V. 2009. An optimal solution for the stochastic version of the Wagner–Whitin dynamic lot-size model. *European Journal of Operational Research* 198(2), 447–451.
- Vidal, R. V. V. 1970. Operations research in production planning. IMSOR, Technical University of Denmark, Lyngby, Denmark, 95–99.
- Wagelmans, A., S. van Hoesel, and A. Kolen. 1992. Economic lot sizing: An $O(n \log n)$ algorithm that runs in linear time in the Wagner–Whitin case. *Operations Research* 40(1), S145–S156.
- Wagner, H. and T. M. Whitin. 1958. Dynamic problems in the theory of the firm. *Naval Research Logistics Quarterly* 5, 53–74.
- Wee, H.-M. 1995. A deterministic lot-size inventory model for deteriorating items with shortages and a declining market. *Computers and Operations Research* 22(3), 345–356.
- Wemmerlöv, U. 1981. The ubiquitous EOQ—Its relations to discrete lot sizing heuristics. *International Journal of Operations and Production Management* 1, 161–179.
- Wemmerlöv, U. 1983. The part-period balancing algorithm and its look ahead–look back feature: A theoretical and experimental analysis of a single stage lot-sizing procedure. *Journal of Operations Management* 4(1), 23–39.
- Zhou, Y. 1996. Optimal production policy for an item with shortages and increasing time-varying demand. *Journal of the Operational Research Society* 47, 1175–1183.

Chapter 6

Individual Items with Probabilistic Demand

In the preceding two chapters, which dealt with the determination of replenishment quantities, the decision rules resulted from analyses that assumed deterministic demand patterns. In several places, we showed that the relevant costs associated with the selection of order quantities were relatively insensitive to inaccuracies in the estimates of the various factors involved. However, the costs of insufficient capacity in the short run—that is, the costs associated with shortages or with averting them—were not included in the analyses. When demand is no longer assumed to be deterministic, these costs assume a much greater importance. Clearly, the assumption of deterministic demand is inappropriate in many production and distribution situations. Therefore, this chapter is devoted to the theme of how to develop control systems to cope with the more realistic case of probabilistic demand. We initially restrict our attention to the case (comparable to Chapter 4) where the *average* demand remains approximately constant with time. We address the more difficult setting of time-varying probabilistic demand in Chapter 7.

The introduction of uncertainty in the demand pattern significantly complicates the inventory situation from a conceptual standpoint. This, together with the plethora of the possible shortage-costing methods or customer-service measures, has made the understanding and acceptance of decision rules for coping with probabilistic demand far less frequent than is merited by the importance of the problem. It is safe to say (as will be illustrated in Section 6.11.2 of this chapter) that in most organizations, an appropriate reallocation of buffer (or safety) stocks (which are kept to meet the unexpected fluctuations in demand) can lead to a significant improvement in the service provided to customers.

It should be emphasized that in this and several succeeding chapters, we are still dealing with a single-stage problem. A control system different from those proposed here is usually in order for a multiechelon situation. We have more to say on this topic in Chapters 11 and 15.

We begin in Section 6.1 with a careful discussion of some important issues and terminology relevant to the case of probabilistic demand. In Section 6.2, we return to the topic of classification of items into A, B, or C categories. In Section 6.3, the important dichotomy of continuous versus periodic review is discussed. This is followed in Section 6.4 by an explanation of the four most common types of control systems. Section 6.5 is concerned with the many different ways of

measuring service or costing shortages that may be relevant to any particular stocking situation. Section 6.6 deals with an illustrative selection of the reorder point for one type of continuous-review system (order-point, order-quantity system). This section builds the conceptual foundation for the remainder of the chapter. In Section 6.7, we present a detailed treatment of the order-point, order-quantity system, showing the decision rules for a wide range of methods of measuring service and costing shortages. Section 6.8 addresses the implied performance that is obtained when a reorder point is chosen by a different service or cost objective. Section 6.9 is more briefly devoted to another commonly used control system—namely, the periodic-review, order-up-to-level system. This is followed, in Section 6.10, by a discussion of how to deal with appreciable variability in the replenishment lead time. Finally, Section 6.11 presents exchange curves displaying the total buffer stock plotted against the aggregate measures of service as a function of a policy variable; the policy variable is the numerical value of a particular service level or shortage cost.

6.1 Some Important Issues and Terminology

6.1.1 Different Definitions of Stock Level

When demand is probabilistic, it is useful to conceptually categorize the inventories as follows:

1. *On-hand (OH) stock* This is stock that is physically on the shelf; it can never be negative. This quantity is relevant in determining whether a particular customer demand is satisfied directly from the shelf.
2. *Net stock* is defined by the following equation:

$$\text{Net stock} = (\text{On hand}) - (\text{Backorders}) \quad (6.1)$$

This quantity can become negative (namely, if there are backorders). It is used in some mathematical derivations and is also a component of the following important definition.

3. *Inventory position* (sometimes also called the *available stock*)* The inventory position is defined by the relation

$$\begin{aligned} \text{Inventory position} = & (\text{On hand}) + (\text{On order}) \\ & - (\text{Backorders}) - (\text{Committed}) \end{aligned} \quad (6.2)$$

The on-order stock is that stock which has been requisitioned but not yet received by the stocking point under consideration. The “committed” quantity in Equation 6.2 is required if such stock cannot be used for other purposes in the short run. As we will see, the inventory position is a key quantity in deciding when to replenish.

4. *Safety stock (SS)* The SS (or buffer) is defined as the average level of the net stock just before a replenishment arrives. If we planned to just run out, on the average, at the moment when the replenishment arrived, the SS would be zero. A positive SS provides a cushion or buffer against larger-than-average demand during the effective replenishment lead time. Throughout this book, we consistently use the definition provided here, although the appropriate method of calculation, and thus the numerical value of the SS, depends on what happens to

* We have chosen to not use the words “available stock” because of the incorrect connotation that such stock is immediately available for satisfying customer demands.

demands when there is a stockout. If no backorders are permitted, net stock (and thus SS) is always nonnegative. With backorders, it is possible that SS, by the definition provided here, could be negative.

6.1.2 Backorders versus Lost Sales

What happens to a customer's order when an item is temporarily out of stock is of obvious importance in inventory control. There are two extreme cases:

1. *Complete backordering.* Any demand, when out of stock, is backordered and filled as soon as an adequate-sized replenishment arrives. This situation corresponds to a captive market, common in government organizations (particularly the military) and at the wholesale–retail link of some distribution systems (e.g., exclusive dealerships).
2. *Complete lost sales.* Any demand, when out of stock, is lost; the customer goes elsewhere to satisfy his or her need. This situation is most common at the retail–consumer link. For example, a person is unlikely to backorder a demand for a loaf of bread. We focus in this chapter on single-item inventory control models, but it is worth noting that when multiple items are available for purchase, a customer facing an out-of-stock situation may choose a substitute product. This would clearly be a lost sale for the originally intended item, but may affect the anticipated demand.

In most practical situations, there is a combination of these two extremes, whereas most inventory models have been developed for one or the other of the extremes. Nevertheless, most of these models serve as reasonable approximations because the decisions they yield tend to be relatively insensitive to the degree of backordering possible in a particular situation. This is primarily a consequence of the common use in practice of high customer-service levels; high service levels imply infrequent *stockout* occasions. When we use the term stockout, we mean a stockout occasion or event. The number of units backordered or lost is a measure of the impact of the stockout.

We now show that the numerical value of the SS depends on the degree to which backorders or lost sales occur. Consider a particular replenishment lead time in which a stockout occurs. Under complete backordering, if demand occurs during the stockout, the net stock will be negative just before the next replenishment arrives. On the other hand, if all demands that occur during the stockout are lost, then, the net stock will remain at the zero level throughout the stockout period. In other words, in a cycle when a stockout occurs, the value of the net stock just before the replenishment arrives depends on whether backorders can occur. Because SS is defined to be the *average* net stock just before a replenishment arrives, its numerical value is influenced by whether backordering is actually possible.

6.1.3 Three Key Issues to Be Resolved by a Control System under Probabilistic Demand

The fundamental purpose of a replenishment control system is to resolve the following three issues or problems:

1. How often the inventory status should be determined.
2. When a replenishment order should be placed.
3. How large the replenishment order should be.

Under the conditions of deterministic demand (discussed in the previous two chapters), the first issue is trivial because knowing the inventory status at any one point allows us to calculate it at all points in time (at least out to a reasonable horizon). Furthermore, under deterministic demand, the second problem is answered by placing an order such that it arrives precisely when the inventory level hits some prescribed value (usually set at zero). Finally, under deterministic demand, the use of the EOQ, one of its variations, or one of the procedures discussed in Chapter 5 provides the solution to the third problem.

Under probabilistic demand, the answers are more difficult to obtain. Regarding the first point, it takes resources (labor, computer time, etc.) to determine the inventory status. On the other hand, the less frequently the status is determined, the longer is the period over which the system must protect against unforeseen variations in demand in order to provide the desired customer service. The answer to the second problem rests on a trade-off between the costs of ordering somewhat early (hence, carrying extra stock) and the costs (implicit or explicit) of providing inadequate customer service. The factors relevant in answering the third problem are similar to those discussed in the derivation of the basic EOQ, except that under some service criteria, specified by the management, there is an interaction; the answer to the question “When to replenish?” may be affected by the replenishment quantity used.

To respond to these three fundamental issues, a manager must determine several things. We pose these as four questions that managers can use to systematically establish inventory policies.

1. How important is the item?
2. Can, or should, the stock status be reviewed continuously or periodically?
3. What form should the inventory policy take?
4. What specific cost or service objectives should be set?

The next four sections address these four questions in turn.

6.2 The Importance of the Item: A, B, and C Classification

Managers must first establish how critical the item under consideration is to the firm. As mentioned in Chapter 2, many firms segment their items into the categories A, B, or C. Recall that A items make up roughly 20% of the total number of items, but represent 80% of the dollar sales volume. B items comprise roughly 30% of the items, but represent 15% of the dollar volume; C items comprise roughly 50% of the items, and represent only 5% of the dollar volume. Firms often include slow-moving and inexpensive items in the A category if the items are critical to the business. It is possible that including many slow-moving or inexpensive items in the A category could shift the typical 80–20 distribution, but our experience is that this is rarely the case.

The importance of the item helps direct the response to the remaining three questions. The remainder of this chapter is devoted to B items. Chapter 7 covers A items, and Chapter 8 addresses C items.

6.3 Continuous versus Periodic Review

The answer to the problem of how often the inventory status should be determined specifies the review interval (R), which is the time that elapses between two consecutive moments at which we

know the stock level. An extreme case is where there is continuous review; that is, the stock status is always known. In reality, continuous surveillance is usually not required; instead, each transaction (shipment, receipt, demand, etc.) triggers an immediate updating of the status. Consequently, this type of control is often called transactions reporting. Point-of-sale (POS) data collection systems (involving electronic scanners), which permit transactions reporting, are having a profound impact at the retail level. With periodic review, as the name implies, the stock status is determined only every R time units; between the moments of review, there may be considerable uncertainty as to the value of the stock level. Even with POS scanning, inventory decisions are often made periodically (e.g., end of the day). Thus, from the perspective of the inventory reordering decision, the review period, R , is 1 day.

A common example of periodic-review systems is the soda machine in a dormitory. The driver comes regularly, say once every week, to refill the machine. If the machine stocks out of an item between visits, no action is taken. As another example, note that some small firms assign one person to review inventory and make purchase decisions. Often, this person is so busy that he or she can allocate only one afternoon each week to perform these tasks. As these examples show, physical constraints or limits on time often determine the review interval.

We now comment on the advantages and disadvantages of continuous and periodic review. Items may be produced on the same piece of equipment, purchased from the same supplier, or shipped in the same transportation mode. In any of these situations, coordination of replenishments may be attractive. In such a case, periodic review is particularly appealing because all items in a coordinated group can be given the same review interval (e.g., all items purchased from a particular supplier might be scheduled for review every Thursday). Periodic review also allows a reasonable prediction of the level of the workload on the staff involved. In contrast, under continuous review, a replenishment decision can be made at practically any moment in time; hence the load is less predictable. A rhythmic, rather than random, pattern is usually appealing to the staff.

Another disadvantage of continuous review is that it is generally more expensive in terms of reviewing costs and reviewing errors. This is particularly true for fast-moving items where there are many transactions per unit of time. Today, POS data collection systems have dramatically reduced reviewing costs and errors. For extremely slow-moving items, very little reviewing costs are incurred by continuous review because updates are only made when a transaction occurs. On the other hand, we have the anomalous condition that periodic review may be more effective than continuous review in detecting the spoilage (or pilferage) of such slow-moving items; periodic review forces an occasional review of the situation, whereas, in transactions reporting, no automatic review will take place without a transaction occurring.

The major advantage of continuous review is that, to provide the same level of customer service, it requires less SS (hence, lower carrying costs) than does periodic review. This is because the period over which safety protection is required is longer under periodic review (the stock level has the opportunity to drop appreciably between review instants without any reordering action being possible in the interim).

6.4 The Form of the Inventory Policy: Four Types of Control Systems

Once the manager has determined whether the item falls in the A, B, or C category, and he or she has settled the question of continuous versus periodic review, it is time to specify the *form* of the

inventory control policy. The form of the inventory policy will begin to resolve the second and third issues: *When* should an order be placed and *what quantity* should be ordered.

There are a number of possible control systems. The physical operation of the four most common ones will be described in the next section, along with a brief discussion of the advantages and disadvantages of each. Our discussion of the advantages and disadvantages will be rather general in that they are dependent on the specific environment in which the systems are to be implemented.

6.4.1 Order-Point, Order-Quantity (s, Q) System

This is a continuous-review system (i.e., $R = 0$). A fixed quantity Q is ordered whenever the inventory position drops to the reorder point s or lower. Note that the inventory position, and not the net stock, is used to trigger an order. The inventory position, because it includes the on-order stock, takes proper account of the material requested but not yet received from the supplier. In contrast, if net stock was used for ordering purposes, we might unnecessarily place another order today even though a large shipment was due in tomorrow. A good example of ordering on the basis of inventory position is the way a person takes medicine to relieve a headache. After taking the medicine, it is not necessary to take more every 5 minutes until the headache goes away. Rather, it is understood that the relief is *on order*.

The (s, Q) system is often called a two-bin system because one physical form of implementation is to have two bins for storage of an item. As long as units remain in the first bin, demand is satisfied from it. The amount in the second bin corresponds to the order point. Hence, when this second bin is opened, a replenishment is triggered. When the replenishment arrives, the second bin is refilled and the remainder is put into the first bin. It should be noted that the physical two-bin system will operate properly only when no more than one replenishment order is outstanding at any point in time. Thus, to use the system, it may be necessary to adjust Q upward so that it is appreciably larger than the average demand during a lead time.

The advantages of the fixed order-quantity (s, Q) system include that it is quite simple, particularly in the two-bin form, for the stock clerk to understand, that errors are less likely to occur, and that the production requirements for the supplier are predictable. The primary disadvantage of an (s, Q) system is that in its unmodified form, it may not be able to effectively cope with the situation where individual transactions are large; in particular, if the transaction that triggers the replenishment in an (s, Q) system is large enough, then, a replenishment of size Q won't even raise the inventory position above the reorder point. (As a numerical illustration, consider a Q value of 10 together with a demand transaction of size 15 occurring when the position is just 1 unit above s .) Of course, in such a situation, one could instead order an integer multiple of Q where the integer was large enough to raise the inventory position above s .

6.4.2 Order-Point, Order-Up-to-Level (s, S) System

This system again assumes continuous review; and, like the (s, Q) system, a replenishment is made whenever the inventory position drops to the order point s or lower. However, in contrast to the (s, Q) system, a variable replenishment quantity is used, ordering enough to raise the inventory position to the order-up-to-level S . If all demand transactions are unit sized, the two systems are identical because the replenishment requisition will always be made when the inventory position is exactly at s ; that is, $S = s + Q$. If transactions can be larger than unit size, the replenishment quantity in the (s, S) system becomes variable. The difference between (s, Q) and (s, S) systems is illustrated in Figure 6.1. The (s, S) system is frequently referred to as a min-max system because

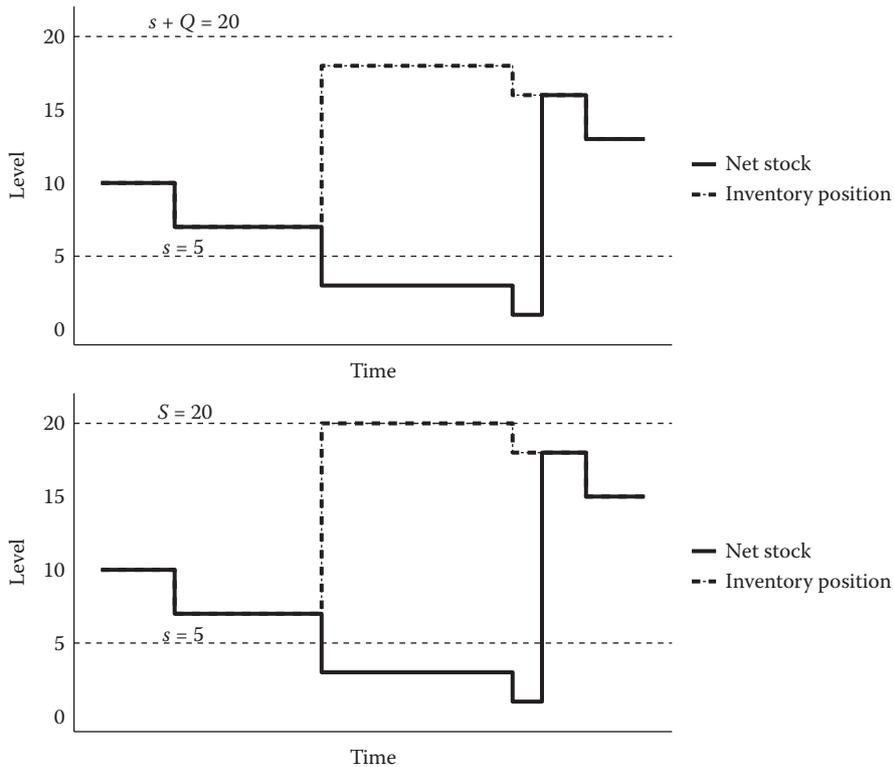


Figure 6.1 Two types of continuous-review systems.

the inventory position, except for a possible momentary drop below the reorder point, is always between a minimum value of s and a maximum value of S .

The best (s, S) system can be shown to have the total costs of replenishment, carrying inventory, and shortage that are not larger than those of the best (s, Q) system. However, the computational effort to find the best (s, S) pair is substantially greater. Thus, the (s, Q) may be the better choice, except perhaps when dealing with an item where the potential savings are appreciable (i.e., an A item). It is interesting that (s, S) systems are frequently encountered in practice. However, the values of the control parameters are usually set in a rather arbitrary fashion. For B items (and even most A items), mathematical optimality does not make sense; instead, we need a fairly simple way of obtaining reasonable values of s and S . This will be discussed further in the next chapter, which deals with A items. One disadvantage of the (s, S) system is the variable order quantity. Suppliers could make errors more frequently; and they certainly prefer the predictability of a fixed order quantity, particularly if the predetermined lot size is convenient from a packaging or handling standpoint (e.g., pallet, container, or truckload.)

6.4.3 Periodic-Review, Order-Up-to-Level (R, S) System

This system, also known as a replenishment cycle system, is in common use, particularly in companies without sophisticated computer control. It is also frequently seen when items are ordered from the same supplier, or require resource sharing. The control procedure is that every R units of

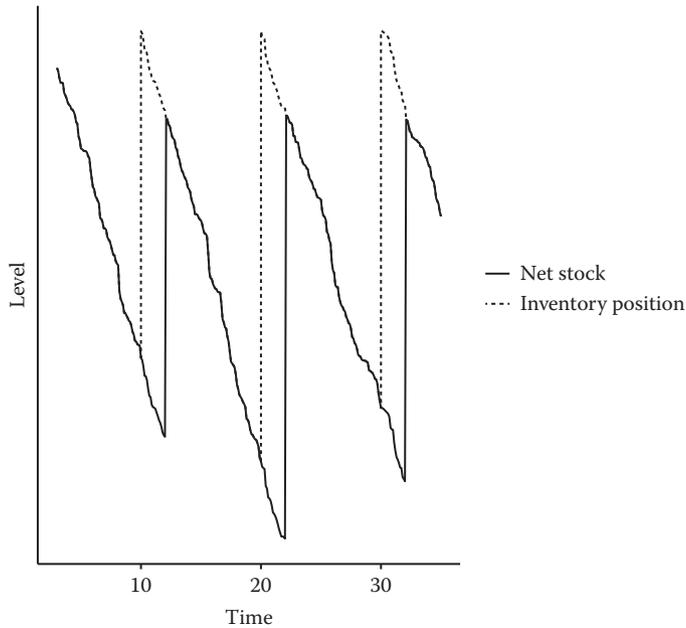


Figure 6.2 The (R, S) system. Orders placed every 10 periods. Lead time of 2 periods.

time (i.e., at each review instant) enough is ordered to raise the inventory position to the level S . A typical behavior of this type of system is shown in Figure 6.2.

Because of the periodic-review property, this system is much preferred to order point systems in terms of coordinating the replenishments of related items. For example, when ordering from overseas, it is often necessary to fill a shipping container to keep shipping costs under control. The coordination afforded by a periodic-review system can provide significant savings. In addition, the (R, S) system offers a regular opportunity (every R units of time) to adjust the order-up-to-level S , a desirable property if the demand pattern is changing with time. The main disadvantages of the (R, S) system are that the replenishment quantities vary and that the carrying costs are higher than in continuous-review systems.

6.4.4 (R, s, S) System

This is a combination of (s, S) and (R, S) systems. The idea is that every R units of time we check the inventory position. If it is at or below the reorder point s , we order enough to raise it to S . If the position is above s , nothing is done until at least the next review instant. The (s, S) system is the special case where $R = 0$, and the (R, S) is the special case where $s = S - 1$. Alternatively, one can think of the (R, s, S) system as a periodic version of the (s, S) system. Also, the (R, S) situation can be viewed as a periodic implementation of (s, S) with $s = S - 1$.

It has been shown (see, e.g., Scarf 1960) that, under quite general assumptions concerning the demand pattern and the cost factors involved, the best (R, s, S) system produces a lower total of replenishment, carrying, and shortage costs than does any other system. However, the computational effort to obtain the best values of the three control parameters is more intense than that

Table 6.1 Rules of Thumb for Selecting the Form of the Inventory Policy

	<i>Continuous Review</i>	<i>Periodic Review</i>
A items	(s, S)	(R, s, S)
B items	(s, Q)	(R, S)

for other systems, certainly for class B items. Therefore, for such items, simplified methods often are used to find reasonable values (again, a topic we discuss in the next chapter). This system is also more difficult to understand than some of the previously mentioned systems. (R, s, S) systems are found in practice where R is selected largely for convenience (e.g., 1 day) even when POS equipment permits continuous review of the inventory position.

Table 6.1 gives a simple rule of thumb for choosing the form of the policy. For C items, firms generally use a more manual and simple approach (which can be equivalent to simple (s, Q) or (R, S) systems). Less effort is devoted to inventory management because the savings available are quite small.

6.5 Specific Cost and Service Objectives

When demand (or delivery capability) is probabilistic, there is a definite chance of not being able to satisfy some of the demand on a routine basis directly out of stock. If demand is unusually large, a stockout may occur or emergency actions may be required to avoid a stockout. On the other hand, if demand is lower than anticipated, the replenishment arrives earlier than needed and inventory is carried. Managers have different perspectives on how to balance these two types of risks. There are four possible methods of modeling these management perspectives to arrive at appropriate decision rules. The choice among these should be consistent with the customers' perceptions of what is important. For example, Kumar and Sharman (1992) describe a firm that sets different service objectives for different items based on cost—higher cost items had higher service targets. Their customers held more inventory of lower cost items, allowing the firm to adjust its policies accordingly. It is critical, however, to recognize that a stockout of a \$1 item could shut down a customer's operation.

1. *SSs Established through the Use of a Simple-Minded Approach.* These approaches typically assign a common safety factor or a common time supply as the SS of each item. Although more easily understood than most of the other procedures to be discussed, we will find that there is a logical flaw in the use of these methods.
2. *SSs Based on Minimizing Cost.* These approaches involve specifying (explicitly or implicitly) a way of costing a shortage and then minimizing the total cost. For instance, air freight may be used to meet a customer demand, even though it costs more. Holding more inventory reduces the probability that air freight will be required, but increases the inventory holding cost. The cost-minimization approach trades off these costs to find the lowest cost policy. Of the many alternative ways of costing a shortage, several will be illustrated shortly. As discussed in Chapter 2, it is important to include only the relevant costs.
3. *SSs Based on Customer Service.* Recognizing the severe difficulties associated with costing shortages, an alternative approach is to introduce a control parameter known as the service

level. The service level becomes a constraint in establishing the SS of an item; for example, it might be possible to minimize the carrying costs of an item subject to satisfying, routinely from stock, 95% of all demands. Again, there is considerable choice in the selection of a service measure. In fact, often in practice, an inventory manager, when queried, may say that company policy is to provide a certain level of service (e.g., 95%) and yet not be able to articulate exactly what is meant by service. Later, we illustrate several of the more common service measures.

4. *SSs Based on Aggregate Considerations.* The idea of this general approach is to establish the SSs of individual items, using a given budget, to provide the best possible aggregate service across a population of items. Equivalently, the selection of the individual SSs is meant to keep the total investment in stocks as low as possible while meeting a desired aggregate service level.

6.5.1 Choosing the Best Approach

Unfortunately, there are no hard-and-fast rules for selecting the appropriate approach and/or measure of service. Which one to use depends on the competitive environment of the particular company. For instance, if the products are in the mature phase of the product life cycle, competition is often based on cost and delivery performance. The presence of many competitors suggests that the cost of a lost sale is simply the lost contribution margin of that sale. Knowing the explicit cost of a shortage allows the firm to minimize inventory costs (or maximize profit). On the other hand, if the products are relatively new to the market, delivery performance may have significant implications for capturing market share; cost is less important. Managers in these cases tend to set stringent service targets. Other factors that can influence this decision include the amount of substitutability of products (i.e., standardized vs. customized products), whether the products are purchased or manufactured, and how the company actually reacts to an impending shortage—for example, does it expedite? In every case, managers should focus on customer needs and desires. Many firms formulate service objectives and work hard to meet them, only to discover that their service objectives are not consistent with what their customers' desire.

Later, we will see equivalencies between certain service measures and methods of costing shortages (a further reference on this topic is Boylan and Johnston 1994). Moreover, it is quite possible that different types of shortage penalties or service measures are appropriate for different classes of items within the same organization.

Our treatment in this section is not meant to be exhaustive; rather, we wish to present the more common measures used in establishing SSs. Because of the wide variety of options for establishing SSs, we provide a summary guide in Table 6.2. We shall show later in the chapter that some of the criteria listed in Table 6.2 are similar to one another (e.g., P_2 and B_3 ; B_2 and TBS), or equivalent (e.g., B_1 and expected total stockout occasions per year [ETSOPY]; B_2 , expected total value of shortages per year [ETVSPY], and TBS). As we have said elsewhere in this chapter, it is not necessary precisely to prescribe the numerical value of a shortage cost or service level.

Finally, the reader should now be aware that maximizing turnover (i.e., minimizing the level of inventories), in itself, is an inadequate criterion for selecting SSs in that it does not take into account of the impact of shortages.

6.5.2 SSs Established through the Use of a Simple-Minded Approach

We illustrate with two of the most frequently used approaches.

Table 6.2 Summary of Different Methods of Selecting the SSs in Control Systems under Probabilistic Demand

<i>Criterion</i>	<i>Discussed in Section</i>	<i>Sections in Which Decision Rules for (s, Q) System Can Be Found^a</i>
Equal time supplies	6.5.2	–
Fixed safety factor	6.5.2	6.7.4
Cost (B_1) per stockout occasion	6.5.3	6.7.5
Fractional charge (B_2) per unit short	6.5.3	6.7.6
Fractional charge (B_3) per unit short per unit time	6.5.3	6.7.7
Cost (B_4) per customer line item backordered	6.5.3	6.7.8
Specified probability (P_1) of no stockout per replenishment cycle	6.5.4	6.7.9
Specified fraction (P_2) of demand to be satisfied directly from the shelf (i.e., the fill rate)	6.5.4	6.7.10
Specified ready rate (P_3)	6.5.4	–
Specified average time (TBS) between stockout occasions	6.5.4	6.7.11
Minimization of ETSOPY subject to a specified total safety stock (TSS)	6.5.5	6.7.12
Minimization of expected total value short per year (ETVSPY) subject to a specified TSS	6.5.5	6.7.13

^a As will be shown in Section 6.9, the decision rules for (R, S) systems are easily obtained from those for (s, Q) systems.

6.5.2.1 Equal Time Supplies

This is a simple, commonly used approach. In fact, a large U.S.-based international consulting firm estimates that 80%–90% of its clients use this approach for setting SS levels. The SSs of a broad group of (if not all) items in an inventory population are set equal to the same time supply; for example, reorder any item when its inventory position minus the forecasted lead time demand drops to a 2-month supply or lower. This approach is seriously in error because it fails to take into account of the difference in the uncertainty of forecasts from item to item. In applying this policy to a number of items, the policy variable is the common number of time periods of supply. In other words, all items in the group have SS set to the same number of periods of supply. One of the authors has worked with a fastener manufacturer that was required by a major U.S. automotive assembler to hold 2 weeks of forecasted demand as SS. This policy worked reasonably well for the assembler, but did not prevent stockouts of items with highly variable demand, and it certainly did not optimize inventory costs.

6.5.2.2 *Equal Safety Factors*

As we will see later, it is convenient to define the safety stock as the product of two factors as follows:

$$SS = k\sigma_L \quad (6.3)$$

where k is called the safety factor and σ_L , as first defined in Chapter 3, is the standard deviation of the errors of forecasts of total demand over a period of duration L (the replenishment lead time). The equal-safety factors approach uses a common value of k for a broad range of items.

6.5.3 *SSs Based on Minimizing Cost*

We present four illustrative cases.

6.5.3.1 *Specified Fixed Cost (B_1) per Stockout Occasion*

Here, it is assumed that the only cost associated with a stockout occasion is a fixed value B_1 , independent of the magnitude or duration of the stockout. This cost arises when a firm expedites to avert an impending stockout.

6.5.3.2 *Specified Fractional Charge (B_2) per Unit Short*

Here, one assumes that a fraction B_2 of unit value is charged per unit short; that is, the cost per unit short of an item i is B_2v_i , where v_i is the unit variable cost of the item. A situation where this type of costing would be appropriate is where units short are made during overtime production (costing a per unit premium).

6.5.3.3 *Specified Fractional Charge (B_3) per Unit Short per Unit Time*

The assumption here is that there is a charge B_3 per dollar short (equivalently, B_3v per unit short) per unit time. An example would be where the item under consideration is a spare part and each unit short results in a machine being idled (with the idle time being equal to the duration of the shortage).

6.5.3.4 *Specified Charge (B_4) per Customer Line Item Short*

In this case, there is a charge B_4 for each customer line item backordered. A customer line item is one of perhaps many lines of a customer's order. Therefore, if a customer orders 10 items, but only nine are in stock, the supplier is short of one customer line item. The assumption here is that there is a charge B_4 per line item short. Many firms impose such penalty payments on their suppliers through a service-level agreement. We revisit the implications of such service-level agreements in Chapter 12.

6.5.4 *SSs Based on Customer Service*

The following are among the more common measures of service.

6.5.4.1 Specified Probability (P_1) of No Stockout per Replenishment Cycle: Cycle Service Level

Equivalently, P_1 is the fraction of cycles in which a stockout does not occur. A stockout is defined as an occasion when the OH stock drops to the zero level. As we will see later, using a common P_1 across a group of items, is equivalent to using a common safety factor k . P_1 service is often called the *cycle service level*.

6.5.4.2 Specified Fraction (P_2) of Demand to Be Satisfied Routinely from Available Inventory: Fill Rate

The fill rate is the fraction of customer demand that is met routinely; that is, without backorders or lost sales. This form of service has considerable appeal to practitioners (particularly where a significant portion of the replenishment lead time is unalterable; e.g., a branch warehouse where the major part of the lead time is the transit time by rail car). It can be shown under certain conditions that the use of the B_3 shortage-costing measure leads to a decision rule equivalent to that for the P_2 service measure, where the equivalence is given by the relation

$$P_2 = \frac{B_3}{B_3 + r} \quad (6.4)$$

with r , as earlier, being the carrying charge.*

6.5.4.3 Specified Fraction of Time (P_3) during Which Net Stock Is Positive: Ready Rate

The ready rate is the fraction of time during which the net stock is positive; that is, there is some stock on the shelf. The ready rate finds common application in the case of equipment used for emergency purposes (e.g., military hardware or bank computer system). Under Poisson demand, this measure is equivalent with the P_2 measure. Solving for the optimal inventory policy in this case is more complex in general than for other service measures, and will not be pursued in this chapter. See Schneider (1981) for a discussion of the ready rate and other service measures. Hopp et al. (1999) employ this service measure, using approximations similar to those developed in this chapter, for a distribution center that stocks 30,000 parts.

6.5.4.4 Specified Average Time (TBS) between Stockout Occasions

Equivalently, one could use the reciprocal of TBS, which represents the desired average number of stockout occasions per year. If each stockout occasion is dealt with by an expediting action, then, a specific TBS value can be selected to result in a tolerable number of expediting actions.

Boylan and Johnston (1994) define other service measures, including a fraction of order lines filled, and illustrate the relationships among them. See also Aardal et al. (1989). A generalization of the fill rate is the fraction of demand that must be met within a specified time (see Problem 6.8 as well as Van der Veen 1981, 1984). Another generalization is the fraction of demand met in each

* This relationship holds when all unmet demand is backlogged and specific assumptions are made about how demand occurs over time. See Rudi et al. (2009).

period, rather than as a long-run average. See Chen and Krass (1995). Song (1998) investigates the order fill rate, or the probability that an entire customer order, composed of possibly more than one item, is met from the shelf. See also Ernst and Pyke (1992). Rudi et al. (2009) investigate how the method of inventory accounting (continuous-time vs. end-of-period accounting) affects the appropriate service-level choice.

6.5.5 *SSs Based on Aggregate Considerations*

We present two common methods of establishing SSs based on aggregate considerations. In both cases, rather than just aggregating cost or service effects across individual items, the item-level measure can be weighted by a factor known as the essentiality of the item. For example, if a particular item was deemed twice as important as another from the perspective of inventory availability, then, its essentiality would be double that of the other item.

6.5.5.1 *Allocation of a Given Total Safety Stock among Items to Minimize the ETSOPY*

As will be shown in Section 6A.4 of the Appendix of this chapter, allocating a fixed total safety stock (TSS) among several items to minimize the expected total number of stockout occasions per year leads to a decision rule for selecting the safety factor of each item based on B_1 . In fact, the decision rule is identical with that obtained by assuming a value (the same for all items) of B_1 , and then selecting the safety factor to keep the total of carrying and stockout costs as low as possible. This allocation interpretation is probably more appealing to management.

6.5.5.2 *Allocation of a Given TSS among Items to Minimize the ETVSPY*

It can be shown that allocating a fixed TSS among a group of items to minimize the ETVSPY (in dollars) leads to a decision rule for selecting the safety factor that is identical with either

1. The one obtained by assuming a value (the same for all items) of the factor B_2 and then selecting the safety factor for each item to minimize the total of carrying and shortage costs.
2. The one obtained by specifying the same average time (TBS) between stockout occasions for every item in the group.

Again, this aggregate view of allocating a limited resource may be considerably more appealing to management than attempting to explicitly ascertain a B_2 value or the somewhat arbitrary specification of a TBS value.

6.6 Two Examples of Finding the Reorder Point s in a Continuous-Review, Order-Point, Order-Quantity (s, Q) System

We now present a simple illustration of the use of SS to protect against a stockout over a replenishment lead time. Then, we use an empirical distribution to build intuition about the operation of an (s, Q) system. In Section 6.7, we will present decision rules for a broad range of service measures and methods of costing shortages.

the replenishment lead time is less than the reorder point. Except where explicitly noted otherwise, we make the simplifying assumption that the undershoots are small enough to be neglected. We explicitly consider undershoots in the next chapter.

This, and the subsequent, discussion assumes that the manager will not take expediting action when faced with an impending stockout. In other words, the replenishment lead time is fixed and cannot be changed. In many cases, however, it is possible to expedite some or all of the replenishment order. If the cost of expediting is low relative to the cost of carrying the inventory, managers should carry less inventory and expedite more often. Likewise, if the cost of expediting is low relative to the cost of stocking out, managers should provide very high service by expediting when necessary. When expediting is expensive, however, more inventory should be carried, and it may be optimal to provide somewhat lower service. See Moinzadeh and Nahmias (1988) for extensions to the standard (s, Q) policy for this case. Also, see Moinzadeh and Schmidt (1991) for the case in which $Q = 1$. Jain et al. (2010, 2011) investigate the impact of the capability to expedite on continuous- and periodic-review models, respectively, and Veeraraghavan and Scheller-Wolf (2008) develop an easily implementable policy for a periodic review inventory system with two modes of supply.

6.6.2 An Example Using a Discrete Distribution

To develop intuition, we now present a stylized example using a discrete demand distribution. We will formalize the treatment of some of these issues in the next section, where we present a more realistic demand distribution. In answering the four questions introduced in Section 6.1, assume that we have a B item, and that it is possible to review the inventory status continuously. Based on Table 6.1, we choose an (s, Q) policy. The lead time demand distribution is presented in Table 6.3.

Assume that the lead time is 1 week, and that the order quantity is 20 units. We can see from the demand distribution that the expected demand per week (or per lead time) is 2.2 units.* Thus, if the firm operates for 50 weeks per year, the annual demand is $50(2.2) = 110$ units.

Let us also assume that the firm has established that the fixed cost per order, A , is \$18, and the cost per unit of inventory held per year is \$10. Furthermore, assume that all unmet demand is lost rather than backordered. This will simplify our example as net stock for this setting must

Table 6.3 Lead Time Demand

<i>Lead Time Demand</i>	<i>Fraction of Occurrences at This Value</i>	<i>Cumulative Fraction at This Value or Lower</i>
0	0.1	0.1
1	0.2	0.3
2	0.3	0.6
3	0.2	0.8
4	0.2	1

* The expected value of a discrete, nonnegative random variable, X , is $\sum_{x=0}^{\infty} x \Pr(X = x)$, where $\Pr(X = x)$ is the probability that X takes on the value x .

be nonnegative. If we wanted to achieve a cycle service level, P_1 , of 90%, it is evident from the lead time demand distribution that a reorder point of 4 is necessary to be certain that no stockout occurs in (at least) 90% of replenishment cycles. A reorder point of 3 would provide a P_1 of only 80%. Thus, if the P_1 requirement were, say, 75%, a reorder point of 3 would be sufficient.

Now, let us consider in more detail a shortage-costing method using the cost per unit short, B_2v where the firm has established that B_2v is \$20. The total annual cost is comprised of ordering cost, holding cost, and shortage cost.

6.6.2.1 Ordering Cost

First, let us determine how many orders are placed each year. In a similar fashion to Chapter 4, the number of orders (or replenishment cycles) per year is

$$\text{Annual demand/Order size} = 110/20 = 5.5 \text{ orders per year}$$

Because the cost per order is \$18, the annual fixed cost of ordering is $\$18(5.5) = \99 . See Section 6A.6 of the Appendix to the chapter for a more technical discussion on finding the expected costs.

6.6.2.2 Holding Cost

To determine the holding cost, we refer to aspects of the development of the EOQ from Chapter 4. Recall that the inventory level follows the familiar sawtooth of Figure 6.4. The average inventory OH, in that simple case, is $Q/2$. Now, when we add the possibility of SS and probabilistic demand,

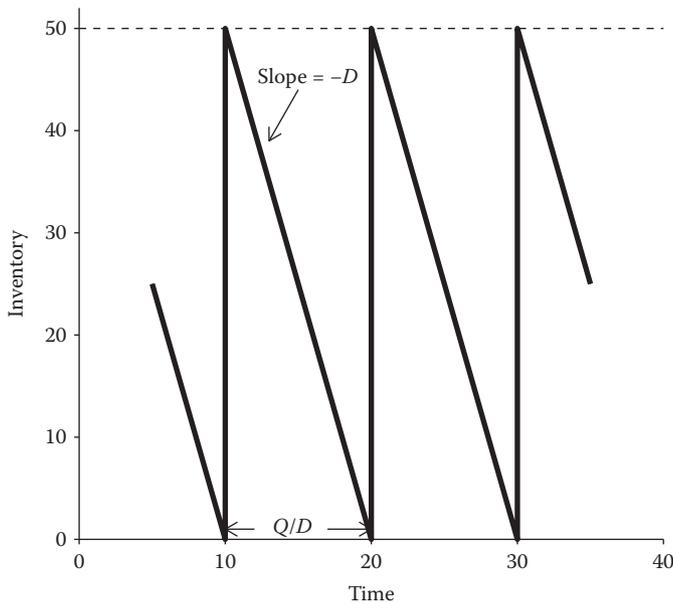


Figure 6.4 Behavior of inventory level with time.

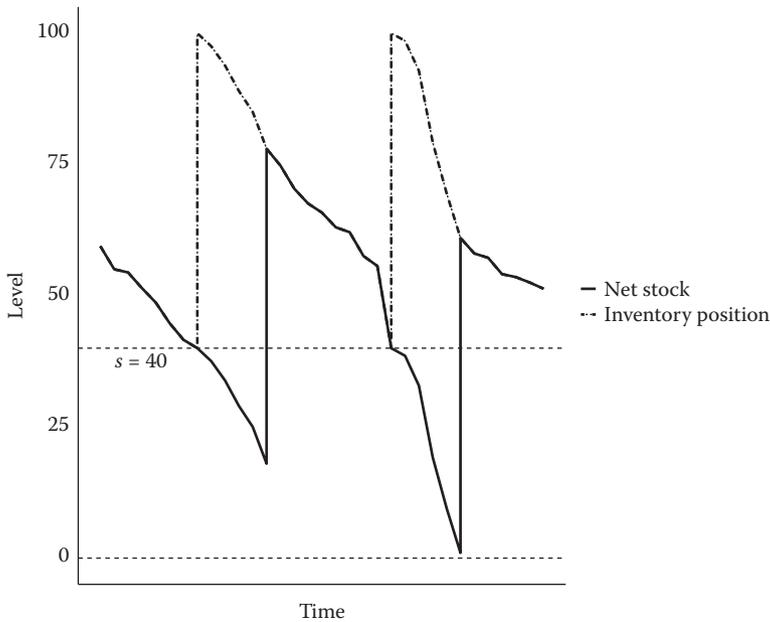


Figure 6.5 Behavior of inventory level with time: probabilistic demand.

the picture looks more like Figure 6.5. Intuitively, imagine the sawtooth of Figure 6.4 floating on a “buffer”—which we call SS. Hence, we can informally suggest that the average inventory OH is $Q/2 + \text{safety stock}$. (We shall present a more formal treatment in the next section.)

Now we know that $Q/2$ is $20/2 = 10$. But what is the SS? For the moment, let the reorder point, s , be 2. Then when lead time demand is 0, we will have 2 units left when the replenishment order arrives. When lead time demand is 1 unit, we will have 1 unit left. When lead time demand is 2, 3, or 4, there is no OH inventory when the replenishment arrives. Since no backorders are allowed, OH inventory and net stock are the same. Table 6.4 fills out the details.

Thus, the safety stock, or the expected amount OH (or net stock) just before the replenishment arrives, is 0.4 units $((0.1)(2) + (0.2)(1))$. So the expected annual holding cost is

$$(Q/2 + SS)(\$10 \text{ per unit}) = (10.4)(\$10) = \$104$$

Table 6.4 Safety Stock

<i>Lead Time Demand</i>	<i>Probability</i>	<i>Amount OH Just before Replenishment Arrives</i>
0	0.1	2
1	0.2	1
2	0.3	0
3	0.2	0
4	0.2	0

Note that we could express the expected SS as

$$\sum_{x=0}^{s-1} (s - x) \Pr(X = x) \tag{6.5}$$

if X is the lead time demand.

6.6.2.3 Shortage Cost

To determine the shortage cost, we need to establish how many units we would expect to be short in each replenishment cycle; and we follow an approach similar to that for the holding cost. If s is 2, and lead time demand is 0, there will be no shortages. This occurs with probability 0.1. If demand is 3, say, we will be one unit short. Again, a simple table describes the events (Table 6.5).

Notice that when lead time demand is 2, and the reorder point is 2, there are no shortages. No customers are turned away without having their demand satisfied, even though there is no inventory left when the replenishment order arrives. The expected number of units short in a cycle is therefore 0.6 units. Because there are $110/20 = 5.5$ cycles per year, and the cost per unit short is \$20, the expected annual shortage cost is

$$\begin{aligned} & (\text{Expected number of units short per cycle}) (\text{Number of cycles per year}) \\ & (\text{Cost per unit short}) = (0.6)(5.5)(\$20) = \$66 \end{aligned}$$

Note that we could express expected units short as

$$\sum_{x=s+1}^{\infty} (x - s) \Pr(X = x) \tag{6.6}$$

6.6.2.4 Total Cost

The total annual cost is therefore $\$99 + \$104 + \$66 = \269 . We can now search for the optimal value of s . These equations and search can be easily implemented on a spreadsheet. The results are given in Table 6.6. We can see that the optimal reorder point is 4.

There are several weaknesses of the simple procedure illustrated here:

Table 6.5 Shortages

<i>Lead Time Demand</i>	<i>Probability</i>	<i>Units Short</i>
0	0.1	0
1	0.2	0
2	0.3	0
3	0.2	1
4	0.2	2

Table 6.6 Optimal Reorder Point

<i>Reorder Point</i>	<i>Ordering Cost</i>	<i>Shortage Cost</i>	<i>Inventory Cost</i>	<i>Total Cost</i>
0	\$99	\$242	\$100	\$441
1	\$99	\$143	\$101	\$343
2	\$99	\$66	\$104	\$269
3	\$99	\$22	\$110	\$231
4	\$99	\$0	\$118	\$217
5	\$99	\$0	\$128	\$227

1. Demand information is rarely presented in the manner described in this section. If it is, there is a good chance that the right tail of the distribution (extremely high demand events) will not be adequately represented, unless we have information from a considerable number of historical lead times. Because service levels are often high, it is this right tail that is the most important. Thus, this simple procedure may not be accurate. (We will have more to say about a similar procedure in Section 6.10.)
2. If the replenishment lead time changes, for some reason, the current information about lead time demand would be of limited value. (We will have more to say about changing the givens in Section 6.7.1.)

To help overcome these weaknesses, it is helpful to fit a member of a family of probability distributions to the available data. In the next section, we do this for the so-called normal probability distribution. Nevertheless, the basic ideas presented in this section are the conceptual foundation for the more formal and extensive treatment presented next.

6.7 Decision Rules for Continuous-Review, Order-Point, Order-Quantity (s, Q) Control Systems

As discussed in Section 6.5, there is a wide choice of criteria for establishing Ss. The choice of a criterion is a strategic decision, to be executed on a relatively infrequent basis and involving senior management directly (the exchange curves of Section 6.11 will be helpful in this regard). Once a criterion (and the implied associated decision rule) is chosen, there is then the tactical issue of the selection of a value of the associated policy variable (e.g., choosing that the target value of P_2 should be 98%).

Besides the choice of a suitable criterion, there is also a choice of the probability distribution of lead time demand. As discussed in Section 3.5.5, from a pragmatic standpoint, we recommend, at least for most B items, the use of a normal distribution of lead time demand. In Section 6.7.14, we discuss the use of other distributions. In particular, we will see that the Gamma distribution can be used almost as easily as the Normal to fit demand during lead time and may be a better choice in certain circumstances.

The assumptions, notation, general approach to the selection of the reorder point, and the portion of the derivation of the rules common to all criteria will be presented prior to showing the

individual decision rules. Further details of some of the derivations can be found in the Appendix of this chapter.

6.7.1 Common Assumptions and Notation

There are a number of assumptions that hold independent of the method of costing shortages or measuring service. These include:

1. Although demand is probabilistic, the average demand rate changes very little with time. Although this assumption may appear to be somewhat unrealistic, note that the decision rules can be used adaptively (i.e., the parameters are updated with the passage of time). The situation where the demand rate changes over time is addressed further in Section 7.6.
2. A replenishment order of size Q is placed when the inventory position is exactly at the order point s . This assumption is tantamount to assuming that all demand transactions are of unit size or else that the undershoots of the order point are of negligible magnitude compared with the total lead time demand. (In Chapter 7, we will consider a situation where the undershoots are not neglected.)
3. If two or more replenishment orders for the same item are simultaneously outstanding, then, they must be received in the same order in which they were placed; that is, crossing of orders is not permitted. A special case satisfying this assumption is where the replenishment lead time is a constant.
4. Unit shortage costs (explicit or implicit) are so high that a practical operating procedure will always result in the average level of backorders being negligibly small when compared with the average level of the OH stock. In practice, this is not a very limiting assumption since, if the level of backorders far exceeds the average OH stock, we are really managing a make (or buy)-to-order business rather than a make-to-stock business.
5. Forecast errors have a normal distribution with no bias (i.e., the average value of the error is zero) and a known standard deviation σ_L for forecasts over a lead time L . In actual fact, as discussed in Chapter 3, the forecast system only provides us with an estimate $\hat{\sigma}_L$ of σ_L . However, extensive simulation studies by Ehrhardt (1979) have revealed that performance is not seriously degraded by using $\hat{\sigma}_L$ instead of σ_L (see Problem 6.5). (Recall that we use the standard deviation of forecast errors rather than the standard deviation of the true lead time demand distribution.) Appendix II, at the end of the book, is devoted to a discussion of the normal distribution, and a graphical representation of a typical distribution is shown in Figure 6.6.
6. Where a value of Q is needed, it is assumed to have been predetermined. In most situations, the effects of the two decision variables, s and Q , are not independent; that is, the best value of Q depends on the s value, and vice versa. However, as will be shown in Chapter 7, where we will take a closer look at the simultaneous determination of the two control parameters, the assumption of Q being predetermined without the knowledge of s makes very good practical sense, particularly for B items. See also Zheng (1992) and Axsäter (1996).
7. The operating costs of the control system do not depend on the specific value of s selected.

Throughout the book, we have been emphasizing the possibility, and requirement, to “change the givens.” In Chapter 4, we noted that there are many ways to reduce the optimal order quantity, such as the use of technological infrastructure (such as EDI) to effectively reduce the ordering cost.

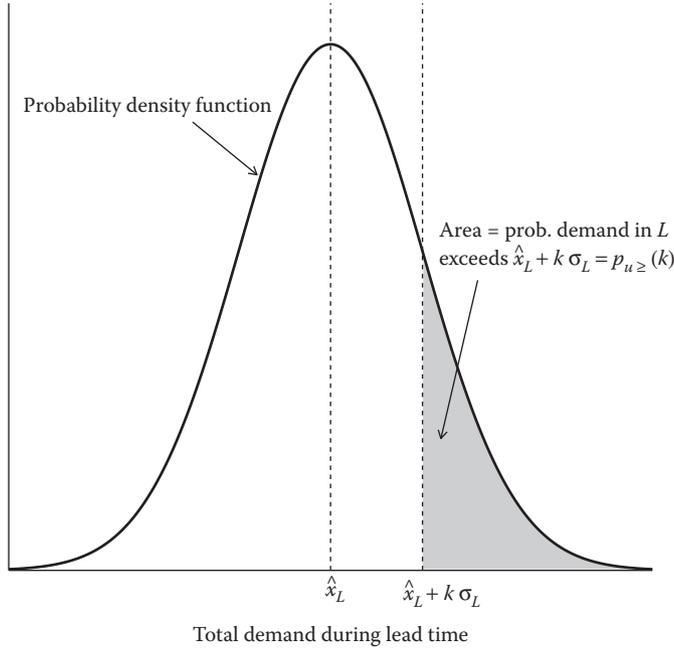


Figure 6.6 Normally distributed forecast errors.

Here, we comment on ways to reduce SS by changing the givens. It will become evident below that some of the important drivers for the amount of SS are the replenishment lead time, the variability of demand, and the service level required. Many firms have focused (rightly) on reducing SS by reducing the replenishment lead time. Choosing a supplier that is closer to your facility and shipping via a faster transportation mode such as air freight are the two ways of reducing this lead time. Improving forecast accuracy and providing customer incentives for specific purchase times and quantities are examples of effective ways to reduce demand variability. Occasionally, it is possible to reduce the required service level by better understanding customer needs and expectations. Once again, our recommendation is to optimize inventory levels given the parameters as they are; and then devote resources to changing the givens. We shall see in Section 6.11.2 that in many cases, it is possible to reduce inventory costs, and improve the service, simultaneously. This is why we suggest initially optimizing inventory levels before applying resources to changing the givens. Then, however, managers should focus attention on reducing inventories by removing the reasons for them.

The common notation includes:

D = demand per year in units/year.

$G_u(k)$ = a special function of the unit normal (mean 0, standard deviation 1) variable.* $G_u(k)$ is used in finding the expected shortages per replenishment cycle (ESPRC). See Equation 6.6 for intuition on this function.

* $G_u(k)$ or $p_{u \geq}(k)$ can be obtained from k , or vice versa, using a table lookup (a table of the unit normal distribution is shown in Appendix II) or a rational approximation presented in Appendix III. Also, as discussed in Appendix III, $p_{u \geq}(k)$ can also be obtained on Excel spreadsheet from Microsoft: given k , $p_{u \geq}(k) = 1 - \text{NORMSDIST}(k)$. Given $p_{u \geq}(k)$, $k = \text{NORMSINV}(1 - p_{u \geq}(k))$.

k = safety factor.

L = replenishment lead time, in years.

$p_{u \geq}(k)$ = probability that a unit normal (mean 0, standard deviation 1) variable takes on a value of k or larger. (A graphical interpretation of $p_{u \geq}(k)$ is shown in Figure 6.6.) Note that $p_{u \geq}(k)$ is often expressed as $1 - \Phi(k)$, where $\Phi(k)$ is the cumulative distribution function (or the left tail) of the unit normal evaluated at k .

Q = prespecified order quantity, in units.

r = inventory carrying charge, in \$/\$/year.

s = order point, in units.

SS = safety stock, in units.

v = unit variable cost, in \$/unit.

\hat{x}_L = forecast (or expected) demand over a replenishment lead time, in units.

σ_L = standard deviation of errors of forecasts over a replenishment lead time, in units.

6.7.2 General Approach to Establishing the Value of s

In Section 6.6.2, we set about directly determining the required reorder point s . Here, it will turn out to be more appropriate to work indirectly using the following relationships:

$$\text{Reorder point, } s = \hat{x}_L + (\text{Safety stock}) \quad (6.7)$$

and

$$\text{Safety stock} = k\sigma_L \quad (6.8)$$

where k is known as the safety factor. The determination of a k value leads directly to a value of s through the use of these two relations.*

The general logic used in computing the appropriate value of s (via k) is portrayed in Figure 6.7. Of particular note is the manual override. The user should definitely have the option of adjusting the reorder point to reflect factors that are not included in the model. However, care must be taken to not again adjust for factors for which manual adjustments were already made as part of the forecasting system (see Figure 3.1 of Chapter 3).

Figure 6.7 suggests specifying the lowest allowable value for k . Management must specify this value because the inventory calculations may recommend, in effect, holding no inventory at all. When the shortage cost is lower than the inventory holding cost, for instance, the logical choice would be to backorder Q customers and then fill their demands with an order of size Q . However, managers may resist providing such poor service. (It is not unlikely that, in this case, the shortage cost was incorrectly specified.) Johnson et al. (1995) point out, however, that k can be negative in multilevel distribution systems. Retailers hold buffer stocks, so that their supplier, a central warehouse, is not required to provide high levels of service. In any case, managers may set a limit on the service provided by specifying that k must be greater than, say, 0.

* The definition of safety stock in Equation 6.8 assumes that unmet demand is backlogged; thus, net stock can be negative. Below, we discuss the adjustments to safety stock calculations when unmet demand is partially or completely lost.

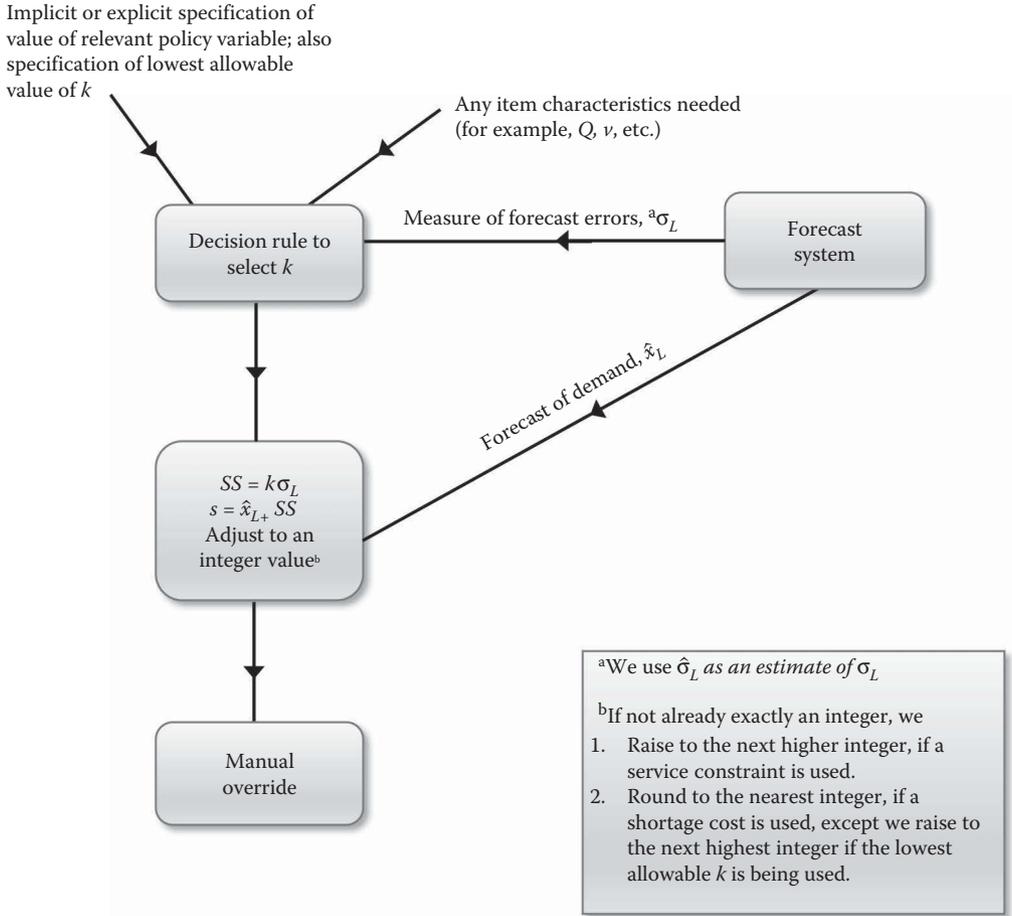


Figure 6.7 General decision logic used in computing the value of *s*.

6.7.3 Common Derivation

Recall that an (s, Q) system operates in the following manner. Any time that the inventory position drops to s or lower, a replenishment of size Q is placed. Because of the assumption of no crossing of orders, if an order is placed at some time t when the inventory position is at level s , then, all previous orders outstanding at that time must have arrived prior to the moment (call it $t + L$) at which the current order arrives. Furthermore, any orders placed after the current one cannot arrive before $t + L$. In other words, all of the inventory position of s units at time t , and no other stock, must have reached the stocking shelf by time $t + L$. Therefore, the service impact of placing the current order when the inventory position is at the level s is determined by whether the total demand x in the replenishment lead time exceeds s . (If the OH level happens to be very low at time t , a stockout may be incurred early in the replenishment cycle before an earlier outstanding order arrives. However, this event is independent of the current replenishment, the one at time t , and should not be considered in evaluating the consequences of using an order point s for the current order.)

If the demand (x) in the replenishment lead time has a probability density function $f_x(x_0)$ defined such that

$$f_x(x_0)dx_0 = \Pr\{\text{total demand in the lead time lies between } x_0 \text{ and } x_0 + dx_0\}$$

then the above arguments lead to the following three important results:

1. Safety stock is the expected net stock just before the replenishment arrives. If unmet demand is backlogged,

$$SS = \int_0^{\infty} (s - x_0)f_x(x_0)dx_0 = s - \hat{x}_L \tag{6.9}$$

This has a particularly simple interpretation for the case where no more than one order is ever outstanding. The average OH inventory level just before a replenishment arrives is equal to the inventory level when the replenishment is placed reduced by the average demand during the lead time.

If unmet demand is lost (no backorders), net stock cannot be negative and an adjustment must be made to the SS calculation:

$$E(\text{OH just before a replenishment arrives}) = \int_0^s (s - x_0)f_x(x_0)dx_0$$

When expected unmet demand is small, the difference in the SS calculation from incorporating this adjustment is also small. We discuss in point 3 below how to operationalize this adjustment.

2. Prob {stockout in a replenishment lead time}

$$\begin{aligned} &= \Pr\{x \geq s\} \\ &= \int_s^{\infty} f_x(x_0)dx_0 \end{aligned} \tag{6.10}$$

= the probability that lead time demand is at least as large as the reorder point (illustrated graphically in Figure 6.6).

3. Expected shortage per replenishment cycle,

$$ESPRC = \int_s^{\infty} (x_0 - s)f_x(x_0)dx_0 \tag{6.11}$$

Using the expression in Equation 6.11, we can develop an expression for SS for the lost sales case mentioned in point 1 above.

$$\begin{aligned}
 E(\text{OH}) &= \int_0^s (s - x_0) f_x(x_0) dx_0 \\
 &= \int_0^s (s - x_0) f_x(x_0) dx_0 + \int_s^\infty (s - x_0) f_x(x_0) dx_0 - \int_s^\infty (s - x_0) f_x(x_0) dx_0 \\
 &= s - \hat{x}_L + \int_s^\infty (x_0 - s) f_x(x_0) dx_0 \\
 &= s - \hat{x}_L + \text{ESPRC}
 \end{aligned} \tag{6.12}$$

From here, we proceed assuming the simpler case of backlogged demand (with SS given by $s - \hat{x}_L$), but note that it is a simple adjustment to the SS expression (the inclusion of ESPRC) to accommodate the lost sales case.

Now, the mean rate of demand is constant with time. Therefore, on average, the OH level drops linearly during a cycle from $(s - \hat{x}_L + Q)$ right after a replenishment arrives to $(s - \hat{x}_L)$ immediately before the next replenishment arrives. Thus

$$E(\text{OH}) \approx \frac{Q}{2} + (s - \hat{x}_L) = \frac{Q}{2} + k\sigma_L \tag{6.13}$$

where, as mentioned earlier, we have chosen to express the safety stock as the multiple of two factors:

$$\text{SS} = k\sigma_L \tag{6.14}$$

A useful graph of Equation 6.13 is shown in Figure 6.8. One other common feature, independent of the service measure or shortage-costing method used, is the expected number of replenishments per year. Each replenishment is of size Q and the mean rate of demand is D , as earlier. Therefore,

$$\text{Expected number of replenishments per year} = \frac{D}{Q} \tag{6.15}$$

To this point, the results hold for any probability distribution of lead time demand (or forecast errors). To proceed further requires specifying the particular distribution. As shown in Appendix II, when forecast errors are assumed to be normally distributed and the SS is expressed as in Equation 6.14, then Equations 6.10 and 6.11 simplify to

$$\text{Pr}\{\text{stockout in a replenishment lead time}\} = p_{u \geq}(k) \tag{6.16}$$

and

$$\text{ESPRC} = \sigma_L G_u(k) \tag{6.17}$$

At this point, the safety factor derivations diverge based on the particular shortage cost or service measure used. Some illustrative details can be found in the Appendix of this chapter. We now turn

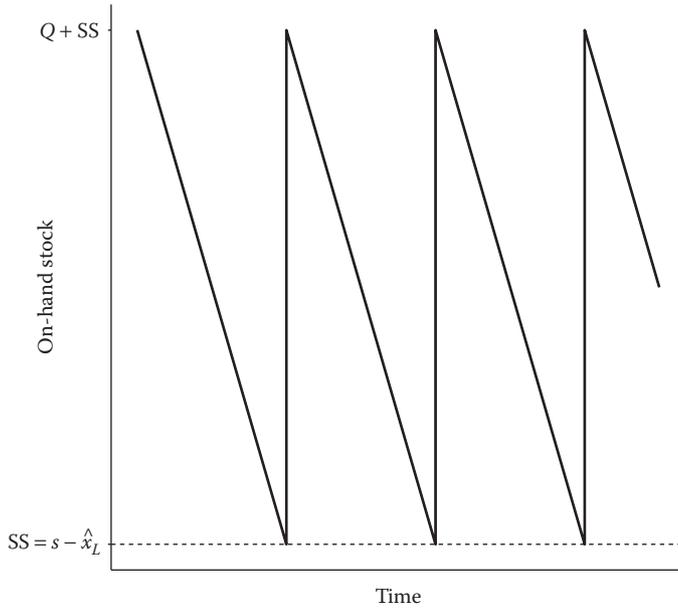


Figure 6.8 Average behavior of OH stock in an (s, Q) system.

to the presentation of the individual decision rules. In each case, we first present the rule. In most cases, this is followed by a numerical illustration of its use, and then by a discussion of the behavior of the rule in terms of how the safety factor k varies from item to item in a population of items. As discussed in Section 6.5, there is a wide choice of criteria for establishing SSs; our treatment is only meant to be illustrative of the many possibilities.

6.7.4 Decision Rule for a Specified Safety Factor (k)

Once a k value is specified, the choice of s immediately follows from Equations 6.7 and 6.8:

The Rule

Step 1 Safety stock, $SS = k\sigma_L$.

Step 2 Reorder point, $s = \hat{x}_L + SS$, increased to the next higher integer (if not already exactly an integer).

6.7.5 Decision Rule for a Specified Cost (B_1) per Stockout Occasion

Two approaches to the derivation of the following rule are shown in Section 6A.1 of the Appendix of this chapter. Here, we simply present the total cost function and then give the decision rule. The total cost is again comprised of ordering, holding, and shortage costs. Relevant expected annual ordering (or replenishment) and holding costs are, as before,

$$C_r = AD/Q$$

and

$$C_c = \left(\frac{Q}{2} + k\sigma_L \right) vr$$

The expected stockout costs C_s per year are obtained by multiplying three factors together—namely, (1) the expected number of replenishment cycles per year, (2) the probability of a stockout per cycle, and (3) the cost per stockout. Thus, using Equations 6.15 and 6.16, we have

$$C_s = \left(\frac{D}{Q} \right) p_{u \geq}(k) B_1$$

The total cost is the sum of these three terms. Because spreadsheets have the standard normal distribution built in, computing the total cost for a given k is quite simple. Searching for the optimal k is easy as well; but when there are thousands of items, it is useful to have a simple decision rule:

Step 1 Is

$$\frac{DB_1}{\sqrt{2\pi}Qv\sigma_L r} < 1? \quad (6.18)$$

where Q has been predetermined, presumably by one of the methods of Chapter 4, B_1 is expressed in dollars (or any other currency), and all the other variables (with units consistent such that the left-hand side of the equation is dimensionless) are as defined in Section 6.7.1. If yes, then go to Step 2. If no, then continue with*

$$k = \sqrt{2 \ln \left(\frac{DB_1}{\sqrt{2\pi}Qv\sigma_L r} \right)} \quad (6.20)$$

If Equation 6.20 gives a value of k lower than the minimum allowable value specified by the management, then go to Step 2. Otherwise, proceed directly to Step 3.

Step 2 Set k at its lowest allowable value (specified by management).

Step 3 Reorder point $s = \hat{x}_L + k\sigma_L$, rounded to the nearest integer (except raised to the next highest integer if Step 2 was used).

6.7.5.1 Numerical Illustration

Suppose an item is one of a number for which a B_1 value of \$300 has been specified. Other relevant characteristics of the item are

$$D = 200 \text{ units/year}$$

$$Q = 129 \text{ units}$$

* An alternative to the use of Equation 6.20 is to find by a table lookup the k value that satisfies

$$f_u(k) = \frac{Qv\sigma_L r}{DB_1} \quad (6.19)$$

where $f_u(k)$ is the probability density function of a unit normal variable evaluated at k . (See Table II.1 in Appendix II.)

$v = \$2/\text{unit}$
 $\hat{x}_L = 50 \text{ units}$
 $\sigma_L = 21.0 \text{ units}$
 $r = 0.24 \text{ \$/\$/year}$
 $A = \$20/\text{order}$

Step 1

$$\frac{DB_1}{\sqrt{2\pi}Qv\sigma_L r} = \frac{200 \text{ units/year} \times \$300}{\sqrt{2\pi} \times 129 \text{ units} \times \$2/\text{unit} \times 21/\text{units} \times 0.24/\text{yr}}$$

$$= 18.4 > 1$$

Hence, from Equation 6.20

$$k = \sqrt{2 \ln(18.4)} = 2.41$$

Step 3 (Step 2 is bypassed in this example.)

$$s = 50 + 2.41(21) = 101 \text{ units}$$

The high value of B_1 has led to a rather large safety factor.

6.7.5.2 Discussion

Notice in Figure 6.9 the behavior of the carrying, shortage, and total costs as k is varied from 1.3 to 3.3 for the numerical example.* Now suppose that σ_L is reduced from 21 to 7 units due to a closer relationship with customers. The effects of changing this “given” are shown graphically in the same figure. (The shortage costs are the same for both values of σ_L , as are the ordering costs AD/Q .) For k values greater than 1.7, the entire total cost curve for $\sigma_L = 7$ is below the low point of the $\sigma_L = 21$ curve. In other words, reducing σ_L from 21 to 7, without much attempt to optimize over k , is much more beneficial than choosing the best value of k for σ_L fixed at 21.

We can see from Equation 6.20 that k decreases as σ_L or v goes up. Intuitively, the behavior of k with v makes sense under the assumed stockout-costing mechanism. If there is only a fixed cost per stockout occasion that is the same for all the items, then, it makes sense to allocate a greater proportional SS to the less-expensive items where an adequate level of protection is achieved with relatively little investment. Furthermore, as shown in Section 6A.1 of the Appendix of this chapter, it follows from Equation 6.20 that k decreases as Dv increases (this is not obvious from a quick look at the equations because both $\sigma_L v$ and Qv depend on Dv). What this means is that higher safety factors are provided to the slower-moving items.

The appearance of Q in the decision rule for finding the safety factor k indicates that a change in Q will affect the required value of the SS. Such effects of Q were ignored in the derivations of Chapter 4. For B items, this turns out to be a reasonable approximation, as we will see in Chapter 7, where such interactions between Q and k will be more closely examined in the context of control of A items.

* The total costs shown for both $\sigma_L = 7, 21$ include ordering cost AD/Q with $A = \$20$.

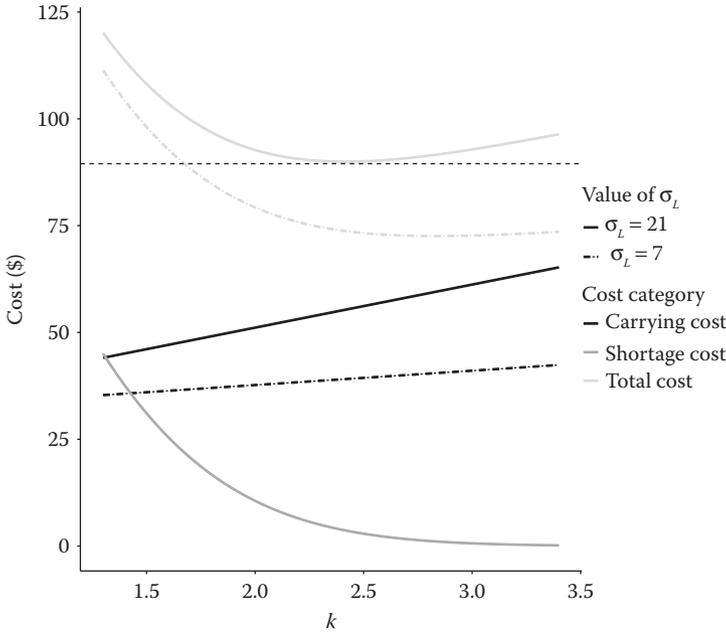


Figure 6.9 Effects of changing σ_L with B_1 costing (dashed lines for $\sigma_L = 7$).

There is no solution to Equation 6.20 when the condition of Equation 6.18 is satisfied. In such a situation, as shown in the derivation in Section 6A.1 of the Appendix, the best solution is the lowest allowable value of k . This situation may arise when B_1 is low relative to vr ; that is, when the shortage cost is low relative to the holding cost.

Note that in establishing the reorder point in Step 3, we advocate rounding to the nearest integer rather than always going to the next highest integer. This is because, in contrast with service measures, here, we are not bound by a service constraint.

6.7.6 Decision Rule for a Specified Fractional Charge (B_2) per Unit Short

Derivations of the following decision rule can be found in McClain and Thomas (1980). Again, we provide the cost function and then the decision rule. The total expected annual cost is

$$TC = AD/Q + \left(\frac{Q}{2} + k\sigma_L\right) vr + \frac{B_2 v \sigma_L G_u(k) D}{Q}$$

The shortage cost term is composed of the ESPRC, $\sigma_L G_u(k)$, the number of cycles per year, D/Q , and the cost per unit short, $B_2 v$.

Step 1 Is

$$\frac{Qr}{DB_2} > 1? \tag{6.21}$$

where Q has been predetermined and the units of the variables are such that the left-hand side of the equation is dimensionless (B_2 itself is dimensionless). If yes, then go to Step 2. If no, then continue with the following. Select k so as to satisfy

$$p_{u \geq}(k) = \frac{Qr}{DB_2} \quad (6.22)$$

If the use of Equation 6.22 gives a k value lower than the minimum allowable safety factor specified by the management, then go to Step 2. Otherwise, move to Step 3.

Step 2 Set k at its lowest allowable value.

Step 3 Reorder point, $s = \hat{x}_L + k\sigma_L$, rounded to the nearest integer (except raised to the next highest integer in the event that Step 2 has been used).

6.7.6.1 Numerical Illustration

A firm wishes to allocate a fixed amount of SS among a number of products to keep the total value of backorders per year as low as possible. Management feels that it is reasonable to use a B_2 value of 0.25; that is, each unit short incurs a cost equal to 25% of its unit value. The replenishment quantity of an item under consideration has been predetermined at 85 units. This value is determined from annual demand of 200, a fixed order cost of \$21.50, a carrying charge of 20%, and a unit cost of \$6. Other quantities of interest include:

$$\hat{x}_L = 50 \text{ units}$$

$$\sigma_L = 10 \text{ units}$$

Step 1

$$\frac{Qr}{DB_2} = \frac{85 \text{ units} \times 0.2/\text{year}}{200 \text{ units/year} \times 0.25} = 0.34 < 1$$

Then, Equation 6.22 gives

$$p_{u \geq}(k) = 0.34$$

From Table II.1 in Appendix II (or from the function NORMSINV(1-0.34) in Excel) $k \approx 0.41$ (presumably larger than the lowest allowable value, hence we go to Step 3).

Step 3

$$s = 50 + 0.41(10) = 54.1 \rightarrow 54 \text{ units}$$

The expected total cost of this policy is

$$TC = AD/Q + \left(\frac{Q}{2} + k\sigma_L \right) vr + \frac{B_2 v \sigma_L G_u(k) D}{Q} = \$51 + \$56 + \$8 = \$115$$

where $G_u(k) = 0.2270$ from Appendix II. Note that holding and ordering costs are not equal, unlike in Chapter 4. We can see from Figure 6.10 the form of the total cost as a function of k . With $B_2 = 0.25$, the decrease in total cost is small as k is increased from 0 to its optimal value ($k \approx 0.41$). Costs can increase sharply if k is set far too high. Now with $B_2 = 1.0$ shortages are much more expensive, and it is beneficial to hold more SS. The result

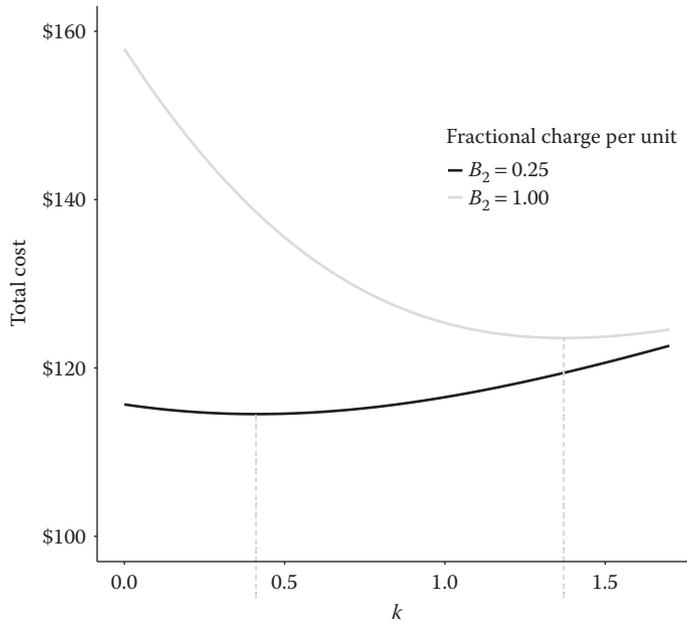


Figure 6.10 Sensitivity of total cost to the safety factor (with $B_2 = 0.25, 1.00$).

is that the decrease in cost as k increases from 0 is much more pronounced. With $B_2 = 1.0$, the optimal k is approximately 1.37.

6.7.6.2 Discussion

As seen in Equation 6.22, the safety factor k increases as B_2/r increases, as one would expect. Moreover, under the use of EOQs, we found that Q/D decreases as Dv increases. Therefore, at least under the use of EOQs, we can see from Equation 6.22 that k increases as Dv increases; that is, under this decision rule, larger safety factors are given to the faster-moving items, all other things being equal.

The $p_{u \geq}(k)$ value represents a probability; hence, there is no solution to Equation 6.22 when the right-hand side of the equation exceeds unity; that is, when Equation 6.21 is satisfied. It can be shown (in a manner paralleling the derivation in Section 6A.1 of the Appendix of this chapter) that, when this happens, the lowest permissible value of the safety factor should be used. Intuitively, the condition is satisfied if the shortage cost rate B_2 is smaller than the holding cost rate, r (for $Q = D$, to be precise). If $r = B_2$, the condition is satisfied if the order quantity is larger than the annual demand.

6.7.7 Decision Rule for a Specified Fractional Charge (B_3) per Unit Short per Unit Time

As mentioned in Section 6.5.4, there is an equivalence between the B_3 costing measure and the P_2 service measure. Thus, we merely present the decision rule here, leaving further comment to the P_2

section (the derivation involves differentiating the total relevant costs with respect to k and using the results in Hadley and Whitin 1963). See also Shore (1986).

Step 1 Select the safety factor k that satisfies

$$G_u(k) = \frac{Q}{\sigma_L} \left(\frac{r}{B_3 + r} \right) \quad (6.23)$$

where Q has been predetermined, presumably by one of the procedures of Chapter 4 (and must be expressed in the same units as σ_L), and B_3 and r are in \$/\$/unit time. Make sure that the k value is at least as large as the lowest allowable (management specified) value (e.g., zero) of the safety factor.

Step 2 Reorder point $s = \hat{x}_L + k\sigma_L$, rounded to the nearest integer (but increased to the next highest integer if the minimum k value is used in Step 1).

6.7.8 Decision Rule for a Specified Charge (B_4) per Customer Line Item Short

Assume that customers order a number of products from a given supplier. The supplier may be interested in minimizing the total cost when there is a charge per line item short. If \hat{z} is the average number of units ordered per customer line item, we can write the following equation for the approximate expected total relevant cost (due to Hausman 1969):

$$\text{ETRC} = AD/Q + \left(\frac{Q}{2} + k\sigma_L \right) vr + B_4 \frac{D\sigma_L G_u(k)}{Q\hat{z}} \quad (6.24)$$

Step 1 Select the safety factor k that satisfies

$$p_{u \geq}(k) = \frac{Qrv\hat{z}}{B_4D} \quad (6.25)$$

where Q has been predetermined, presumably by one of the procedures of Chapter 4 (and must be expressed in the same units as σ_L), and r is in \$/\$/unit time. Make sure that the k value is at least as large as the lowest allowable value (e.g., zero) of the safety factor.

Step 2 Reorder point $s = \hat{x}_L + k\sigma_L$, rounded to the nearest integer (but increased to the next highest integer if the minimum k value is used in Step 1).

6.7.9 Decision Rule for a Specified Probability (P_1) of No Stockout per Replenishment Cycle

Suppose management has specified that the probability of no stockout in a cycle should be no lower than P_1 (conversely, the probability of a stockout should be no higher than $1 - P_1$). Then, we have the following simple decision rule (whose derivation is shown in Section 6A.2 of the Appendix of this chapter).

Step 1 Select the safety factor k to satisfy

$$p_{u \geq}(k) = 1 - P_1 \quad (6.26)$$

where

$p_{u \geq}(k) = \Pr\{\text{Unit normal variable (mean 0, standard deviation 1) takes on a value of } k \text{ or larger}\}$, a widely tabulated function (see Table II.1 in Appendix II)

Step 2 Safety stock, $SS = k\sigma_L$.

Step 3 Reorder point, $s = \hat{x}_L + SS$, increased to the next higher integer (if not already exactly an integer).

6.7.9.1 Numerical Illustration

Suppose that the forecast system generates the following values:

$$\hat{x}_L = 58.3 \text{ units}$$

$$\sigma_L = 13.1 \text{ units}$$

Management also desires a service level of $P_1 = 0.90$. From Equation 6.26, we know that

$$p_{u \geq}(k) = 0.10$$

From Table II.1 in Appendix II (or from the function $\text{NORMSINV}(1 - p_{u \geq}(k))$ on Excel)

$$k \approx 1.28$$

$$\begin{aligned} SS &= k\sigma_L = 1.28 \times 13.1 \text{ units} \\ &= 16.8 \text{ units} \end{aligned}$$

Reorder point, $s = 58.3 + 16.8 = 75.1$, say 76 units.

6.7.9.2 Discussion of the Decision Rule

From Equation 6.26, we can see that the safety factor k depends only on the value of P_1 ; in particular, it is independent of any individual-item characteristics such as the order quantity Q . Therefore, all items for which we desire the same service level, P_1 , will have identical values of the safety factor k . Thus, we see an equivalence between two criteria, namely, using a specified value of k and a specified value of P_1 .

Because of the discrete nature of the reorder point, we are not likely to be able to provide the exact level of service desired since this usually would require a noninteger value of s . Therefore, a noninteger value of s found in Equation 6.26 is rounded up to the next higher integer, with the predicted service level then being slightly higher than required. Such differences are often less pronounced than the error in the input data.

The fact that the safety factor does not depend on any individual-item characteristics may necessitate reexamining the meaning of service here. Recall that service under the measure used in this section is prob no stockout per replenishment cycle. Consider two items, the first being replenished 20 times a year, the other once a year. If both of them are given the same safety factor so that both have a probability of 0.10 of stockout per replenishment cycle, then, we would expect $20 \times (0.10)$ or 2 stockouts per year for the first item and only 1 stockout every 10 years (0.1 per year) for the second item. Therefore, depending on management's definition of service, we, in fact,

may not be giving the same service on these two items. Rules based on other service measures will now be presented.

6.7.10 Decision Rule for a Specified Fraction (P_2) of Demand Satisfied Directly from Shelf

The following decision rule is derived in Section 6A.2 of the Appendix to this chapter.

Step 1 Select the safety factor k that satisfies*

$$G_u(k) = \frac{Q}{\sigma_L}(1 - P_2) \quad (6.28)$$

where Q has been predetermined, presumably by one of the procedures of Chapter 4 (and must be expressed in the same units as σ_L). The other relevant variables are defined in Section 6.7.1.

Make sure that the k value is at least as large as the lowest allowable value of the safety factor.

Step 2 Reorder point, $s = \hat{x}_L + k\sigma_L$, increased to the next higher integer (if not already exactly an integer).

A brief comment on the order quantity is in order. It is common practice to use the EOQ for the order quantity in an (s, Q) system. Although this practice is often very close to the optimal, Platt et al. (1997) have shown that a simple modification of the order quantity outperforms the EOQ when sequentially choosing the order quantity and reorder point. They recommend an order quantity based on the limit as the ratio $EOQ/\sigma \rightarrow \infty$, when using a fill rate as the performance measure:

$$Q = \frac{1}{P_2} \sqrt{\frac{2AD}{rv} + \sigma_L^2}$$

Equation 6.28 underestimates the true fill rate if σ_L is large relative to Q , because Equation 6.28 double counts backorders from a previous cycle that are not met at the start of the next cycle. A more accurate formula for small values of Q/σ_L is

$$G_u(k) - G_u(k + Q/\sigma_L) = \frac{Q}{\sigma_L}(1 - P_2) \quad (6.29)$$

For further discussion of the fill rate, see Silver (1970), Johnson et al. (1995), Zipkin (1986), Platt et al. (1997), and Disney et al. (2015).

* Equation 6.28 applies for the case of complete backordering. The only difference for the case of complete lost sales is that $1 - P_2$ is replaced by $(1 - P_2)/P_2$:

$$G_u(k) = \frac{Q}{\sigma_L} \left(\frac{1 - P_2}{P_2} \right) \quad (6.27)$$

6.7.10.1 Numerical Illustration

Consider a particular type of developing liquid distributed by a manufacturer of laboratory supplies and equipment. Management has specified that 99% of demand is to be satisfied without backordering. A replenishment quantity of 200 gallons has been predetermined and the forecast system provides us with $\hat{x}_L = 50$ gallons and $\sigma_L = 11.4$ gallons.

Step 1 Equation 6.28 yields

$$G_u(k) = \frac{200}{11.4}(1 - 0.99) = 0.175$$

and Table II.1 in Appendix II gives

$$k = 0.58$$

Step 2

$$s = 50 + 0.58(11.4) = 56.6 \rightarrow 57 \text{ gallons}$$

In this case, Equation 6.29 gives the same result as Equation 6.28. For the case of complete lost sales (rather than backordering), the use of Equation 6.27 would lead us to the same value of the reorder point. Recall that $G_u(k)$ can be obtained using a rational approximation presented in Appendix III at the end of the book. This is especially appropriate for spreadsheet use, and is extremely accurate.

6.7.10.2 Discussion

Intuitively, we would expect that the required SS would increase if (1) Q decreased (more opportunities for stockouts), (2) σ_L increased (higher uncertainty of forecasts), or (3) P_2 increased (better service desired). If any of these changes take place, $G_u(k)$ decreases, but as seen from the table in Appendix II, a decrease in $G_u(k)$ implies an increase in k , that is, exactly the desired behavior. These relationships can be seen in Figure 6.11.

In addition, on average, σ_L tends to increase with D ; therefore, the increase in k with increasing σ_L says that, on the average under this decision rule, faster-moving items get higher safety factors than do the slower-moving items.

If the right-hand side of Equation 6.28 is large enough, a negative value of k is required to produce the equality; to give a service level as poor as P_2 , one must deliberately plan to be short on the average when a replenishment arrives. In particular, if Q is large, many customers have their demand satisfied before there is even an opportunity for a stockout. When a replenishment order is placed, so many demands have been met from the shelf that a stockout will only serve to lower P_2 to the target. Management may find this intolerable and, instead, set a lower limit on k (e.g., zero). If k is set at zero when Equation 6.28 calls for a negative value of k , the service provided will be better than P_2 .

The P_2 value is usually quite close to unity. Therefore, as evidenced by the numerical example, Equations 6.27 and 6.28 normally give very similar values of $G_u(k)$, that is, of k itself. In other words, the value of the safety factor is little influenced by whether the model assumes complete backordering or complete lost sales (or any mix of these two extremes).

Finally, a comparison of Equations 6.23 and 6.28 shows the analogy between the P_2 and B_3 measures mentioned earlier.

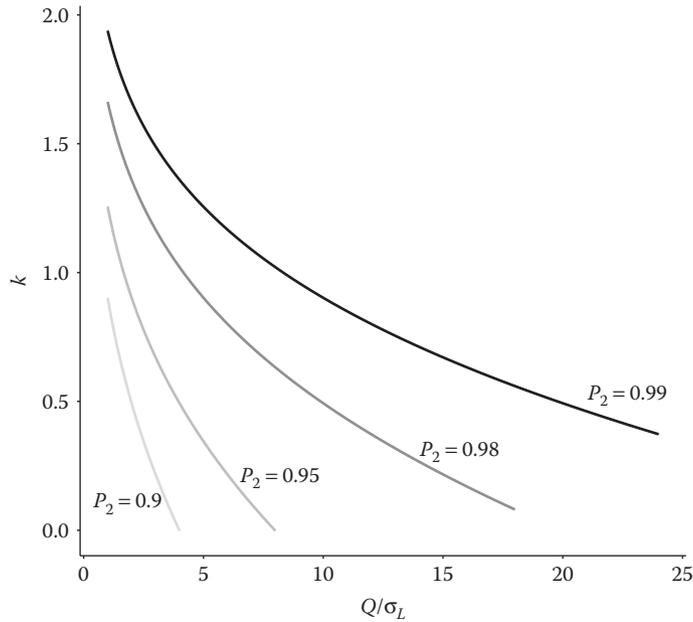


Figure 6.11 Sensitivity of the P_2 service measure.

6.7.11 Decision Rule for a Specified Average Time (TBS) between Stockout Occasions

Step 1 Is

$$\frac{Q}{D(\text{TBS})} > 1? \quad (6.30)$$

where Q has been predetermined (presumably by the methods of Chapter 4) and the units of the variables are such that the left side is dimensionless. If yes, then go to Step 2. If no, then continue with the following. Select the safety factor k to satisfy

$$p_{u \geq}(k) = \frac{Q}{D(\text{TBS})} \quad (6.31)$$

If the resulting k is lower than the minimum allowable value specified by the management, then go to Step 2. Otherwise, move to Step 3.

Step 2 Set k at its lowest permitted value.

Step 3 Reorder point, $s = \hat{x}_L + k\sigma_L$ raised to the next highest integer (if not already exactly an integer).

6.7.11.1 Numerical Illustration

For the item used in Section 6.7.9, suppose that

$$\hat{x}_L = 58.3 \text{ units}$$

$$\sigma_L = 13.1 \text{ units}$$

$$D = 200 \text{ units/year}$$

$$Q = 30 \text{ units}$$

and the management specifies a desired TBS value of 2 years. Then, Equation 6.31 gives

$$p_{u \geq}(k) = \frac{300}{200(2)} = 0.075$$

From Table II.1 of Appendix II, we have

$$k = 1.44$$

Then,

$$SS = 1.44 \times 13.1 = 18.9 \text{ units}$$

and,

$$s = 58.3 + 18.9 = 77.2, \text{ say } 78 \text{ units}$$

6.7.11.2 Discussion

A comparison of Equations 6.22 and 6.31 reveals that there is an equivalence between using the B_2 costing method and using the TBS service measure. Specifically, we have, for equivalence, that

$$\text{TBS} = \frac{B_2}{r}$$

From Equation 6.31, we see that increasing TBS leads to a lower value of $p_{u \geq}(k)$, and hence a larger value of the safety factor k , which is certainly an appealing behavior.

6.7.12 Decision Rule for the Allocation of a TSS to Minimize the ETSOPY

As shown in Section 6A.4 of the Appendix of this chapter, allocation of a total safety stock (TSS) among a group of items to minimize the ETSOPY leads to a decision rule identical to that under the B_1 shortage-costing method. Thus, simply use the procedure of Section 6.7.5, adjusting the value of B_1 until the total implied SS is equal to the given quantity to be allocated. (Conversely, when we discuss SS exchange curves later in this chapter, we will see that the selection of an aggregate operating point on an appropriate exchange curve will imply a value of B_1 to be used). The total number of stockout occasions per year is simply the product of the probability of a stockout in a replenishment cycle multiplied by the number of cycles per year:

$$p_{u \geq}(k) \left(\frac{D}{Q} \right)$$

6.7.13 Decision Rule for the Allocation of a TSS to Minimize the ETVSPY

In a fashion similar to that shown in Section 6A.4 of the Appendix of this chapter, it can be shown that there is an equivalence between the decision rules for three different criteria:

1. Allocating a TSS among a group of items to minimize the expected total value of shortages per year (ETVSPY)

2. Use of a B_2 shortage cost
3. Use of the TBS service measure

Thus, for the allocation criterion, one needs to only use the procedure of Section 6.7.6, adjusting the value of B_2 until the total implied SS is equal to the given quantity to be allocated. The total value of shortages per year is the product of the number of shortages per replenishment cycle, multiplied by the value of those shortages, multiplied by the number of cycles per year:

$$\sigma_L G_u(k)v \left(\frac{D}{Q} \right)$$

6.7.14 Nonnormal Lead Time Demand Distributions

As mentioned earlier, for pragmatic reasons, we have focused on the case of normally distributed demands (or forecast errors) in this chapter. The normal distribution often provides a good empirical fit to the observed data: it is convenient from an analytic standpoint, it is widely tabulated and is now built into spreadsheets, and, finally, the impact of using other distributions is usually quite small (see, e.g., Naddor 1978; Fortuin 1980; Lau and Zaki 1982; Tyworth and O'Neill 1997), particularly when recognizing the other inaccuracies present (estimates of the parameters of the distribution, estimates of cost factors, and so on). Tyworth and O'Neill (1997) and Lau and Zaki (1982) show that, while the errors in SS may be high when using the normal, the errors in total cost are quite low. If the ratio σ_L/\hat{x}_L is greater than 0.5, it may be desirable to use another distribution for lead time demand, such as the Gamma. (When this ratio equals 0.5, the probability of negative values of demand for a normal distribution is 2.3%; but the real issue in choosing the reorder point is the right tail of the distribution.) Therefore, we often use the following rule of thumb: If the ratio σ_L/\hat{x}_L is greater than 0.5, consider a distribution other than the normal. However, so long as this ratio is less than 0.5, the normal is probably an adequate approximation. Finally, if daily demand is not normal, the normal will generally be a good approximation for lead time demand if the lead time is at least several days in duration. (If the lead time itself follows some probability distribution, consult Section 6.10.) As discussed in Chapter 3, it is advisable to at least test the normal distribution using data from a sample of items.

We now comment specifically on three other distributions—namely, the Gamma, the Laplace, and the Poisson. As discussed by Burgin (1975), the Gamma distribution has considerable intuitive appeal for representing the distribution of lead time demand. In particular, if the demand distribution is skewed to the right, or if the ratio σ_L/\hat{x}_L is greater than 0.5, use of the Gamma should be considered. The distribution is not as tractable as the normal; thus, considerable effort has been devoted to developing approximations and tables (see Burgin and Norman 1976; Das 1976; Taylor and Oke 1976, 1979; Johnston 1980; Van der Veen 1981; Tyworth et al. 1996). We derive an expression for expected units short per cycle in the Appendix to this chapter, and we present a simple spreadsheet approach for finding reorder points using the Gamma distribution in Appendix III.

The Laplace or pseudoexponential distribution was first proposed by Presutti and Trepp (1970). It is analytically very simple to use and may be an appropriate choice for slow-moving items (see Ng and Lam 1998).

The Poisson distribution is also a candidate for slower-moving items. In fact, in Chapter 7, we will illustrate its use for expensive, low-usage items. Many other distributions of lead time

demand have been proposed in the literature. We now simply provide a listing of several of them accompanied by suitable references:

1. Exponential: Brown (1977)
2. Geometric: Carlson (1982)
3. Logistic: Fortuin (1980), Van Beek (1978)
4. Negative binomial: Ehrhardt (1979), Agrawal and Smith (1996)
5. Pearson: Kottas and Lau (1980)
6. Tukey's Lambda: Silver (1977)
7. Split-normal (for asymmetric lead time demand): Lefrancois (1989)

In addition, there has been a significant amount of research on approaches that can model a number of distributions. Zipkin (1988) describes the use of phase-type distributions that can approximate many other distributions. Strijbosch and Heuts (1992) provide nonparametric approximations. See also Sahin and Sinha (1987), Shore (1995), and Kumaran and Achary (1996). A powerful method of fitting many distributions using mixtures of Erlang distributions is presented by Tijms (1986). Much work has been done on distribution-free approaches. The original work by Scarf (1958) proposed finding the worst-possible distribution for each decision variable and then finding the optimal inventory policy for that distribution. Thus, it is a conservative approach. It requires only that the mean and variance of the lead time demand are known. Bulinskaya (1990), Gallego (1992), Gallego and Moon (1993), and Moon and Choi (1994, 1995) have advanced the research, and have shown that, in some cases, applying a distributional form incorrectly can lead to large errors. The distribution-free approach could generate significant savings. If the distribution appears to be very different from a known form, one should consider the distribution-free formulas, or those described in the next paragraph. Silver and Rahnama (1986) find the cost penalty for using the mean and standard deviation from a statistical sample rather than the true values. Silver and Rahnama (1987) show that the safety factor, k , should be biased upward when using a statistical sample.

Finally, we briefly present a procedure that appears to be quite useful for a number of different distributions. Lau and Lau (1993) observe that the optimal order quantities in (s, Q) systems are very similar regardless of the distribution of demand. Likewise, the cumulative distribution function of the lead time demand distributions evaluated at the reorder point s , or $F_x(s) = \int_0^s f_x(x_0) dx_0$, is quite close. Building on these insights, they develop a simple approach for the case of B_2 shortage costing, relatively few stockout occasions, and $Q > s$. They employ the uniform distribution, which is easy to evaluate, for use in finding the order quantity. Then, they use the actual lead time demand distribution for finding the reorder point.

Let $g_2 = 2DB_2/r$ and find

$$Q = \sqrt{\frac{\text{EOQ}^2(g_2)}{g_2(g_2 - 4\sigma_L\sqrt{3})}}$$

and

$$F_x(s) = 1 - 2\sqrt{\frac{\text{EOQ}^2}{g_2(g_2 - 4\sigma_L\sqrt{3})}}$$

Finally, use the actual lead time demand distribution to find s from $F_x(s)$. See Problem 6.26 for an example. Lau and Lau demonstrate that the cost error of using this approach rather than the exact approach is nearly always less than 1%. See also Shore (1995) for a similar approach.

6.8 Implied Costs and Performance Measures

When choosing a cost minimization or a service objective approach, it is important to realize that the obtained inventory policy necessarily implies values for alternate cost or service measures. These should be reported to the management if there is a possibility that the alternate measures are of interest. An example will be helpful.

Assume that the fill rate, $P_2 = 0.98$, is chosen as the service objective, and that the item in question has the following characteristics:

$$D = 4000 \text{ units/year}$$

$$A = \$20.25$$

$$r = 0.30 \text{ \$/\$/year}$$

$$v = \$6/\text{unit}$$

$$L = 1 \text{ week}$$

$$\hat{x}_L = 80 \text{ units}$$

$$\sigma_L = 20 \text{ units}$$

We can calculate the EOQ to be 300 units, and using Equation 6.28 we find that $G_u(k) = 0.30$. Therefore, $k = 0.22$ and $s = 84.4$, or rounded to $s = 85$. Now, it is possible that managers are also interested in the probability of a stockout in a replenishment cycle. For this (s, Q) policy of $(85, 300)$, we know that $k = (85 - 80)/20 = 0.25$. Note that the actual k is greater than 0.22 because we rounded s up to 85. Thus, $p_{u \geq}(k) = 0.4013$ (from Table II.1 in the Appendix II). So $P_1 = 1 - 0.4013 = 0.5987$. In other words, a fill rate of 98% yields a cycle service level of only 59.87%. In this case, the relatively large order quantity implies that many customers have their demands met from the shelf before a lead time ever begins. Thus, frequent stockouts at the end of the replenishment cycle do not significantly lower the fill rate. They do, on the other hand, imply a relatively low cycle service level. See Snyder (1980). This may be acceptable to management; but it may also be problematic, and the inventory policy should be adjusted.

A similar approach can be used to determine the implied values for costs, such as B_2 . In the preceding example, the fact that $p_{u \geq}(k) = 0.4013$ implies, using Equation 6.22, that $B_2 = 0.056$ —or the cost of a unit short is 5.6% of the unit value. Again, this value may be consistent with management's perspective, or it may seem to be completely in error. If the latter is true, it would be wise to consider other inventory policies that are more consistent.

In general, we can see that for any input performance objective (B_1, B_2, P_1 , and so on), we will obtain a value for k , which will be used to find the reorder point, s . The value of k can then be used to find the implied values of any of the other input performance objectives.

6.9 Decision Rules for Periodic-Review, Order-Up-to-Level (R, S) Control Systems

Recall that in an (R, S) system every R units of time, a replenishment order is placed of sufficient magnitude to raise the inventory position to the order-up-to-level S . Fortunately, for such systems, there will be no need to repeat all the detail that was necessary for (s, Q) systems in the previous section, because there is a rather remarkable, simple analogy between (R, S) and (s, Q) systems. Specifically, the (R, S) situation is exactly equivalent to the (s, Q) situation if one makes the following transformations:

(s, Q)	(R, S)
s	S
Q	DR
L	$R + L$

It therefore follows that the decision rule for determining the S value in an (R, S) system for a particular selection of shortage cost or service measure is obtained from the corresponding rule for determining s in an (s, Q) system by simply making the above three substitutions. Prior to proving this equivalence, we will discuss the choice of R as well as the key time period involved in the selection of S .

6.9.1 The Review Interval (R)

In an (R, S) control system, a replenishment order is placed every R units of time; and when computing the value of S , we assume that a value of R has been predetermined. The determination of R is equivalent to the determination of an EOQ expressed as a time supply, except for two minor variations. First, the cost of reviewing the inventory status must be included as part of the fixed setup cost A . Second, it is clear that it would be senseless to attempt to implement certain review intervals, for example, 2.36 days; that is, R is obviously restricted to a reasonably small number of feasible discrete values. In fact, as mentioned in Section 6.3, the value of R is often dictated by external factors, such as the frequency of truck deliveries.

6.9.2 The Order-Up-to-Level (S)

The key time period over which protection is required is now of duration $R + L$, instead of just a replenishment lead time L . This is illustrated in Figure 6.12 and Table 6.7 with an example where $S = 50$ units and where two consecutive orders (called X and Y) are placed at times t_0 and $t_0 + R$, respectively, and arrive at $t_0 + L$ and $t_0 + R + L$, respectively. In selecting the order-up-to-level at time t_0 we must recognize that, once order X has been placed, no other later orders (in particular Y) can be received until time $t_0 + R + L$. Therefore, the order-up-to level at time t_0 must be sufficient to cover demand through a period of duration $R + L$. A stockout could occur in the early portion of the period (prior to $t_0 + L$) but that would be a consequence of the setting of the order-up-to-level on the order preceding order X . What is of interest to us is that a stockout will occur toward the end of the period (after time $t_0 + L$) if the total demand in an interval of length $R + L$ exceeds the order-up-to-level S . Another way of putting this is that any stockouts up to time $t_0 + L$ will not be influenced by our ordering decision at time t_0 ; however, this decision certainly influences the likelihood of a stockout at time $t_0 + R + L$. Of course, this does not mean that we would necessarily take no action if at time t_0 a stockout appeared likely by time $t_0 + L$. Probably, one would expedite any already-outstanding order or try to get the current order (that placed at time t_0) delivered as quickly as possible. Note, however, that this still does not influence the size of the current order.

We should note that is often the case, in our experience, that an impending stockout generates an overreaction on the part of the person responsible for purchasing. The current order size is inflated, causing excess inventory when the order arrives. If the impending stockout is due to a real

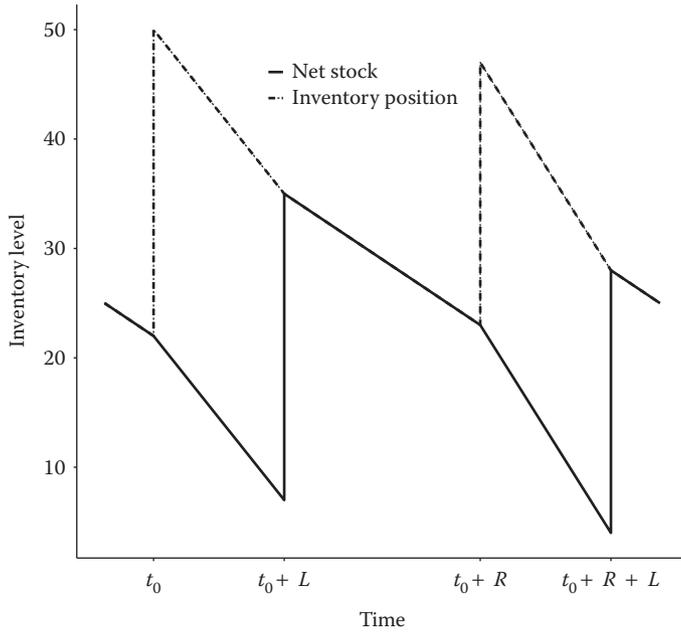


Figure 6.12 The time period of protection in an (R, S) system.

Table 6.7 Illustration of Why (R + L) Is the Crucial Time Interval in an (R, S) System (Illustrated for S = 50)

Event	Time or Interval ^a	Demand	(OH - BO) ^b Net Stock	On Order	Inventory Position
Place order for 30	$t_0 - \epsilon$ $t_0 - \epsilon$ to $t_0 + \epsilon$	-	20 20	- 30	20 50 ^(a)
Order of 30 arrives	$t_0 + \epsilon$ to $t_0 + L - \epsilon$ $t_0 + L - \epsilon$ to $t_0 + L + \epsilon$	15 ^(b) -	5 35	30 -	35 35
Place order for 27	$t_0 + L + \epsilon$ to $t_0 + R - \epsilon$ $t_0 + R - \epsilon$ to $t_0 + R + \epsilon$	12 ^(c) -	23 23	- 27	23 50
Order of 27 arrives	$t_0 + R + \epsilon$ to $t_0 + R + L - \epsilon$ $t_0 + R + L - \epsilon$ to $t_0 + R + L + \epsilon$	19 ^(d) -	4 ^(e) 31	27 -	31 31

$$(e) = (a) - \underbrace{[(b) + (c) + (d)]}_{\text{Total demand in } t_0 \text{ to } t_0 + R + L}$$

^a ϵ is a shorthand notation for a very small interval of time. Thus $t_0 - \epsilon$ represents a moment just prior to time t_0 , while $t_0 + \epsilon$ represents a moment just after time t_0 .

^b (On hand) - (Backorders).

increase in demand, the value of S should be recomputed. Otherwise, the manager should resist the temptation to inflate inventories.

As mentioned above, a stockout will occur at the end of the current cycle (i.e., at time $t_0 + R + L$) if the total demand in an interval of duration $R + L$ exceeds S . A little reflection shows a very close analogy with the determination of s in the continuous-review system. In fact, the only difference is the duration of the time period of concern, namely, $R + L$ instead of just L .

6.9.3 Common Assumptions and Notation

The assumptions are very similar to those listed in Section 6.7.1, with suitable replacements of $R + L$ for L , DR for Q , and S for s . We note only two additional assumptions:

1. There is a negligible chance of no demand between reviews; consequently, a replenishment order is placed at every review.
2. The value of R is assumed to be predetermined. In most situations, the effects of the two decision variables, R and S , are not independent, that is, the best value of R depends on the S value, and vice versa. However, as will be shown in Chapter 7, it is quite reasonable for practical purposes when dealing with B items to assume that R has been predetermined without knowledge of the S value.

Unlike in the (s, Q) situation, the assumption of unit-sized demand transactions is not needed here.

We mention only notation that has not been used extensively thus far in this chapter:

R = prespecified review interval, expressed in years

S = order-up-to-level, in units

\hat{x}_{R+L} = forecast (or expected) demand over a review interval plus a replenishment lead time, in units

σ_{R+L} = standard deviation of errors of forecasts over a review interval plus a replenishment lead time, in units

6.9.4 Common Derivation

We abbreviate this section because it closely parallels the development of (s, Q) systems. Because of Assumption 1, we have

$$(\text{Number of reviews per year}) = 1/R$$

and

$$(\text{Number of replenishment orders placed per year}) = 1/R \quad (6.32)$$

We now present the relevant equations for SS and stockouts. (Reference to Figure 6.12 will be helpful.) Recall that two orders, X and Y , are considered. If the demand (x) in $R + L$ has a probability density function $f_x(x_0)$ defined such that $f_x(x_0)dx_0 = \Pr \{\text{Total demand in } R + L \text{ lies}$

between x_0 and $x_0 + dx_0$ then the reasoning of Section 6.9.2 leads to the following three important results:

1. Safety stock = E (Net stock just before order Y arrives)

$$= \int_0^{\infty} (S - x_0) f_x(x_0) dx_0$$

that is,

$$SS = S - \hat{x}_{R+L} \tag{6.33}$$

2. Prob {Stockout in a replenishment cycle}

$$\Pr\{x \geq S\} = \int_S^{\infty} f_x(x_0) dx_0 \tag{6.34}$$

the probability that the total demand during a review interval plus lead time is at least as large as the order-up-to-level.

3. Expected shortage per replenishment cycle, ESPRC

$$= \int_S^{\infty} (x_0 - S) f_x(x_0) dx_0 \tag{6.35}$$

On the average, the on-hand level is

$$E(OH) \approx S - \hat{x}_{R+L} + DR/2 \tag{6.36}$$

It is convenient to again set

$$SS = k\sigma_{R+L} \tag{6.37}$$

If forecast errors are normally distributed, the results of Appendix II give that Equations 6.34 and 6.35 reduce to

$$\Pr\{\text{Stockout in a replenishment cycle}\} = p_{u \geq}(k) \tag{6.38}$$

and

$$\text{ESPRC} = \sigma_{R+L} G_u(k) \tag{6.39}$$

A comparison of pairs of equations follows:

<i>Compare</i>	<i>With</i>
6.33	6.9
6.39	6.13
6.37	6.14
6.32	6.15
6.38	6.16
6.39	6.17

The comparison reveals the validity of our earlier assertion that the (R, S) situation is exactly equivalent to the (s, Q) situation if one makes the transformations listed at the beginning of this section.

6.10 Variability in the Replenishment Lead Time Itself

In all the decision systems discussed in this chapter, the key variable in setting a reorder point or order-up-to-level is the total demand in an interval of length $R + L$ (which reduces to just L for the case of continuous review). So far, our decisions have been based on a known replenishment lead time L with the only uncertain quantity being the demand rate during L or $R + L$. If L itself is not known with certainty, it is apparent that increased SS is required to protect against this additional uncertainty.*

It should be noted that where the pattern of variability is known, for example, seasonally varying lead times, there is no additional problem because the lead time at any given calendar time is known and the SS (and reorder point) can be adjusted accordingly. Where lead times are increasing in a known fashion, such as in conditions of reduced availability of raw materials, again, SSs (and reorder points) should be appropriately adjusted. We discuss these issues in more detail in Chapter 7.

Every reasonable effort should be made to eliminate variability in the lead time, and one can use the models developed in this section to quantify the value of reduction in lead time variability. A key component of the lead time in which uncertainty exists is the shipping time from the supplier to the stocking point under consideration. Choices among transportation modes can affect the variability, as well as the average duration, of the lead time (see Herron 1983, and Problem 6.10). Ordering process and information technology improvements can also reduce the average lead time and lead time variability. Blackburn (2012) and De Treville et al. (2014) provide recent discussions of the potential value of reductions in lead time and lead time variability.

* When inventory OH just after the arrival of a replenishment order may not be positive, de Kok (1990) has shown that the transformation of Equation 6.29 is not straightforward. In fact, the more accurate equation for P_2 is

$$P_2 = 1 - \frac{\sigma_{R+L} G_u(k) - \sigma_L G_u\left(\frac{\sigma_{R+L}}{\sigma_L} k + \frac{\hat{x}_R}{\sigma_L}\right)}{\hat{x}_R} \quad (6.40)$$

Lead time improvements, of course, require close cooperation with suppliers. In return for firm commitments well ahead of time, a reasonable supplier should be prepared to promise a more dependable lead time. The projection of firm commitments well into the future is not an easy task. We return to this point in Chapter 15 in the context of Material Requirements Planning, which offers a natural means of making such projections. Firms also have been increasingly employing strategic alliances—long-term relationships with close partners—to facilitate the required cooperation.

Where there is still some residual variability in the lead time that cannot be eliminated, two basic courses of action are possible:

1. Measure the distribution of the lead time (or $R + L$), and the distribution of demand per period, separately. Then combine the two to obtain the distribution of total demand over the lead time.
2. Try to ascertain the distribution of total demand over the full lead time by measuring or estimating actual demand over that period.

In each case, there are at least two ways to use the information. First, assume (or fit) a specific functional form—such as the Normal—to the distribution of total demand over lead time. Or, second, use the exact distribution to set an inventory policy.

6.10.1 Approach 1: Use of the Total Demand over the Full Lead Time

We begin by discussing a simple, but effective, method that is useful when it is possible to measure the total demand over the lead time. This procedure, due to Lordahl and Bookbinder (1994), uses these measurements directly and is based on the distribution-free properties of order statistics. The objective is to find the reorder point, s , when the parameters, and form, of the lead time demand distribution are unknown and P_1 is the desired performance measure. The procedure requires only that the user records, over time, the actual demand during lead time. These steps describe the procedure:

1. Rank order the observed lead time demand from the least to the greatest. So $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ if there are n observations of lead time demand.
2. Let $(n + 1)P_1 = y + w$, where $0 \leq w < 1$, and y is an integer.
3. If $(n + 1)P_1 > n$, set $s = x_{(y)}$. Otherwise, set $s = (1 - w)x_{(y)} + wx_{(y+1)}$, a weighted average of two of the observations. In effect, the reorder point is estimated from two of the larger values in the sample.

6.10.1.1 Numerical Illustration

A new entrepreneurial firm is trying to set reorder points for several items that are important to its business. Because lead times are roughly 2–3 weeks, and order sizes are not too large, the manager has gathered a number of observations of lead time demand. These are 15, 18, 12, 22, 28, 25, 22, 23, 11, and 19. The manager has set $P_1 = 0.85$.

Step 1 Rank order the data. 11, 12, 15, 18, 19, 22, 22, 23, 25, 28.

Step 2 $n = 10$

$P_1 = 0.85$

$$(n + 1)P_1 = (11)(0.85) = 9.35.$$

Thus $y + w = 9.35$.

So $y = 9$ and $w = 0.35$.

Step 3 $(n + 1)P_1 < n$, so

$$s = (1 - w)x_{(y)} + wx_{(y+1)}$$

$x_{(y)}$ = the 9th observation = 25

$x_{(y+1)}$ = the 10th observation = 28

$$s = (1 - 0.35)25 + 0.35(28) = 26.05 \text{ or } 26.$$

Lordahl and Bookbinder (1994) show that this procedure is better than using the normal distribution in many cases. In particular, if the firm must achieve the targeted service level, this approach works well.

Karmarkar (1994) provides a similar methodology in which the general shape of the demand distribution is ignored because it is irrelevant to the SS decision. Rather, attention is focused on the right tail of the distribution. His method uses a mixture of two distributions, with the right tail modeled by a translated exponential. The method works quite well in simulation experiments.

The second way to apply Approach 1 is to measure the mean and variance of the total demand over the lead time, and apply a specific functional form. Then, if the total demand appears to fit a normal distribution, the methods from this chapter can be used to find the appropriate reorder point.

6.10.2 Approach 2: Use of the Distribution of Demand Rate per Unit Time Combined with the Lead Time Distribution

The second approach assumes that the lead time (L) and the demand (D) in each unit time period are independent random variables, and requires measurements or estimates of each. The assumption of independence is probably a reasonable approximation to reality despite the fact that, in some cases, high demand is likely to be associated with long lead times (i.e., positive correlation), because of the heavy workload placed on the supplier.* Likewise, low demand can be associated with long lead times (i.e., negative correlation) because the supplier has to wait longer to accumulate sufficient orders for his or her desired run size. If L and D are assumed to be independent random variables, then

$$E(x) = E(L)E(D) \quad (6.41)$$

and

$$\sigma_x = \sqrt{E(L)\text{var}(D) + [E(D)]^2\text{var}(L)} \quad (6.42)$$

where x , with mean $E(x)$ and standard deviation σ_x , is the total demand in a replenishment lead time, in units; L , with mean $E(L)$ and variance $\text{var}(L)$, is the length of a lead time (L is the number of unit time periods, i.e., just a dimensionless number); and D with mean $E(D)$ and variance $\text{var}(D)$ is the demand, in units, in a unit time period. Derivations for Equations 6.41 and 6.42 are in Section 6A.5 of the Appendix.

Where variability in L exists, the $E(x)$ and σ_x quantities found from Equations 6.41 and 6.42 should be used in place of \hat{x}_L and σ_L in all the decision rules of this chapter.

* Wang et al. (2010) analyze the case where demand per period and lead times are correlated.

6.10.2.1 Numerical Illustration

We first illustrate the use of this approach when the combined distribution is (or is assumed to be) normally distributed. Let us consider two situations:

Situation 1: No Uncertainty in the Lead Time Itself (i.e., $\text{var}(L) = 0$ in Equation 6.42)

Suppose the unit time period is 1 week with $E(L) = 4$ and $E(D) = 100$ units and $\text{var}(D) = 300$ (units)². Then Equations 6.41 and 6.42 give

$$E(x) = 400 \text{ units}$$

and

$$\sigma_x = 34.64 \text{ units}$$

Then, to provide a desired probability (P_1) of no stockout per replenishment cycle equal to 0.95, we require a reorder point of $s = E(x) + k\sigma_x = 400 + 1.64(34.64) = 456.81$, say, 457 units.

Situation 2: Uncertainty in the Lead Time

Suppose that $E(L)$, $E(D)$, and $\text{var}(D)$ are as above, so that $E(x)$, the expected total demand over a lead time, is still 400 units. However, now we assume that L itself is random, in particular

$$\text{var}(L) = 1.44$$

(The standard deviation of L is $\sqrt{\text{var}(L)}$, or 1.2.) Then, Equation 6.42 gives

$$\sigma_x = \sqrt{(4)(300) + (100)^2(1.44)} = 124.90 \text{ units}$$

Now we find that the reorder point must be 605 units. In other words, the uncertainty in L has increased the reorder point by 148 units. Knowing the values for v and r would permit us to calculate the additional carrying costs implied by the extra 148 units. These extra costs would provide a bound on how much it would be worth to reduce or eliminate the variability in L .

6.10.3 Nonnormal Distributions

The expressions for mean and standard deviation of demand during lead time in Equations 6.41 and 6.42 do not necessarily assume a Normal distribution. So, these values could be used in conjunction with another distribution, such as the Gamma distribution (e.g., Tyworth et al. 1996; Moors and Strijbosch 2002; Leven and Segerstedt 2004). Such an approach may be justified when the demand during lead time is positively skewed or the ratio of $\sqrt{\text{var}(L)}/E(L)$ is large*. Use of the Gamma distribution is a good alternative in these cases. In Appendix III, we present the details of how one can implement the use of the Gamma distribution in a spreadsheet.

When combining the demand and lead time distribution using the approach in the previous section, it is possible to draw on some compelling research. Eppen and Martin (1988) show by an example that if the lead time takes on just a few values, say 2 or 4 weeks with equal probability, the use of the normal distribution provides very poor and unusual results. In fact, the reorder point in their example actually increases when the lead time decreases. (See Problem 6.27.)

* Note that if this ratio is much greater than 0.4, there is a significant chance of negative demand values.

Other research on the problem of variable lead time includes Burgin (1972), McFadden (1972), Magson (1979), Ray (1981), Mayer (1984), Bagchi and Hayya (1984), Bagchi et al. (1984, 1986), Zipkin (1988), Bookbinder and Lordahl (1989), Tyworth (1992), Chen and Zheng (1992), Federgruen and Zheng (1992), Wang and Rao (1992), Strijbosch and Heuts (1994), Heuts and de Klein (1995), and Keaton (1995). Most of this research assumes that the lead time and demand distributions are given (even if not known). Wu and Tsai (2001) and Chu et al. (2004) investigate the more complicated but more flexible approach of using mixtures of distributions for modeling lead times.

It is often possible, however, to “change the givens,” and some work on this problem has been done. Das (1975) shows that the total relevant costs are more sensitive to changes in the expected lead time if $\sqrt{\text{var}(L)}/E(L) > 1$. Otherwise, they are more sensitive to changes in the variance of lead time. This study has implications for where improvement resources should be allocated—to reduce the average lead time, or to reduce the variability of lead time. See also Nasri et al. (1990), Paknejad et al. (1992), Song (1994a,b), Lee et al. (2000), and Ray et al. (2004).

6.11 Exchange Curves Involving SSs for (s, Q) Systems

As was discussed in Chapters 1 and 2, management is primarily concerned with aggregate measures of effectiveness. In Chapter 4, exchange curves were developed that showed the aggregate consequences, in terms of the total average cycle stock versus the total number (or costs) of replenishments per year, when a particular policy (the EOQ) is used across a population of items. The curve was traced out by varying the value of the policy variable A/r . In a similar fashion, exchange curves can be developed showing the aggregate consequences, in terms of TSS (in dollars) versus aggregate service, when a particular decision rule for establishing SSs (i.e., reorder points) is used across a population of items.

Exchange curves are also useful as a means of determining the preferred level of service for a given investment in the inventory. Many inventory formulas assume that the firm has an infinite amount of working capital available. Obviously, this is never true. If the firm can achieve its service objectives without being constrained by dollar investment in the inventory, it is in an enviable position. Unfortunately, many firms face constraints on the amount of money available for investment in inventories, and must trade off inventory investment with service.

We begin, in Section 6.11.1, our discussion of exchange curves with an example of the inventory/service trade-off for a single item. Then, we show, in Section 6.11.2, a simple three-item example that illustrates the aggregate benefits that are usually possible in moving away from the commonly used approach of basing SSs (or reorder points) on equal time supplies. We also describe the general method of developing exchange curves.

As in Chapter 4, there is an input policy variable for each type of decision rule. For example, the management may decide to compute reorder points based on a common P_2 objective—that is, each item’s reorder point is determined using the formulas from Section 6.7.10. Management’s selection of an aggregate operating point, based on the exchange curve developed, implies a value of the policy variable—say, $P_2 = 0.97$. The use of this implicit value in the corresponding individual-item decision rule then leads to the desired aggregate consequences. An outline of the derivation of exchange curves is presented in Section 6.11.3. We restrict attention to (s, Q) systems but, because of the analogy developed in Section 6.9, the same type of approach can be used for (R, S) systems. In each case, note that by developing the formulas on a spreadsheet, it is straightforward to use a data table or similar functionality to quickly build exchange curves.

Management's ability to meaningfully utilize the SS versus service exchange curve hinges on their capacity to think of inventories as being split into the two components of cycle stocks and SSs and to be able to make separate tradeoffs of cycle stocks versus replenishments and SSs versus aggregate service. Based on our experience, these conditions are not always met. Therefore, in Section 6.11.4, we discuss the so-called composite exchange curves that show a three-way exchange of the total stock (cycle plus safety, in dollars) versus the number of replenishments per year versus aggregate service.

6.11.1 Single Item Exchange Curve: Inventory versus Service

When a manager is faced with a constraint on the amount of money available to invest in the inventory, it is necessary to determine the best customer service possible for the given investment. Often, however, the constraint is not rigid—it is possible to divert funds from other activities if it is determined that the gains in customer service are worth the loss in those other activities. An exchange curve, even for a single item, is extremely valuable in such situations. Figure 6.13 presents one example, using a fill rate (P_2) service objective and dollar investment in SS. Notice that the relationship is not linear; in fact, there are diminishing returns. In other words, to attain perfect customer service, the cost is exceedingly high. Also, notice that because of a relatively large order quantity, the actual SS required for low service levels is very low. The authors have used curves like

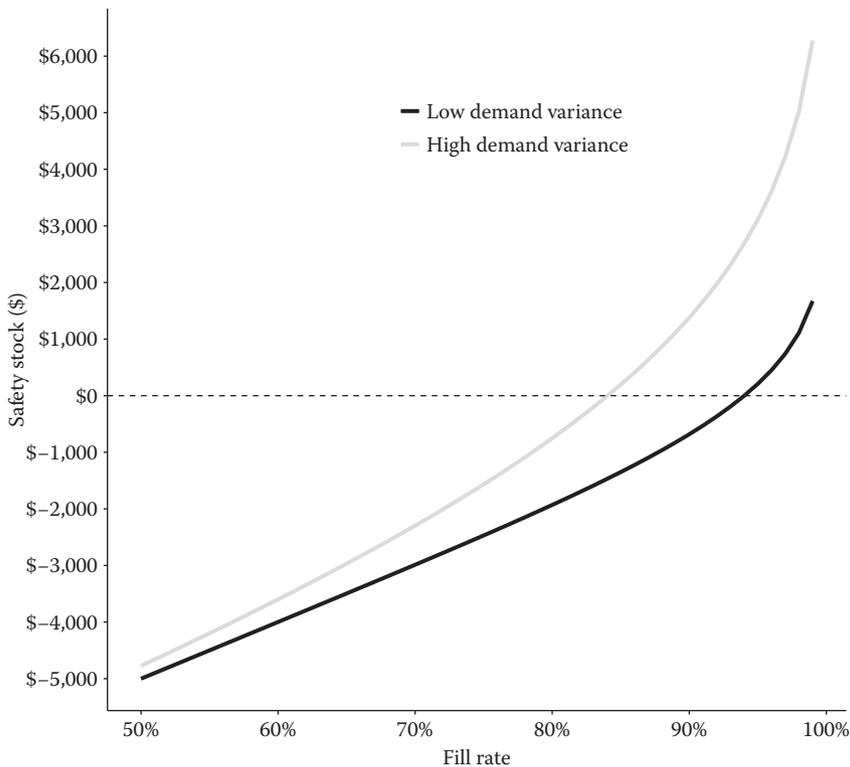


Figure 6.13 A single item exchange curve.

Figure 6.13 to show clients how much money is required to achieve their desired level of customer service. Alternatively, the manager can choose a service level based on the capital available.

It is critical, however, when using exchange curves in this way, to consider the possibility of changing the givens. What would happen, for instance, if the lead time were reduced? Or, what would be the effect of a closer relationship to the customers so that demand variability were reduced? Figure 6.13 illustrates the result of reducing demand variability by half for the particular item considered. The savings in SS are significant, especially for high service levels, implying that managers should consider investing resources to reduce demand variability.

6.11.2 An Illustration of the Impact of Moving away from Setting Reorder Points as Equal Time Supplies

Single item exchange curves can be useful; but, by far, the more common use of exchange curves is with multiple items. We now briefly discuss the general approach to developing exchange curves, an approach that is easy to use even with spreadsheets. As noted in Section 6.8, when an input objective is specified for an item, implied measures of performance are obtained. To develop exchange curves, one simply specifies an input objective (such as B_1 , B_2 , P_1 , or P_2). The computations required to find the reorder point for each item involve finding a k value for each item. This k value then determines the implied value of the other performance measures, such as B_1 , B_2 , P_1 , or P_2 . Managers may also be interested in TSS, the expected total number of stockout occasions per year (ETSOPY), or the expected total value of shortages per year (ETVSPY). See Table 6.8. These can then be graphed on an exchange curve.

It is not uncommon for organizations to use the following type of rule for setting reorder points: reorder when the inventory position has dropped to a 2-month time supply. Without going into detail, we wish to show the impact of moving away from this type of decision rule. To illustrate, let us consider three typical items produced and stocked by MIDAS International (a parent company, serving Canadian as well as other subsidiaries and direct customers throughout the world). Relevant characteristics of the items are given in Table 6.9. The order quantities, assumed as prespecified, are as shown. The current reorder points are each based on a 2-month time supply (i.e., $D/6$).

Assuming normally distributed forecast errors, it can be shown (by the procedure to be discussed in Section 6.11.3) that, under the use of the current reorder points, the SSs, the expected

Table 6.8 Exchange Curve Methodology

Input		Outputs
		B_1
		P_2
		P_1
B_1 or B_2 or P_1 or P_2 or ...	$\rightarrow k \rightarrow$	P_2
		TSS
		ETSOPY
		ETVSPY

Table 6.9 Three-Item Example

Item Identification	Demand Rate D (Units/Year)	Unit Value v (\$/Unit)	Lead Time L (Months)	σ_L (Units)	Current Order Quantity Q (Units)	Current Reorder Points s (Units)
PSP-001	12,000	20	1.5	300	2,000	2,000
PSP-002	6,000	10	1.5	350	1,500	1,000
PSP-003	4,800	12	1.5	200	1,200	800

stockout occasions per year, and the expected values short per year are as listed in columns (2), (3), and (4) of Table 6.10. If the TSS of \$14,900 is instead allocated among the items to give equal probabilities of stockout per replenishment cycle (i.e., equal P_1 values), the resulting expected stockout occasions per year and the expected values short per year are as in columns (6) and (7) of the table. The reallocation has reduced both aggregate measures of disservice substantially without increasing inventory investment. The reason for this is that the current equal-time supply strategy does not take into account of the fact that the forecasts for PSP-002 and PSP-003 are considerably more uncertain relative to their average demand during the lead time than PSP-001. Note the shift of SS away from PSP-001 to the other items under the revised strategy. The marked service improvement for PSP-002 and PSP-003 outweighs the reduction in service of PSP-001 caused by this reallocation of SS. The type of service improvement indicated here is typical of what can be achieved in most actual applications of logical decision rules for computing reorder points.

For either of the two service objectives shown in Table 6.10, we can achieve even greater improvement by determining reorder points using a B_1 or B_2 rule. As discussed in Section 6.7.12,

Table 6.10 Numerical Example of Shifting away from Reorder Points as Equal Time Supplies

Item No.	Equal Time Supply Reorder Points			Reorder Points Based on Equal Probabilities of Stockout per Cycle		
	Safety Stock ^a	Expected Stockout Occasions per Year	Expected Total Value Short per Year	Safety Stock	Expected Stockout Occasions per Year	Expected Total Value Short per Year
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	\$10,000	0.287	\$714	\$7,513	0.632	\$1,813
2	2,500	0.950	1,952	4,382	0.421	705
3	2,400	0.635	800	3,005	0.421	484
Totals	\$14,900	1.871/year	\$3,466/year	\$14,900	1.474/year	\$3,002/year

^a Under the given strategy safety stock (in units) = $s - \hat{x}_L = s - DL$. Then, safety stock (in \$) = $(s - DL)v$.

Table 6.11 Numerical Example of Setting Reorder Points with B_1 and B_2 Rules

Item No.	Reorder Points Based on a B_1 Rule			Reorder Points Based on a B_2 Rule		
	Safety Stock	Expected Stockout Occasions per Year	Expected Total Value Short per Year	Safety Stock	Expected Stockout Occasions per Year	Expected Total Value Short per Year
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	\$6,852	0.760	\$2,272	\$8,210	0.514	\$1,415
2	4,387	0.420	703	3,969	0.514	898
3	3,660	0.254	266	2,721	0.514	616
Totals	\$14,900	1.435/year	\$3,240/year	\$14,900	1.541/year	\$2,929/year

a B_1 rule will minimize the expected total stockouts per year for a given inventory budget. As indicated in Section 6.7.13, a B_2 will minimize ETVSPY for a given inventory budget. Keeping the same order quantities as above (see Table 6.9), we can calculate the SS allocations that minimize these service objectives for this numerical example. These allocations are shown in Table 6.11.

It is not a coincidence in Table 6.11 that the expected stockout occasions per year are the same for all three items when using a B_2 rule. Rearrangement of Equation 6.22 gives

$$\text{ESOPY} = \left(\frac{D}{Q} \right) p_{u \geq}(k) = \frac{r}{B_2}$$

So, if items are subject to the same holding cost rate r and the same fractional charge per unit short B_2 , the expected stockout occasions per year will be the same.

We can compare the performance of these different approaches for setting reorder points across a wide range and generate a multi-item exchange curve as shown in Figures 6.14 and 6.15. As we would expect, allocating SS according to a B_1 rule generates the best (lowest SS levels for a given level of service) exchange curve when the performance measure is the expected stockout occasions per year. In addition, this figure indicates how poorly a simplistic rule such as equal time supply can perform. Figure 6.15 shows an analogous result where an allocation based on a B_2 rule performs best when we seek to minimize the expected total value short.

6.11.3 Derivation of the SS Exchange Curves

We now provide a more formal treatment of exchange curve derivation. In all cases, for a given value of the relevant policy variable, the corresponding point on each exchange curve is found in the following manner. First, the value of the safety factor k_i is ascertained for each item by use of the appropriate decision rule. Then, any number of output performance measures are computed

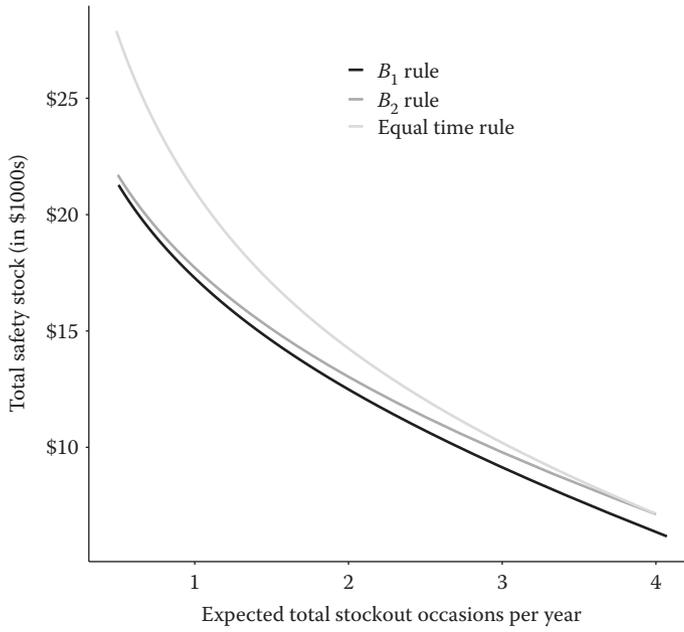


Figure 6.14 Exchange curves of TSS versus the expected total stockout occasions.

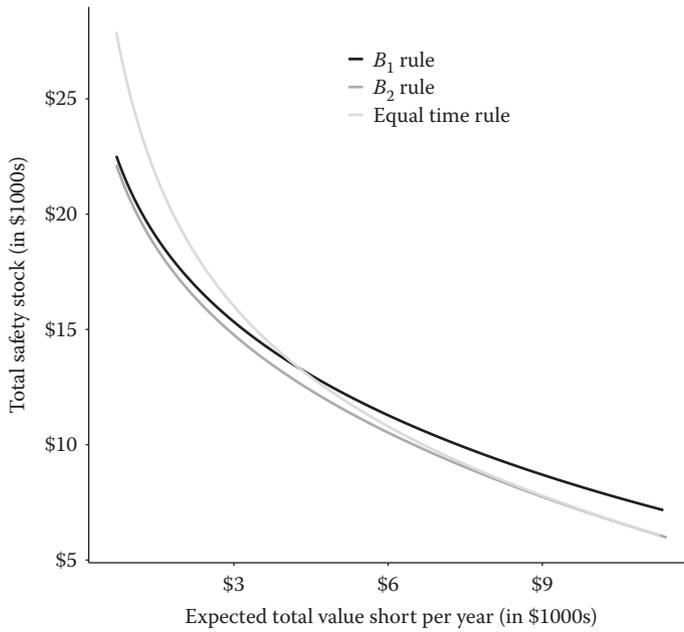


Figure 6.15 Exchange curves of TSS versus the expected total value short.

for each item i :

$$\text{Safety stock (in dollars) of item } i = SS_i v_i = k_i \sigma_{L_i} v_i \tag{6.43}$$

$$\text{Expected stockout occasions per year} = \frac{D_i}{Q_i} p_{u \geq}(k_i) \text{ for item } i \tag{6.44}$$

$$\text{Expected value short per year} = \frac{D_i}{Q_i} \sigma_{L_i} v_i G_u(k_i) \text{ for item } i^* \tag{6.45}$$

$$\text{Expected fill rate} = 1 - \frac{\sigma_{L_i} G_u(k_i)}{Q_i} \text{ for item } i \tag{6.46}$$

$$\text{Expected cycle service level} = 1 - p_{u \geq}(k_i) \text{ for item } i \tag{6.47}$$

These equations are, of course, based on an assumption of normally distributed errors. However, one could develop comparable results for any of the other distributions discussed in Section 6.7.14.

Equation 6.44 follows from multiplying the expected number of replenishment cycles per year (Equation 6.15) by the probability of a stockout per cycle (Equation 6.16). Similarly Equation 6.45 is a result of the product of the expected number of replenishment cycles per year (Equation 6.15) and the expected shortage per replenishment cycle, expressed in dollars (v_i times Equation 6.17). Finally, each of these quantities is summed across all items. In the last two cases, it is common to take a demand-weighted summation:

$$\text{Total safety stock (in dollars)} = \sum_i k_i \sigma_{L_i} v_i \tag{6.48}$$

$$\text{Expected total stockout occasions per year (ETSOPY)} = \sum_i \frac{D_i}{Q_i} p_{u \geq}(k_i) \tag{6.49}$$

$$\text{Expected total value short per year (ETVSPY)} = \sum_i \frac{D_i}{Q_i} \sigma_{L_i} v_i G_u(k_i) \tag{6.50}$$

$$\text{Demand-weighted fill rate} = \frac{\sum_i D_i \left(1 - \frac{\sigma_{L_i} G_u(k_i)}{Q_i}\right)}{\sum_i D_i} \tag{6.51}$$

$$\text{Demand-weighted cycle service level} = \frac{\sum_i D_i (1 - p_{u \geq}(k_i))}{\sum_i D_i} \tag{6.52}$$

The results of Equations 6.48 and 6.49 give a point on the first exchange curve (e.g., Figure 6.14) and those of Equations 6.48 and 6.50 give a point on the second curve (e.g., Figure 6.15) for the particular SS decision rule. Repeating the whole process for several values of the policy

* Strictly speaking, the expected value short per year for item i is

$$= \frac{D_i}{Q_i} \sigma_{L_i} v_i \left[G_u(k_i) - G_u\left(k_i + \frac{Q_i}{\sigma_{L_i}}\right) \right]$$

The approximation of Equation 6.45 is quite accurate as long as $Q_i/\sigma_{L_i} > 1$, because then the additional term becomes negligible. When $Q_i/\sigma_{L_i} < 1$, it is advisable to use the more accurate expression. See Equation 6.29.

variable generates a set of points for each of the two curves. Similar curves can be generated for demand-weighted fill rate and demand-weighted cycle service level.

In some cases, it is not necessary to go through all the calculations of Equations 6.48 through 6.52. To illustrate, for the case of a specified fraction (P_2) of demand to be satisfied directly off the shelf for all items,

$$\text{Expected total value short per year} = (1 - P_2) \sum_i D_i v_i$$

provided, of course, that no individual k_i has to be adjusted up to a minimum allowable value (thus providing a better level of service than P_2 for that particular item).

6.11.4 Composite Exchange Curves

For any decision rule for selecting SSs, we can combine the EOQ exchange curve showing the total average cycle stock versus the number of replenishments per year (recall Figure 4.12) with any of the SS exchange curves (TSS vs. aggregate service level). The result is a set of composite exchange curves showing a three-way trade-off of the total average stock versus the number of replenishments per year versus the aggregate service level. A separate line is required for each combination of decision rule and aggregate service measure. For the three-item numerical example from the previous section, we have shown an illustrative set of curves (for three levels of only one aggregate service measure) in Figure 6.16. They represent the use of the P_2 decision rule and an EOQ ordering strategy.

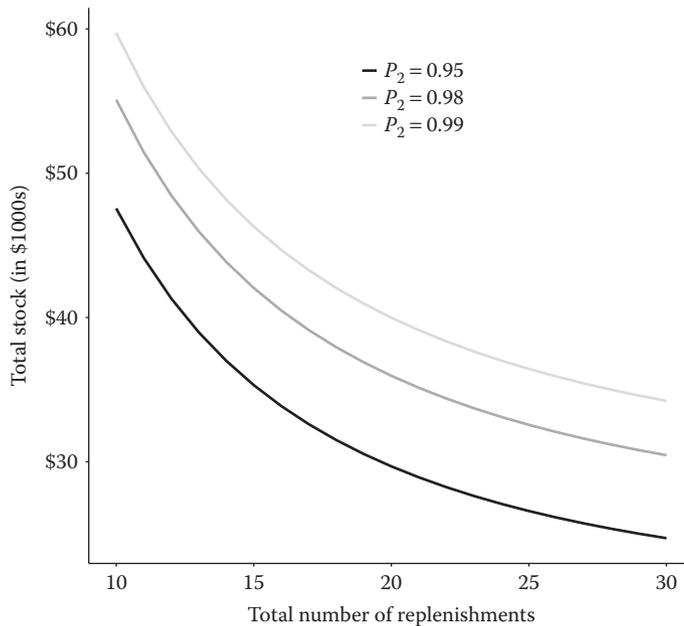


Figure 6.16 Composite exchange curves with a P_2 rule.

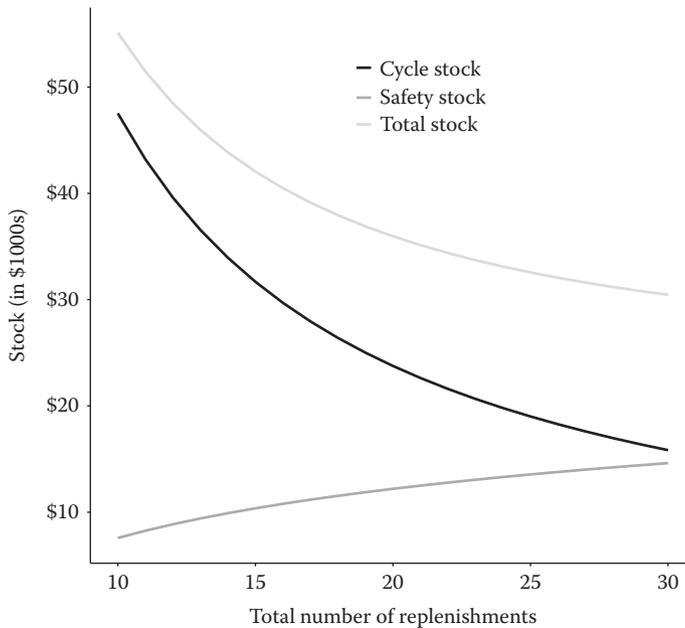


Figure 6.17 Breakdown of cycle stock and SS for an exchange curve with a P_2 rule.

Each curve of Figure 6.16 decreases as the number of replenishments N increases for a given aggregate service level (the value of P_2). This is a consequence of the reduction in cycle stock that results from more frequent replenishment dominating the increase in SS needed to maintain the desired P_2 level.

Figure 6.17 shows the behavior of cycle stock and SS separately for $P_2 = 0.98$. It is clear in this figure that more frequent replenishments reduce cycle stock and increase SS. It is quite possible that in some settings, the increase in SS exceeds the reduction in cycle stock, and the aggregate curves may eventually increase as the number of replenishments increases.

Selection of a desired aggregate operating point on a composite exchange curve implies:

1. A value of A/r to be used in the EOQ decision rule (this is not shown directly on the graphs; instead, we show the total number of replenishments per year N . A/r can be found from N using Equation 4.30 of Chapter 4).
2. A value of the parameter to be used in the particular SS decision rule under consideration (e.g., P_1 , TBS, B_1 , etc.).

For extensions of these concepts see Gardner and Dannenbring (1979) and Gardner (1983).

6.12 Summary

In this chapter, we have introduced the reader to the fundamentals of inventory control under probabilistic demand. For a variety of shortage costs and service measures, we have shown the details of two common types of decisions systems:

1. (s, Q) systems
2. (R, S) systems

The material presented included the actual decision rules, illustrative examples, and a discussion of each of the rules. Again, aggregate exchange curve concepts turned out to be an important aspect of the analysis. In Chapter 7, we will address two other, somewhat more complicated control systems—namely, (s, S) and (R, s, S) —in the context of A items.

It should now be apparent that, even for the case of a single item considered in isolation at a single stocking point, it is not an easy task to cope with uncertainty. It is little wonder that *usable* theory for more complex probabilistic situations (e.g., coordinated control of related items at a single location, multi-echelon control of a single item, etc.) is not extensive. We will have more to say on these issues in Chapters 10, 11, 15, and 16.

Problems

- 6.1 It is reasonable to assume that demand over its replenishment lead time of an item, ISF-086, is normally distributed with $\hat{x}_L = 20$ units and $\sigma_L = 4.2$ units.
 - a. For a desired service level of $P_1 = 0.90$, compute the required safety factor, SS, and reorder point.
 - b. Repeat for $P_1 = 0.94, 0.98, \text{ and } 0.999$.
 - c. Prepare a rough sketch of SS versus P_1 . Is the curve linear? Discuss.
- 6.2 The demand for a certain product has shown to be relatively stable; its sales are not subject to any seasonal influence, and demand is not expected to change in the future. In fact, during the last 10 lead times, the following demand was observed:

64	51	48	32	93	21	47	57	41	46
----	----	----	----	----	----	----	----	----	----

Assuming that this demand is a good representative of total demands in future lead times and a reorder point of 51 is used, compute:

- a. The various possible shortage sizes and their probabilities of occurrence (HINT: Assume that these are the only values that the demand can take, each with a 0.1 probability of occurring).
 - b. The expected (or average) shortage.
 - c. What reorder point would you use if the average shortage was to be no larger than 5 units? What is the associated probability of a stockout?
- 6.3 Consider an item with $A = \$25$; $Dv = \$4,000/\text{year}$; $\sigma_{Lv} = \$100$; $B_1 = \$30$; and $r = 0.10$ \$/\$/year.
 - a. Find the following:
 - i. EOQ, in dollars
 - ii. k , using the B_1 criterion
 - iii. SS, in dollars
 - iv. Annual cost of carrying SS
 - v. Total average stock, in dollars
 - vi. Expected number of stockout occasions per year
 - vii. Expected stockout costs per year

- b. The cost equation used to develop the rule for the B_1 criterion is

$$ETRC(k) = k\sigma_Lvr + \frac{D}{Q}B_1p_{u \geq}(k) \tag{6.53}$$

As seen in part 6.3a, the two components of Equation 6.53 are not equal at the optimal k value. (This is in contrast to what we found for the EOQ analysis in Chapter 4.) Why are the two components not equal? What quantities are equal at the optimal k value?

- c. By looking at the basic equations for the items requested in (a), discuss how each would be affected (i.e., whether it would be increased or decreased) by an increase in the r value.
- 6.4 Consider an item with $A = \$25$; $D = 4,000$ units/year; $v = \$8$ /unit; $L = 1$ month; $\sigma_L = 100$ units; TBS = 6 months; and $r = 0.16$ \$/\$/year.
- Compute the reorder quantity of the (s, Q) policy using the EOQ.
 - Compute the safety factor, SS, and reorder point.
 - Graph TBS versus ETVSPY (the expected total value short per year).
- 6.5 Consider an item that has an average demand rate that does not change with time. Suppose that demands in consecutive weeks can be considered as independent, normally distributed variables. Observations of total demand in each of 15 weeks are as follows:

89	102	107	146	155	64	78	122
78	119	76	80	60	115	86.	

- Estimate the mean and standard deviation of demand in a 1-week period and use these values to establish the reorder point of a continuously reviewed item with $L = 1$ week. A B_2 value of 0.3 is to be used, and we have $Q = 1,000$ units, $v = \$10$ /unit, $r = 0.24$ \$/\$/year, and $D = 5,200$ units/year.
 - In actual fact, the above-listed demands were randomly generated from a normal distribution with a mean of 100 units and a standard deviation of 30 units. With these true values, what is the appropriate value of the reorder point? What is the percentage cost penalty associated with using the result of part 6.5a instead?
- 6.6 Canadian Wheel Ltd. establishes SSs to provide a fraction of demand satisfied directly from stock at a level of 0.94. For a basic item with an essentially level average demand, the annual demand is 1,000 units and an order quantity of 200 units is used. The supplier of the item ensures a constant lead time of 4 weeks to Canadian Wheel. The current purchase price of the item from the supplier is \$0.80/unit. Receiving and handling costs add \$0.20/unit. The supplier offers to reduce the lead time to a new constant level of 1 week, but in doing so, she will increase the selling price to Canadian by \$0.05/unit. Canadian management is faced with the decision of whether or not to accept the supplier's offer.
- Qualitatively discuss the economic trade-off in the decision. (Note: Canadian would not increase the selling price of the item nor would they change the order quantity.)
 - Quantitatively assist management in the decision. (Assume that $\sigma_t = \sigma_1 t^{1/2}$ and $\sigma_4 = 100$ units where σ_4 is the standard deviation of forecast errors over the current lead

time of 4 weeks. Also assume that forecast errors are normally distributed and that the annual carrying charge is 20%.)

- 6.7 The 4N Company reacts to shortages at a particular branch warehouse in the following fashion: Any shortage is effectively eliminated by bringing material in from the plant by air freight at a cost of \$2.50/kg. Assume that the average demand rate of an item is essentially constant with time and that forecast errors are normally distributed. Assume that an (s, Q) system is to be used and that Q , if needed, is prespecified. To answer the following questions, introduce whatever symbols are necessary:
- What are the expected shortage costs per cycle?
 - What are the expected shortage costs per year?
 - What are the expected total relevant costs per year associated with the safety factor k ?
 - Develop an equation that the best k must satisfy.
 - If we had two items with all characteristics identical except that item 1 weighed more (on a unit basis) than item 2, how do you intuitively think that the two safety factors would compare? Verify this using the result of part (d).
- 6.8 The Norwich Company has a policy, accepted by all its customers, of being allowed up to 1 week to satisfy each demand. Norwich wishes to use a P_2 service measure with a specified P_2 value. Consider an item controlled by an (s, Q) system with Q prespecified. Suppose that the lead time L was equal to the grace period.
- Under these circumstances, what s value is needed to give $P_2 = 1.00$?
 - More generally, consider an item with a desired $P_2 = 0.99$ and a Q value of 500 units. Other (possibly) relevant characteristics of the item are
 - $D = 1,000$ units/year (the average is relatively level throughout the year)
 - $v = \$5$ /unit
 - $r = 0.20$ \$/\$/year
 - $L = 5$ weeks
 - $\sigma_1 = 10$ units (for a 1-week period); $\sigma_t = \sigma_1 t^{1/2}$
 How would you modify the logic for computing the s value to take into account of the 1-week “grace” period? Illustrate with this item.
 - For the item of part (b), how would you estimate the annual cost savings of having the grace period?
- 6.9 A retail electronic-supply outlet orders once every 2 weeks. For a particular item with unit value (v) of \$3, it is felt that there is a cost of \$1 for each unit demanded when out of stock. Other characteristics of the item are
- $L = 1$ week
 - $r = 30\%$ per year
- Consider a period of the year where $\hat{x}_1 = 30$ units and $MSE_1 = 6$ units (where the unit time period is 1 week). What order-up-to-level S should be used? (Assume $\sigma_t = \sigma_1 t^{1/2}$.)
 - Suppose the average demand rate changed slowly throughout the year. Suggest a reasonable way for the retailer to cope with this situation.
- 6.10 A particular company with production facilities located in Montreal has a large distribution warehouse in Vancouver. Two types of transportation are under consideration for shipping material from the plant to the remote warehouse, namely,

Option A: Send all material by railroad freight cars.

Option B: Send all material by air freight.

If only one of the two pure options can be used, *outline* an analysis that you would suggest to help the management choose between the two options. Indicate the types of data that you would require. *Intuitively* would it ever make sense to use both options, that is, to use rail freight for part of the shipments and air freight for the rest? Briefly explain why or why not.

- 6.11 You have signed a consignment contract to supply a chemical compound to a major customer. This compound costs you \$32.50/gallon and generates a contribution of \$12.50/gallon. Delivery costs you \$175 in shipping and handling. Demand for this year is estimated at 2,000 gallons, standard deviation is 400 gallons. The contract says that the compound is inventoried at your client's facility: he will draw the material from that stock as needed and pay you at such time. It also says that if the client finds that inventory empty when needed, you are subject to a fee of \$0.325/gallon/day short. Inventory carrying cost is 0.20 \$/\$/year.
- Determine the optimal (s, Q) policy, assuming a stochastic lead time, $L = 30$ days, standard deviation of lead time = 6.
 - Your customer has realized that controlling the stockout charge is too burdensome. Two alternative charges have been proposed:
 - A flat charge of \$1,000 per stockout occasion.
 - A \$65 charge per gallon short per cycle.
 Under each alternative, what would your policy be? Are the proposed changes acceptable?
- 6.12 A fast-food restaurant has just launched a new hamburger recipe. All of its products are prepared in small batches, targeting a fill rate $P_1 = 0.7$ at lunch time. The manager observed the demand while the first few batches of the new sandwich were prepared. They were:

3	7	2	0	10	5
4	8	1	6	11	8

- For the next few days (until better information is gathered) what reorder point should be used?
 - What additional information would you need in order to decide the size of the next batch, now that you have the reorder point?
- 6.13 Consider an (r, S) policy where the management adopts a stockout charge B_2 per unit short. Is such a policy sensitive to the individual item's parameters?
- 6.14
- A company manager has decided that the inventory control policy "should have zero stockouts." Is this a realistic policy to pursue? In particular, discuss the implications in terms of various types of costs.
 - On occasion, this company has filled an important customer order by supplying a competitor's product at below the retail price. What form of shortage costing would it be appropriate to apply under such circumstances? Does this type of strategy make sense?
- 6.15 Safety Circuits Inc. stocks and sells small electronics products using an (R, S) type of control system. A manager reviews the stock monthly. The supplier delivers at their location following a lead time of 10 days. Inventory carrying charge is 0.12 \$/\$/year. The manager has observed that demand for the XRL line is as follows:

Item i	Demand D_i (Packs/Year)	v_i (\$/Pack)	$\sigma_{i,1}$ (Packs)
XRL-1	1,200	10	35
XRL-2	350	35	50
XRL-3	700	17	40

where $\sigma_{i,1}$ is the standard deviation of yearly demand for item i . Suppose that a TSS of \$1,200 is to be allocated among the three items. Consider the following service measures:

1. Same k for the 3 items
2. Same P_1 for the 3 items
3. Same TBS for the 3 items
4. Same B_1 for the 3 items
5. Same B_2 for the 3 items

For each, determine

- a. How the \$1,200 is allocated among the three items?
- b. ETSOPY
- c. ETVSPY (expected total value short per year)

Note: Assume that forecast errors are normally distributed and the same r value is applicable to all three items. Negative safety factors are not permitted.

6.16 Safety Circuits Inc. is considering investing in a scanner such that they can move from an (R, S) to an (s, Q) inventory policy (see Problem 6.15). It is expected that with a continuous-review policy, the firm could reduce inventory management costs significantly: holding and shortage costs could be reduced. Management believes that for each unit short, there is an implicit charge $B_2 = 0.15$.

- a. Determine the safety factor, SS , and reorder point for each of the three products, assuming that the same lot sizes are adopted.
- b. Determine the total relevant costs for the XRL product line under an (s, Q) policy and $B_2 = 0.15$.
- c. Determine the total relevant costs for the XRL product line under an (R, S) policy and $B_2 = 0.15$.
- d. Assume that the scanner costs \$30,000, and its operating costs are immaterial. The XRL product line is a good representative of the products carried by SFI, and that it comprises 1% of the total average stock for the firm. Calculate the time before the investment breaks even (disregard cost of money). Would you recommend the change? Discuss.

6.17 Waterloo Warehouse Ltd. (WWL) acts as a distributor for a product manufactured by Norwich Company. WWL uses an (s, Q) type of control system. Demands on Waterloo can be assumed to be unit sized. Norwich has a particular way of delivering the quantity Q : They deliver $Q/2$ at the end of the lead time of 1 month and the other $Q/2$ at the end of 2 months. Assume that this business has the following characteristics:

$$Q = 100 \text{ units}$$

$$s = 40 \text{ units}$$

$$\sigma_1 = 10 \text{ units } (\sigma_1 = \text{the standard deviation of forecast errors over 1 month})$$

$$D = 360 \text{ units/year}$$

$$v = \$1.00/\text{unit}$$

$$r = 0.25 \text{ \$/\$/year}$$

- a. What is the numerical value of the SS? Explain your answer.
 - b. Suppose that a replenishment is triggered at $t = 0$, and no order is outstanding at this time. Describe in words what conditions must hold in order for there to be no stockout up to time $t = 2$ months.
 - c. Suppose that Norwich offered WWL the option to receive the whole quantity Q after a lead time of 1 month, but at a higher unit price. Outline the steps of an analysis to decide whether or not the option should be accepted.
- 6.18 True Colors Co. purchases a number of chemical additives for its line of colors and dyes. After a recent classification of stock, the inventory manager has to decide the purchase criteria for the B items. Consider the pigment Red-105: demand forecast for the next 12 months of 200 gallons, with a standard deviation of 50 gallons. They cost \$25/gallon at the supplier's plant, in addition to \$30 per order for shipping and handling; expected lead time is 45 days. It is believed that each day costs the firm \$5 per gallon short. Inventory carrying charge is 0.18 \$/\$/year.
- a. What (s, Q) policy should be adopted for this product. What is the implicit TBS?
 - b. Looking through some of the recent orders, the manager observed that, in fact, 90% of the shipments arrived in 30 days, the remainder in 75 days (the later arrivals due to unpredictable losses at the supplier). Should the inventory policy be adjusted to account for this lead time uncertainty? Discuss.
- 6.19 Consider a group of n items that are each being independently controlled using an (R, S) type of system (periodic-review, order-up-to-level). Let R_i be the prespecified value (in months) of R for item i . Suppose that we wish to allocate a total given SS in dollars (call it W) among the n items to minimize the dollars of expected backorders per unit time summed across all n items.
- a. Develop the decision rule for selecting k_i , the safety factor for item i . Assume that forecast errors are normally distributed.
 - b. Illustrate for the following two-item case ($r = 0.2 \text{ \$/\$/year}$; $W = \$1,180$; and $L = 1$ month):

Item i	v_i (\$/Units)	A_i (\$)	Demand D_i (Units/Year)	R_i (Months)	σ_{R_i} (Units)	σ_{L_i} (Units)	σ_{R+L_i} (Units)
1	1	2.00	3,000	1.0	200	200	300
2	2	5.00	15,000	0.5	500	300	700

- 6.20 Consider the following product line:

Item i	Demand D_i (Units/Year)	v_i (\$/Units)	$\sigma_{i,1}$ (Units)	A (\$)
1	100	250.00	45	100.00
2	200	30.00	140	100.00
3	300	250.00	210	150.00
4	60	210.00	35	150.00

Inventory carrying charges are 0.12 \$/\$/year, and the lead time is 3 months. Management feels that a fractional charge of 0.3 for each item short should be considered. Assume that lead time demand is normally distributed.

- a. Determine the optimal (s, Q) policy, identifying SS, reorder point, and reorder quantity for each item using the EOQ.
- b. After more careful study, management realized that lead time demand is better described by the Gamma distribution. Adjust the (s, Q) policy in the previous question to account for the actual distribution.
- c. Data from the previous 10 cycles give the actual lead time demand for item 3 as follows:

0	32	308	77	60	105	6	81	208	8
---	----	-----	----	----	-----	---	----	-----	---

What would have been the actual shortage per year under each of the suggested policies?

6.21 Consider three items with the following characteristics:

Item i	Demand D_i (Units/Year)	v_i (\$/Units)	Q_i (Units)	σ_{L_i} (Units)
1	1,000	1.00	200	100
2	500	0.80	200	100
3	100	2.00	100	30

Suppose that a TSS of \$120 is to be allocated among the three items. Consider the following service measures (shortage-costing methods):

1. Same P_1 for the 3 items
2. Same P_2 for the 3 items
3. Same B_1 for the 3 items
4. Same B_2 for the 3 items
5. Same TBS for the 3 items
6. Minimization of total expected stockout occasions per year
7. Minimization of total expected value short occasions per year
8. Equal time supply SSs

For each, determine

- a. How the \$120 is allocated among the three items
- b. ETSOPY
- c. ETVSPY

Note: Assume that forecast errors are normally distributed, that the same r value is applicable to all three items, and that negative safety factors are not permitted.

6.22 Botafogo Circuits imports some electronic components and redistributes them in the Brazilian market. This includes some difficulties, such as delivery uncertainty. Using regular air freight, shipping and handling costs \$250 per item ordered, but the delivery may take anywhere from 10 to 30 days. Because of this uncertainty, the manager contacted a number of carriers looking for more reliable alternatives. Two distinct bids stood up:

- a. An express importer would deliver any order within 15 days, at a cost of \$350 per item ordered.

- b. A regular shipper would pick orders every 15th of the month and deliver them on the 15th of the next month, for a cost of \$200 per order plus \$0.50 per unit, for any number of items ordered.

Demand for these items can be approximated by a Normal distribution. The following table shows the mean and standard deviation of the yearly demand, as well as the unit costs for some of these items:

Item i	Demand D_i (Packs/Year)	v_i (\$/Pack)	$\sigma_{i,1}$ (Packs)
XRL-10	1,200	40.00	860
XRL-12	1,250	35.00	500
XRL-15	1,800	22.00	120
XRL-20	700	17.00	400

In its internal accounts, the manager considers a fractional charge of 0.50 for each unit short, and an inventory carrying charge of 0.30 \$/\$/year. Describe the inventory policy that should be used with each of the transportation alternatives. Which alternative should the firm use?

- 6.23 Consider an item under (s, Q) control. Basic item information is as follows: $D = 40,000$ units/year; $A = \$20$; $r = 0.25$ \$/\$/year; and $v = \$1.60$ /unit. All demand when out of stock is backordered. The EOQ is used to establish the Q value. A service level of 0.95 (demand satisfied without backorder) is desired. The item's demand is somewhat difficult to predict and two forecasting procedures are possible.

System	Cost to Operate per Year (\$/Year)	σ_L (Units)
A (complex)	200	1,000
B (simple)	35	2,300

Which forecasting system should be used? Discuss. *Note:* Forecast errors can be assumed to be normally distributed with zero bias for both models.

- 6.24 Consider two items with the following characteristics:

Item	Unit Value (\$/Unit)	Demand D (Units/Year)	x_L (Units)	σ_L (Units)
1	10.00	300	100	10
2	1.00	300	100	35

Suppose that the inventory controller has set the SS of each of these items as a 1-month time supply.

- What are the SSs in units? In dollars?
- What is the P_1 value associated with each item?
- Reallocate the same TSS (in dollars) so that the two items have the same value of P_1 .
- What reduction in TSS is possible if both items have $P_1 = 0.95$?

6.25 Upper Valley Warehouse adopts an (s, Q) inventory policy where all items meet a safety factor $k = 1.2$. Some of the B items enjoy a deterministic lead time of 1 month from their suppliers. They are the following products:

Item i	Demand D_i (Units/Year)	v_i (\$/Units)	$\sigma_{i,1}$ (Units)	A (\$)
1	1,000	50.00	300	100.00
2	2,000	30.00	700	100.00
3	1,000	100.00	700	100.00
4	800	80.00	500	120.00
5	1,000	35.00	400	120.00
6	3,000	150.00	800	120.00
7	600	210.00	200	150.00

Inventory carrying cost is 0.12 \$/\$/year. Reorder quantity is calculated by the EOQ.

- a. Determine
 - i. Implicit TBS for each product
 - ii. ETSOPY
 - iii. ETVSPY (expected total value short per year)
 - iv. SS in dollars
 - b. The firm has decided to allocate \$50,000 into the SS of these products. What should be the individual safety factor, SS, and the reorder point under each of the following policies:
 - i. Same TBS for the 7 items
 - ii. Minimize ETSOPY
 - iii. Minimize ETVSPY
- 6.26 Consider an item with the following characteristics:
- $D = 1,000$ units/year
 - $L = 4$ weeks
 - $A = \$20$
 - $v = \$5$ per unit
 - $r = 0.20$ \$/\$/year
 - $\hat{x}_L = 80$ units
 - $\sigma_L = 8$ units
 - $B_2 = 0.8$
- a. Use the approach of Lau and Lau in Section 6.7.14 to find the optimal order quantity. Then find $F_x(s)$. Finally, assume that demand is normal, and find k and s .
 - b. Use the approach of Section 6.7.6 to find the optimal s and Q .
- 6.27 Consider the item discussed in Approach 2 in Section 6.10.2. Recall that with $\text{var}(L) = 0$, the reorder point is 457 units if P_1 is 0.95. Now suppose the firm actually reduces the lead time by working with its supplier. Unfortunately, the reduced lead time is not always achieved. In particular, the lead time is 2 weeks half of the time, but remains at 4 weeks half

the time. So

$$L = 4 \text{ with probability } 0.5$$

$$\text{and } L = 2 \text{ with probability } 0.5$$

Thus, $E(L) = 3$, and $\text{var}(L) = 0.5(2 - 3)^2 + 0.5(4 - 3)^2 = 1$.

- a. Use Equations 6.41 and 6.42 to find the total lead time demand, and then find the reorder point assuming that the total demand is normally distributed.
 - b. Eppen and Martin (1988) provide an alternative solution that relies on the actual distributions. Let k_1 be the safety factor when the lead time is 2 weeks; and let k_2 be the safety factor when the lead time is 4 weeks. Find an expression for the reorder point as a function of k_1 and another expression for the reorder point as a function of k_2 . Now find the optimal reorder point by noting that the weighted average service level is 95%.
- 6.28 Tower Bookstore orders books from one particular supplier every Tuesday. Once the order is placed, it takes 3 days for delivery, and an additional four days to get the books on the shelf. This season there is a book—*The One Minute Parent*—that is selling especially well. The purchasing manager has gathered some data that she anticipates will be representative of demand for the next several months. It turns out that weekly demand has been 75 books on average, with a standard deviation of 5. The carrying charge is 18% per dollar invested in books, and the manager has determined that the charge per unit short is 30% of the sales price. Each copy costs Tower \$5; and they apply a 50% markup to books of this type. Find the correct ordering policy.
- 6.29 A materials manager for Missoula Manufacturing has provided the following data for one of their widely used components:

$Q^* = 50$	$D = 500$	$rv = \$10$	$A = \$25$
------------	-----------	-------------	------------

Lead time demand for the component is normal with a mean of 25 and a variance of 25. What is the expected quantity of backordered items per year if the manager uses an optimal (s, Q) policy?:

- a. The desired service level (the probability of no stockout during the lead time) is 90%.
- b. The desired service level is 95%.

Appendix 6A: Some Illustrative Derivations and Approximations

6A.1 B_1 Shortage Costing

We show two different approaches: the first using the total relevant costs; the second based on a marginal analysis. The latter is simpler but does not permit sensitivity analysis (in that a total cost expression is not developed).

6A.1.1 Total Cost Approach

We recall from Section 6.7.5 the expression for the expected total relevant costs per year as a function of the control parameter k . Let us denote this function by ETRC (k). There are three

relevant components of costs: (1) replenishment (C_r), (2) carrying (C_c), and (3) stockout (C_s). So,

$$\begin{aligned} \text{ETRC}(k) &= C_r + C_c + C_s \\ &= AD/Q + (Q/2 + k\sigma_L)vr + \frac{DB_1}{Q}p_{u\geq}(k) \end{aligned}$$

We wish to select the k value that minimizes $\text{ETRC}(k)$. A convenient approach is to set

$$\frac{d\text{ETCR}(k)}{dk} = 0$$

that is,

$$\sigma_Lvr + \frac{DB_1}{Q} \frac{dp_{u\geq}(k)}{dk} = 0 \tag{6A.1}$$

But

$$\frac{dp_{u\geq}(k)}{dk} = -f_u(k)$$

(The derivative of the cumulative distribution function is the density function as shown in Appendix II.) Therefore, Equation 6A.1 gives

$$f_u(k) = \frac{Qv\sigma_Lr}{DB_1} \tag{6A.2}$$

Now the density function of the unit normal is given by

$$f_u(k) = \frac{1}{\sqrt{2\pi}} \exp(-k^2/2)$$

Therefore, Equation 6A.2 gives

$$\frac{1}{\sqrt{2\pi}} \exp(-k^2/2) = \frac{Qv\sigma_Lr}{DB_1} \tag{6A.3}$$

which solves for

$$k = \sqrt{2 \ln \frac{DB_1}{\sqrt{2\pi}Qv\sigma_Lr}} \tag{6A.4}$$

However, it is necessary to proceed with caution. Setting the first derivative to zero does not guarantee a minimum. In fact, Equation 6A.4 will have no solution if the expression inside the square root is negative. This will occur if the argument of the logarithm is less than unity. A case where this happens is shown in the lower part of Figure 6A.1. In such a situation, the model says that the lower the k is, the lower the costs are. The model assumes a linear savings in carrying costs with decreasing k . Obviously, this is not true indefinitely. The practical resolution is to set k at its lowest allowable value.

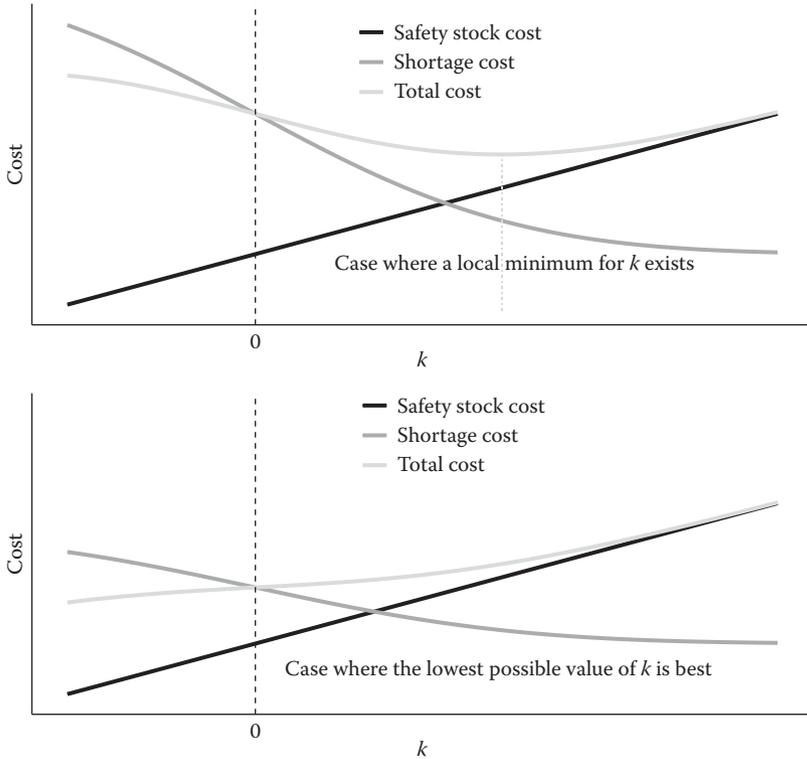


Figure 6A.1 Behavior of the expected total relevant costs for the case of a fixed cost per stockout.

6A.1.2 Marginal Analysis

Let us consider the effects of increasing s by a small amount, say Δs . The extra carrying costs per year are given by

$$\text{Marginal cost/year} = \Delta s vr \tag{6A.5}$$

Savings result when, in a lead time, the total demand lies exactly between s and $s + \Delta s$, so that the extra Δs has avoided a stockout. On the average, there are D/Q replenishment occasions per year. Thus,

$$\begin{aligned} \text{Expected marginal savings/year} &= \frac{D}{Q} B_1 \Pr\{\text{Demand lies in interval } s \text{ to } s + \Delta s\} \\ &= \frac{D}{Q} B_1 f_x(s) \Delta s \end{aligned} \tag{6A.6}$$

Now consider a particular value of s . If the marginal cost associated with adding Δs to it exceeds the expected marginal savings, we would not want to add the Δs . In fact, we might want to go below the particular s value. On the other hand, if the expected marginal savings exceeded the marginal cost, we would definitely want to increase s by Δs . However, we might wish to continue increasing s even higher. This line of argument leads us to wanting to operate at the s value,

where the marginal cost of adding Δs exactly equals the expected marginal savings. Thus, from Equations 6A.5 and 6A.6, we wish to select s so that

$$\Delta svr = \frac{D}{Q} B_1 f_x(s) \Delta s$$

or

$$f_x(s) = \frac{Qvr}{DB_1} \tag{6A.7}$$

This is a general result that holds for any probability distribution of lead time demand. For the normal distribution, we have that

$$f_x(s) = \frac{1}{\sqrt{2\pi}\sigma_L} \exp\left[-(s - \hat{x}_L)^2 / 2\sigma_L^2\right] \tag{6A.8}$$

Moreover, from Equations 6.7 and 6.8,

$$s = \hat{x}_L + k\sigma_L \tag{6A.9}$$

Substitution of Equations 6A.8 and 6A.9 into Equation 6A.7 gives the same result as Equation 6A.3.

6A.1.3 Behavior of k as a Function of Dv

Equation 6A.4 shows that k increases as $D/Qv\sigma_L$ or $Dv/Qv\sigma_L v$ increases. Now, empirically it has been found that, on the average,

$$\sigma_L v \approx c_1 (Dv)^{c_2}$$

where c_1 and c_2 are constants that depend on the particular company involved but, in all known cases, c_2 lies between 0.5 and 1. Also assuming an EOQ form for Qv , we have

$$Qv = c_3 \sqrt{Dv}$$

Therefore, we have that k increases as

$$\frac{Dv}{c_3 \sqrt{Dv} c_1 (Dv)^{c_2}} \quad \text{or} \quad \frac{(Dv)^{1/2-c_2}}{c_1 c_3}$$

increases. But, because $c_2 > \frac{1}{2}$, the exponent of Dv in the above quantity is negative. Therefore, k goes down as Dv increases.

6A.2 P_1 Service Measure

Suppose the reorder point is expressed as the sum of the forecast demand plus SS, as shown in Equation 6.7. Furthermore, suppose, as in Equation 6.8, that the SS is expressed as the product of two factors

1. The safety factor k —the control parameter
2. σ_L , the standard deviation of forecast errors over a lead time

Then from Equation 6.16, we have

$$\Pr\{\text{stockout in a lead time}\} = p_{u \geq}(k) \quad (6A.10)$$

If the desired service level is P_1 , then we must have

$$\Pr\{\text{stockout in a lead time}\} = 1 - P_1 \quad (6A.11)$$

It follows from Equations 6A.10 and 6A.11 that we must select k to satisfy

$$p_{u \geq}(k) = 1 - P_1$$

which is Equation 6.26.

Had we not chosen the indirect route of using Equations 6.7 and 6.8, we would have ended up with a more cumbersome result than Equation 6.26, namely,

$$p_{u \geq}[(s - \hat{x}_L)/\sigma_L] = 1 - P_1$$

6A.3 P_2 Service Measure

6A.3.1 Complete Backordering

Because each replenishment is of size Q , we can argue as follows:

$$\text{Fraction backordered} = \frac{\text{Expected shortage per replenishment cycle, ESPRC}}{Q}$$

Fraction of demand satisfied directly from shelf = $1 - (\text{Fraction backordered})$.

Therefore, we want

$$P_2 = 1 - \frac{\text{ESPRC}}{Q} \quad (6A.12)$$

Substituting from Equation 6.17, we have

$$P_2 = 1 - \frac{\sigma_L G_u(k)}{Q}$$

which reduces to

$$G_u(k) = \frac{Q}{\sigma_L} (1 - P_2)$$

which is Equation 6.28.

6A.3.2 Complete Lost Sales

If demands when out of stock are lost instead of backordered, then, the expected demand per cycle is no longer just Q , but rather is increased by the expected shortage per replenishment cycle, ESPRC. Therefore, we now have

$$\text{Fraction of demand satisfied directly from shelf} = 1 - \frac{\text{ESPRC}}{Q + \text{ESPRC}}$$

The derivation then carries through exactly as above, resulting in Equation 6.27.

6A.4 Allocation of a TSS to Minimize the ETSOPY

The problem is to select the safety factors, the k_i 's, to minimize

$$\sum_{i=1}^n \frac{D_i}{Q_i} p_{u \geq}(k_i)$$

subject to

$$\sum_{i=1}^n k_i \sigma_{L_i} v_i = Y$$

where n is the number of items in the population and Y is the TSS expressed in dollars. One method of the solution is to use a Lagrange multiplier M (see Appendix I), that is, we select the k_i 's to minimize

$$L(k_i's, M) = \sum_{i=1}^n \frac{D_i}{Q_i} p_{u \geq}(k_i) - M \left(Y - \sum_{i=1}^n k_i \sigma_{L_i} v_i \right)$$

This is accomplished by setting all the partial derivatives to zero. Now

$$\frac{\partial L}{\partial k_i} = -\frac{D_i}{Q_i} f_u(k_i) + M \sigma_{L_i} v_i = 0$$

or

$$f_u(k_i) = M \frac{Q_i v_i \sigma_{L_i}}{D_i}$$

But, this has the exact same form as Equation 6A.2 with $M \equiv r/B_1$. Therefore, selection of a particular M value implies a value of B_1 . Also, of course, a given M value leads to a specific value of Y .

The economic interpretation of the Lagrange multiplier M is that it represents the marginal benefit, in terms of reduced total expected stockouts, per unit increase in the total dollar budget available for SSs. r is the cost to carry one dollar in SS for a year. B_1 is the cost associated with a stockout occasion. From a marginal viewpoint, we wish to operate where

$$\begin{aligned} &(\text{Cost to carry last dollar of safety stock}) = \\ &(\text{expected benefit of last dollar of safety stock}) \end{aligned}$$

that is,

$$\begin{aligned} r &= (\text{Reduction in total expected stockouts per year}) \times (\text{Cost per stockout}) \\ &= MB_1 \end{aligned}$$

thus, $M = r/B_1$, as we found above.

6A.5 Standard Deviation of Demand during Lead Time

Demand in any period t is assumed to be a random variable D_t with mean $E(D)$ and standard deviation $\sqrt{\text{var}(D)}$. In this analysis, demands in different periods are assumed to be independent and identically distributed and that the random variable for lead time L takes on only integer values. As in Equations 6.41 and 6.42, x has mean $E(x)$ and standard deviation σ_x , and lead time L has mean $E(L)$ and variance $\text{var}(L)$.

Demand during the lead time x is a sum of random variables D_t where the number of terms in the sum is also random (L).

$$x = \sum_{i=1}^L D_i$$

We can find both the mean and standard deviation of X by conditioning on the value of the lead time L . First, for the mean,

$$\begin{aligned} E[x] &= E_L[E_D[x|L]] \\ &= E_L[(L)E(D)] \\ &= E(D)E(L) \end{aligned}$$

Now, for the standard deviation, note that for any random variable Z , we can find the variance of Z by conditioning on another random variable (e.g., Ross 1983) say Y , as follows:

$$\text{var}[Z] = E_Y[\text{var}_Z(Z|Y = y)] + \text{var}_Y[E_Z(Z|Y = y)]$$

So, again conditioning on the value of the lead time L

$$\begin{aligned} \sigma_x^2 &= E_L[\text{var}_D(x|L)] + \text{var}_L[E_D(x|L)] \\ &= E_L[(L)\text{var}(D)] + \text{var}_L[E(D)(L)] \\ &= (E(L)\text{var}(D) + [E(D)]^2\text{var}(L)) \\ \sigma_x &= \sqrt{E(L)\text{var}(D) + [E(D)]^2\text{var}(L)} \end{aligned}$$

6A.6 Two Approaches to Determining Expected Costs per Unit Time

Throughout the book, the basic approach we use in establishing the expected total relevant costs per unit time will follow these steps:

1. Determine the average inventory level (\bar{I}), in units.
2. Carrying costs per year = $\bar{I}vr$.
3. Determine the average number of replenishments per year (N).
4. Replenishment costs per year = NA .
5. Determine the expected shortage costs per replenishment cycle, say W .
6. Expected shortage costs per year = NW .
7. Total expected costs per year = sum of the results of steps 2, 4, and 6.

If the inventory process has a natural regenerative point where the same state repeats periodically, then, a different approach, using a key result of renewal theory, is possible. Each time that the regenerative point is reached, a cycle of the process is said to begin. The time between consecutive regenerative points is known as the length of the cycle, which we denote by T_c .

There is a set of costs (carrying, replenishment, and shortage) associated with each cycle. Let us denote the total of these costs by C_c . This quantity is almost certainly a random variable because its value depends on the demand process that is usually random in nature.

A fundamental result of renewal theory (see, e.g., Ross 1983), which holds under very general conditions, is

$$\text{Expected total relevant costs per unit time} = \frac{E(C_c)}{E(T_c)}$$

The key point is that in a complicated process, it may be easier to evaluate each of $E(C_c)$ and $E(T_c)$ instead of the left-hand side of this equation directly. See Silver (1976) for an illustration of this approach.

6A.7 Expected Units Short per Cycle with the Gamma Distribution

Suppose demand follows a Gamma distribution described by shape parameter α and scale parameter β , implying a mean of $\alpha\beta$ and standard deviation of $\sqrt{\alpha}\beta$. Assuming a reorder point of s , we write expected units short per replenishment cycle as

$$\text{ESPRC} = \int_s^{\infty} (x - s)f(x; \alpha, \beta)dx$$

where $f(x; \alpha, \beta)$ represents the probability density function of the Gamma distribution with parameters α and β

$$f(x; \alpha, \beta) = \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^{\alpha}\Gamma(\alpha)}$$

Γ is the Gamma function that has the property $x\Gamma(x) = \Gamma(x + 1)$.

Note that

$$\begin{aligned}
 xf(x; \alpha, \beta) &= x \frac{x^{\alpha-1} e^{-x/\beta}}{\beta^\alpha \Gamma(\alpha)} \\
 &= \alpha \beta \frac{x^\alpha e^{-x/\beta}}{\beta^\alpha \alpha \Gamma(\alpha)} \\
 &= \alpha \beta \frac{x^\alpha e^{-x/\beta}}{\beta^{\alpha+1} \Gamma(\alpha+1)} \\
 &= \alpha \beta f(x; \alpha+1, \beta)
 \end{aligned}$$

Using this fact,

$$\begin{aligned}
 \text{ESPRC} &= \int_s^\infty (x-s)f(x; \alpha, \beta) dx \\
 &= \int_s^\infty xf(x; \alpha, \beta) dx - s[1 - F(s; \alpha, \beta)] \\
 &= \alpha \beta [1 - F(s; \alpha+1, \beta)] - s[1 - F(s; \alpha, \beta)]
 \end{aligned}$$

References

- Aardal, K., Ö. Jonsson, and H. Jönsson. 1989. Optimal inventory policies with service-level constraints. *Journal of the Operational Research Society* 40(1), 65–73.
- Agrawal, N. and S. A. Smith. 1996. Estimating negative binomial demand for retail inventory management with unobservable lost sales. *Naval Research Logistics* 43(6), 839–861.
- Axsäter, S. 1996. Using the deterministic EOQ formula in stochastic inventory control. *Management Science* 42(6), 830–834.
- Bagchi, U., J. Hayya, and J. K. Ord. 1984. Modeling demand during lead time. *Decision Sciences* 15, 157–176.
- Bagchi, U. and J. C. Hayya. 1984. Demand during lead time for normal unit demand and Erlang lead time. *Journal of the Operational Research Society* 35(2), 131–135.
- Bagchi, U., J. C. Hayya, and C.-H. Chu. 1986. The effect of lead-time variability: The case of independent demand. *Journal of Operations Management* 6(2), 159–177.
- Blackburn, J. 2012. Valuing time in supply chains: Establishing limits of time-based competition. *Journal of Operations Management* 30(5), 396–405.
- Bookbinder, J. and A. Lordahl. 1989. Estimation of inventory re-order levels using the bootstrap statistical procedure. *IIE Transactions* 21(4), 302–312.
- Boylan, J. E. and F. R. Johnston. 1994. Relationships between service level measures for inventory systems. *Journal of the Operational Research Society* 45(7), 838–844.
- Brown, R. G. 1977. *Materials Management Systems*. New York: Wiley-Interscience.
- Bulinskaya, E. V. 1990. Inventory control in case of unknown demand distribution. *Engineering Costs and Production Economics* 19, 301–306.
- Burgin, T. 1972. Inventory control with normal demand and gamma lead times. *Operational Research Quarterly* 23(1), 73–80.
- Burgin, T. A. 1975. The Gamma distribution and inventory control. *Operational Research Quarterly* 26(3), 507–525.

- Burgin, T. A. and J. M. Norman. 1976. A table for determining the probability of a stockout and potential lost sales for a gamma distributed demand. *Operational Research Quarterly* 27(3i), 621–631.
- Carlson, P. 1982. An alternate model for lead-time demand: Continuous-review inventory systems. *Decision Sciences* 13(1), 120–128.
- Chen, F. Y. and D. Krass. 2001. Inventory models with minimal service level constraints. *European Journal of Operational Research* 134(1), 120–140.
- Chen, F. and Y. Zheng. 1992. Waiting time distribution in (T,S) inventory systems. *Operations Research Letters* 12, 145–151.
- Chu, P., K.-L. Yang, S.-K. Liang, and T. Niu. 2004. Note on inventory model with a mixture of back orders and lost sales. *European Journal of Operational Research* 159(2), 470–475.
- Das, C. 1975. Effect of lead time on inventory: A study analysis. *Operational Research Quarterly* 26(2i), 273–282.
- Das, C. 1976. Approximate solution to the (Q, r) inventory model for gamma lead time demand. *Management Science* 22(9), 1043–1047.
- de Kok, A. G. 1990. Hierarchical production planning for consumer goods. *European Journal of Operational Research* 45, 55–69.
- De Treville, S., I. Bicer, V. Chavez-Demoulin, V. Hagspiel, N. Schürhoff, C. Tasserit, and S. Wager. 2014. Valuing lead time. *Journal of Operations Management* 32(6), 337–346.
- Disney, S. M., G. J. Gaalman, C. P. T. Hedenstierna, and T. Hosoda. 2015. Fill rate in a periodic review order-up-to policy under auto-correlated normally distributed, possibly negative, demand. *International Journal of Production Economics* 170, 501–512.
- Ehrhardt, R. 1979. The power approximation for computing (s, S) inventory policies. *Management Science* 25(8), 777–786.
- Eppen, G. D. and R. K. Martin. 1988. Determining safety stock in the presence of stochastic lead time and demand. *Management Science* 34(11), 1380–1390.
- Ernst, R. and D. Pyke. 1992. Component part stocking policies. *Naval Research Logistics* 39, 509–529.
- Federgruen, A. and Y.-S. Zheng. 1992. An efficient algorithm for computing an optimal (r, Q) policy in continuous review stochastic inventory systems. *Operations Research* 40(4), 808–813.
- Fortuin, L. 1980. The all-time requirement of spare parts for service after sales—Theoretical analysis and practical results. *International Journal of Operations and Production Management* 1(1), 59–70.
- Gallego, G. 1992. A minmax distribution free procedure for the (Q, R) inventory model. *Operations Research Letters* 11(1), 55–60.
- Gallego, G. and I. Moon. 1993. The distribution free newsboy problem: Review and extensions. *Journal of the Operational Research Society* 44(8), 825–834.
- Gardner, E. and D. Dannenbring. 1979. Using optimal policy surfaces to analyze aggregate inventory tradeoffs. *Management Science* 25(8), 709–720.
- Gardner, E. S. 1983. Approximate decision rules for continuous review inventory systems. *Naval Research Logistics Quarterly* 30, 59–68.
- Hadley, G. and T. M. Whitin. 1963. *Analysis of Inventory Systems*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Hausman, W. 1969. Minimizing customer line items backordered in inventory control. *Management Science* 15(12), 628–634.
- Herron, D. P. 1983. Improving productivity in logistics operations. In R. L. Schultz (Ed.), *Applications of Management Science*, pp. 49–85. Greenwich, CT: JAI Press.
- Heuts, R. and J. de Klein. 1995. An (s, q) inventory model with stochastic and interrelated lead times. *Naval Research Logistics* 42(5), 839–859.
- Hopp, W. J., R. Q. Zhang, and M. L. Spearman. 1999. An easily implementable hierarchical heuristic for a two-echelon spare parts distribution system. *IIE Transactions* 31(10), 977–988.
- Jain, A., H. Groenevelt, and N. Rudi. 2010. Continuous review inventory model with dynamic choice of two freight modes with fixed costs. *Manufacturing and Service Operations Management* 12(1), 120–139.
- Jain, A., H. Groenevelt, and N. Rudi. 2011. Periodic review inventory management with contingent use of two freight modes with fixed costs. *Naval Research Logistics* 58(4), 400–409.
- Johnson, M. E., H. Lee, T. Davis, and R. Hall. 1995. Expressions for item fill rates in periodic inventory systems. *Naval Research Logistics* 42(1), 57–80.

- Johnston, F. R. 1980. An interactive stock control system with a strategic management role. *Journal of the Operational Research Society* 31(12), 1069–1084.
- Karmarkar, U. 1994. A robust forecasting technique for inventory and lead time management. *Journal of Operations Management* 12(1), 45–54.
- Keaton, M. 1995. Using the Gamma distribution to model demand when lead time is random. *Journal of Business Logistics* 16(1), 107–131.
- Kottas, J. and H. S. Lau. 1980. The use of versatile distribution families in some stochastic inventory calculations. *Journal of the Operational Research Society* 31(5), 393–403.
- Kumar, A. and G. Sharman. 1992. We love your product, but where is it? *Sloan Management Review* 33(2), 93–99.
- Kumaran, M. and K. K. Achary. 1996. On approximating lead time demand distributions using the generalised l-type distribution. *Journal of the Operational Research Society* 47, 395–404.
- Lau, A. H.-L. and H.-S. Lau. 1993. A simple cost minimization procedure for the (Q, R) inventory model: Development and evaluation. *IIE Transactions* 25(2), 45–53.
- Lau, H.-S. and A. Zaki. 1982. The sensitivity of inventory decisions to the shape of lead time-demand distribution. *IIE Transactions* 14(4), 265–271.
- Lee, H. L., K. C. So, and C. S. Tang. 2000. The value of information sharing in a two-level supply chain. *Management Science* 46(5), 626–643.
- Lefrancois, P. 1989. Allowing for asymmetry in forecast errors: Results from a Monte-Carlo study. *International Journal of Forecasting* 5, 99–110.
- Leven, E. and A. Segerstedt. 2004. Inventory control with a modified Croston procedure and Erlang distribution. *International Journal of Production Economics* 90(3), 361–367.
- Lordahl, A. E. and J. H. Bookbinder. 1994. Order statistic calculation, costs, and service in an (s, Q) inventory system. *Naval Research Logistics* 41(1), 81–97.
- Magson, D. 1979. Stock control when the lead time cannot be considered constant. *Journal of the Operational Research Society* 30, 317–322.
- Mayer, R. M. 1984. Optimum safety stocks when demands and lead times vary. *Journal of Purchasing and Materials Management* 20(1), 27–32.
- McClain, J. O. and L. J. Thomas. 1980. *Operations Management: Production of Goods and Services*. Englewood Cliffs, NJ: Prentice-Hall.
- McFadden, F. R. 1972. On lead time demand distributions. *Decision Sciences* 3, 106–126.
- Moinzadeh, K. and S. Nahmias. 1988. A continuous review model for an inventory system with two supply modes. *Management Science* 34(6), 761–773.
- Moinzadeh, K. and C. P. Schmidt. 1991. An $(S - 1, S)$ inventory system with emergency orders. *Operations Research* 39(2), 308–321.
- Moon, I. and S. Choi. 1994. The distribution free continuous review inventory system with a service level constraint. *Computers and Industrial Engineering* 27(1–4), 209–212.
- Moon, I. and S. Choi. 1995. The distribution free newsboy problem with balking. *Journal of the Operational Research Society* 46, 537–542.
- Moors, J. J. A. and L. W. G. Strijbosch. 2002. Exact fill rates for (R, s, S) inventory control with gamma distributed demand. *Journal of the Operational Research Society* 53, 1268–1274.
- Naddor, E. 1978. Sensitivity to distributions in inventory systems. *Management Science* 24(16), 1769–1772.
- Nasri, F., J. F. Affisco, and M. J. Paknejad. 1990. Setup cost reduction in an inventory model with finite-range stochastic lead times. *International Journal of Production Research* 28(1), 199–212.
- Ng, K. Y. K. and M. N. Lam. 1998. Standardisation of substitutable electrical items. *Journal of the Operational Research Society* 49(9), 992–997.
- Paknejad, M. J., F. Nasri, and J. Affisco. 1992. Lead-time variability reduction in stochastic inventory models. *European Journal of Operational Research* 62, 311–322.
- Platt, D. E., L. W. Robinson, and R. B. Freund. 1997. Tractable (Q, R) heuristic models for constrained service levels. *Management Science* 43(7), 951–965.
- Presutti, V. and R. C. Trepp. 1970. More Ado about EOQ. *Naval Research Logistics Quarterly* 17(2), 243–251.

- Ray, D. L. 1981. Assessing U.K. manufacturing industry's inventory management performance. *Focus on PDM (Journal of the Institute of Physical Distribution Management)* 27, 5–11.
- Ray, S., Y. Gerchak, and E. M. Jewkes. 2004. The effectiveness of investment in lead time reduction for a make-to-stock product. *IIE Transactions* 36(4), 333–344.
- Ross, S. M. 1983. *Stochastic Processes*. New York: John Wiley & Sons.
- Rudi, N., H. Groenevelt, and T. R. Randall. 2009. End-of-period vs. continuous accounting of inventory-related costs. *Operations Research* 57(6), 1360–1366.
- Sahin, I. and D. Sinha. 1987. Renewal approximation to optimal order quantity for a class of continuous-review inventory models. *Naval Research Logistics* 34, 655–667.
- Scarf, H. 1958. A min-max solution of an inventory problem. In K. Arrow, S. Karlin, and H. Scarf (Eds.), *Studies in The Mathematical Theory of Inventory and Production*, pp. 201–209. California: Stanford University Press.
- Scarf, H. 1960. The optimality of (S, s) policies in the dynamic inventory problem. In K. J. Arrow, S. Karlin, and P. Suppes (Eds.), *Mathematical Methods in the Social Sciences*, pp. Chapter 13. Stanford: Stanford University Press.
- Schneider, H. 1981. Effect of service-levels on order-points or order-levels in inventory models. *International Journal of Production Research* 19(6), 615–631.
- Shore, H. 1986. General approximate solutions for some common inventory models. *Journal of the Operational Research Society* 37(6), 619–629.
- Shore, H. 1995. Fitting a distribution by the first two moments (partial and complete). *Computational Statistics and Data Analysis* 19, 563–577.
- Silver, E. 1970. Some ideas related to the inventory control of items having erratic demand patterns. *Canadian Operations Research Society Journal* 8(2), 87–100.
- Silver, E. A. 1976. Establishing the order quantity when the amount received is uncertain. *INFOR* 14(1), 32–39.
- Silver, E. A. 1977. A safety factor approximation based on Tukey's Lambda distribution. *Operational Research Quarterly* 28(3ii), 743–746.
- Silver, E. A. and M. R. Rahnema. 1986. The cost effects of statistical sampling in selecting the reorder point in a common inventory model. *Journal of the Operational Research Society* 37(7), 705–713.
- Silver, E. A. and M. R. Rahnema. 1987. Biased selection of the inventory reorder point when demand parameters are statistically estimated. *Engineering Costs and Production Economics* 12, 283–292.
- Snyder, R. D. 1980. The safety stock syndrome. *Journal of the Operational Research Society* 31(9), 833–837.
- Song, J. 1994a. The effect of leadtime uncertainty in a simple stochastic inventory model. *Management Science* 40(5), 603–613.
- Song, J. 1994b. Understanding the lead-time effects in stochastic inventory systems with discounted costs. *Operations Research Letters* 15, 85–93.
- Song, J.-S. 1998. On the order fill rate in a multi-item, base-stock inventory system. *Operations Research* 46(6), 831–845.
- Strijbosch, L. W. G. and R. Heuts. 1992. Modelling (s, Q) inventory systems: Parametric versus non-parametric approximations for the lead time demand distribution. *European Journal of Operational Research* 63, 86–101.
- Strijbosch, L. W. G. and R. Heuts. 1994. Investigating several alternatives for estimating the lead time demand distribution in a continuous review inventory model. *Kwantitatieve Methoden* 15(46), 57–75.
- Taylor, P. B. and K. H. Oke. 1976. Tables for stock control—Problems of formulation and computation. *Operational Research Quarterly* 27(3ii), 747–758.
- Taylor, P. B. and K. H. Oke. 1979. Explicit formulae for computing stock-control levels. *Journal of the Operational Research Society* 30(12), 1109–1118.
- Tijms, H. C. 1986. *Stochastic Modelling and Analysis: A Computational Approach*. New York: John Wiley & Sons.
- Tyworth, J. E. 1992. Modeling transportation–inventory trade-offs in a stochastic setting. *Journal of Business Logistics* 13(2), 97–124.
- Tyworth, J. E., Y. Guo, and R. Ganeshan. 1996. Inventory control under gamma demand and random lead time. *Journal of Business Logistics* 17(1), 291–304.

- Tyworth, J. E. and L. O'Neill. 1997. Robustness of the normal approximation of lead-time demand in a distribution setting. *Naval Research Logistics* 44(2), 165–186.
- Van Beek, P. 1978. An application of the logistic density on a stochastic continuous review stock control model. *Zeitschrift für Operations Research Band* 22, B165–B173.
- Van der Veen, B. 1981. Safety stocks—An example of theory and practice in OR. *European Journal of Operational Research* 6(4), 367–371.
- Van der Veen, B. 1984. Safety stocks and the paradox of the empty period. *European Journal of Operational Research* 16(1), 19–33.
- Veeraraghavan, S. and A. Scheller-Wolf. 2008. Now or later: A simple policy for effective dual sourcing in capacitated systems. *Operations Research* 56(4), 850–864.
- Wang, M. C. and S. S. Rao. 1992. Estimating reorder points and other management science applications by bootstrap procedure. *European Journal of Operational Research* 56, 332–342.
- Wang, P., W. Zinn, and K. L. Croxton. 2010. Sizing inventory when lead time and demand are correlated. *Production and Operations Management* 19(4), 480–484.
- Wu, J.-W. and H.-Y. Tsai. 2001. Mixture inventory model with back orders and lost sales for variable lead time demand with the mixtures of normal distribution. *International Journal of Systems Science* 32(2), 259–268.
- Zheng, Y. S. 1992. On properties of stochastic inventory systems. *Management Science* 38(1), 87–103.
- Zipkin, P. 1986. Inventory service level measures: Convexity and approximations. *Management Science* 32(8), 975–981.
- Zipkin, P. 1988. The use of phase-type distributions in inventory-control models. *Naval Research Logistics* 35, 247–257.

SPECIAL CLASSES OF ITEMS



In Section II, we were concerned with routine control of B items, that is, the bulk of the items in a typical population of stock-keeping units. Now, we turn our attention to special items, those at the ends of the dollar-volume (Dv) spectrum, namely, the A and C groups as well as those items not having a well-defined continuing demand pattern. The most important items, the A group, are handled in Chapter 7. Chapter 8 is concerned with slow-moving and low-dollar-volume items, the C group. Then, in Chapter 9, we deal with style goods and perishable items, that is, SKUs that can be maintained in inventory for only relatively short periods of time. The well-known newsvendor problem is covered in this chapter. The presentation of this portion of the book continues in the same vein as Section II; again, we are concerned primarily with managing individual-item inventories, although we do discuss several cases of coordinating across multiple items.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Chapter 7

Managing the Most Important Inventories

In this chapter, we devote our attention to the relatively small number of important items typically classified in the A group. For such items, additional effort may be warranted in both selecting control parameters and managing inventory. While not all important items are necessarily high-volume items, we restrict our attention in this chapter to items that are moderate to high volume. Chapter 8 addresses the management of slow-moving items of all levels of importance. Section 7.1 presents a brief review of the nature of A items. This is followed, in Section 7.2, by some general guidelines for the control of such items. Section 7.3 deals with faster-moving items and a more elaborate set of decision rules, which, in an (s, Q) system, take account of the service (shortage)-related effects in selecting the value of the order quantity Q .^{*} Next, in Sections 7.4 and 7.5, we present control logic for two somewhat more complicated systems whose use may be justified for A items. In Section 7.6, we address the issue of demand that is not stationary over time. Finally, in Section 7.7, we briefly discuss the research that analyzes the case of multiple suppliers.

7.1 Nature of Class A Items

Recall from Chapter 2 that class A items are defined to be those at the upper end of the spectrum of importance. That is, the total costs of replenishment, carrying stock, and shortages associated with such an item are high enough to justify a more sophisticated control system than those proposed for the less important B items in Chapters 4 through 6. This usually means a high annual dollar usage (a high value of Dv). By *usage* we mean retail customer demand as well as usage, say, of spare parts in production. In addition, one or more other factors may dictate that an item be placed in the A category. For example, an item, although its sales are relatively low, may be essential to round out the product line. To illustrate, a company may carry a high-end item for its prestige value. Poor customer service on such an item may have adverse effects on customers who buy large amounts

^{*} Equivalently, R in an (R, S) system.

of other products in the line. In essence, the important issue is a trade-off between control system costs (such as the costs of collecting required data, performing the computations, providing action reports, etc.) and the other three categories of costs mentioned above. (For simplicity, let us call this latter group “other” costs.) By definition, these “other” costs for an A item will be appreciably higher than for a B item. On the other hand, *given that the same control system is used* for an A and a B item, the control costs will be essentially the same for the two items. A numerical example is in order.

Consider two types of control systems: one simple, one complex. Suppose we have an A and a B item with estimated “other” costs using the simple system of \$900/year and \$70/year, respectively. It is estimated that the use of the complex control system will reduce these “other” costs by 10%. However, the control costs of this complex system are estimated to be \$40/year/item higher than those incurred in using the simple system. The following table illustrates that the complex system should be used for the A, but not for the B item.

<i>Using Complex System</i>			
<i>Item</i>	<i>Increase in Control Costs</i>	<i>Reduction in Other Costs</i>	<i>Net Increase in Costs</i>
A	40	$0.1(900) = 90$	-50
B	40	$0.1(70) = 7$	+33

As mentioned above, the factor Dv is important in deciding whether or not to put an item in the A category. However, the type of control to use within the A category should definitely depend on the magnitudes of the individual components, D and v . As we see later in this chapter, a high Dv resulting from a low D and a high v value (low unit sales of a high-value item) implies different control from a high Dv resulting from a high D and a low v value (high unit sales of a low-value item). To illustrate, we would certainly control the inventories of the following two items in different ways:

Item 1: Midamatic film processor, valued at \$30,000, which sells at a rate of 2 per year.

Item 2: Developing chemical, valued at \$2/liter, which sells at a rate of 30,000 liters per year.

Inventory control for items such as number 1 is addressed in Section 8.1, and for items such as number 2 is examined in Sections 7.3 through 7.5.

7.2 Guidelines for Control of A Items

In Section II, which was concerned primarily with B items, we advocated a management by exception approach. That is, most decisions would be made by routine (computerized or manual) rules. This should not be the case with A items. The potentially high payoff warrants frequent managerial attention to the replenishment decisions of individual items. Nonetheless, decision rules, based on mathematical models, do have a place in aiding the manager. Normally, these models cannot incorporate all the important factors. In such cases, the manager should incorporate subjective

factors that were omitted from the model. The art of management is very evident in this type of activity.

We now list a number of suggested guidelines for the control of A items. These may be above and beyond what normally can be incorporated in a usable mathematical decision rule:

1. *Inventory records should be maintained on a perpetual (transactions recording) basis, particularly for the more expensive items.* This need not be through the use of a computer system although this will be the most common setting. Even if computer systems are available for all items, our experience is that manual effort is still required to maintain system and data accuracy; thus, this effort may not be warranted for items of lesser importance.
2. *Keep top management informed.* Frequent reports (e.g., monthly) should be prepared for at least a portion of the A items. These should be made available to senior management for careful review.
3. *Estimate and influence demand.* This can be done in three ways:
 - a. Provide manual input to forecasts. For example, call your customers to get advance warning of their needs.
 - b. Ascertain the predictability of demand. Where the demand is of a special planned nature (e.g., scheduled overhaul of a piece of equipment) or where adequate warning of the need for a replacement part is given (e.g., a machine part fails but can be rapidly and inexpensively repaired to last for sufficient time to requisition and receive a new part from the supplier), there is no need to carry protection stock. In some cases, information regarding the impending need for replacement parts can be gathered by monitoring a system. Such information can be helpful in anticipating the need for replacement parts. Jardine et al. (2006) discuss such condition-based maintenance approaches. When the demand can occur without warning (i.e., a random breakdown), some protective stock may be appropriate. When the unit value of an item is extremely high, it may make sense to use a pool of spare parts *shared among several companies within the same industry*. An example is in the ammonia manufacturing industry where certain spare parts are maintained in a common pool.
 - c. Manipulate a given demand pattern. Seasonal or erratic fluctuations can sometimes be reduced by altering price structures, negotiating with customers, smoothing shipments, and so forth.
4. *Estimate and influence supply.* Again, it may not be advisable to passively accept a given pattern of supply. Negotiations (e.g., establishing blanket orders and freeze periods) with suppliers may reduce the average replenishment lead time, its variability, or both. The idea of a freeze period is that the timing or size of a particular order cannot be altered within a freeze period prior to its due date. In the same vein, it is very important to coordinate A item inventory requirements in-house with the production scheduling department.
5. *Use conservative initial provisioning.* For A items that have a very high v value and a relatively low D value, the initial provisioning decision becomes particularly crucial. For such items, erroneous initial overstocking (due to overestimating the usage rate) can be extremely expensive. Thus, it is a good idea to be conservative in initial provisioning.
6. *Review decision parameters frequently.* Generally, frequent review (as often as monthly or bimonthly) of order points and order quantities is advisable for A items.
7. *Determine precise values of control quantities.* In Chapters 4 and 8, we consider the use of tabular aids in establishing order quantities for B and C items. We argued that restricting attention to a limited number of possible time supplies (e.g., 1/2, 1, 2, 3, 6, and 12 months)

results in small cost penalties. This is not the case for A items. The percentage penalties may still be small, but even small percentage penalties represent sizable absolute costs. Therefore, order quantities of A items should be based on the most exact analysis possible.

8. *Confront shortages as opposed to setting service levels.* In Chapter 6, we pointed out that it is often very difficult to estimate the cost of shortages; as an alternative, we suggested specifying a desired level of customer service instead. In most cases, we do not advocate this approach for A items because of the typical management behavior in the short run in coping with potential or actual shortages of such items. The response is not usually to sit back passively and accept shortages of A items (or B items, for that matter). Instead, expediting (both in and out of house), emergency air freight shipments, and other actions are undertaken. That is, the shortages are avoided or eliminated quickly. Such actions, with associated costs that can be reasonably estimated, should be recognized when establishing safety stocks to prevent them. On the other hand, it should be pointed out that, in certain situations, customers are willing to wait a short time for delivery. Because A items are replenished frequently, it may be satisfactory to operate with very little on-hand stock (i.e., low safety stock) and, instead, backorder and satisfy such demand from the next order due in very soon.

7.3 Simultaneous Determination of s and Q for Fast-Moving Items

For fast-moving A items, we can usually rely on the normal distribution. We discuss the use of alternative distributions in Chapter 8 where we address slow-moving items. We covered sequential determination of s and Q in Chapter 6, noting there that the value of Q can often be constrained, or at least influenced, by minimum order sizes or predetermined lot sizes. For very important items, it may be valuable to simultaneously determine s and Q . In this section, we provide an example of simultaneous determination for the case of a B_1 shortage cost and normal lead time demand.*

First, we present the total cost equation for a general lead time demand distribution and a cost per stockout occasion:

$$\text{ETRC}(s, Q) = AD/Q + \left[Q/2 + \int_0^s (s - x_0) f_x(x_0) dx_0 \right] vr + \frac{DB_1}{Q} \int_s^\infty f_x(x_0) dx_0 \quad (7.1)$$

where the second term in the brackets is the safety stock, and the final term is the shortage cost. See the Appendix of Chapter 6 and Section 6.7.3 for more details. For the case of normal demand, we have an approximation

$$\text{ETRC}(k, Q) = AD/Q + (Q/2 + k\sigma_L)vr + \frac{DB_1}{Q} p_{u \geq}(k) \quad (7.2)$$

* Silver and Wilson (1972) develop similar results for a B_2 cost and for (R, S) systems. For P_2 , Yano (1985) treats the case of normal demand, and Moon and Choi (1994) addresses the case where we know only the mean and variance of demand. See also Hill (1992). See Problem 7.13 for the formula for the B_2 case. Finally, see Moon and Gallego (1994) for the \$/unit short case where we know only the mean and variance of demand.

where, as earlier,

$$s = \hat{x}_L + k\sigma_L$$

and $p_{u \geq}(k)$ is a function giving the probability that a unit normal variable takes on a value of k or larger. (As noted in Appendix III, $p_{u \geq}(k)$ is also available on many spreadsheets.)

7.3.1 Decision Rules

The derivation is shown Appendix of this chapter. All of the underlying assumptions discussed in Section 6.7.1 of Chapter 6 still apply, except that a value of Q is no longer prespecified. The two equations in the two unknowns Q and k (the latter being the safety factor) are

$$Q = \text{EOQ} \sqrt{\left[1 + \frac{B_1}{A} p_{u \geq}(k)\right]} \quad (7.3)$$

and

$$k = \sqrt{2 \ln \left[\frac{1}{2\sqrt{2\pi}} \left(\frac{B_1}{A}\right) \left(\frac{\sigma_L}{Q}\right) \left(\frac{\text{EOQ}}{\sigma_L}\right)^2 \right]} \quad (7.4)$$

where

- EOQ = the basic economic order quantity
- A = fixed cost component incurred with each replenishment
- D = demand rate
- v = unit variable cost of the item
- r = carrying charge
- σ_L = standard deviation of lead time demand
- $p_{u \geq}(k)$ = function of the unit normal variable

Note that Equation 7.4 is identical to the formula given by Equation 6.20 in Chapter 6:

$$k = \sqrt{2 \ln \left[\frac{DB_1}{\sqrt{2\pi} Q v \sigma_L r} \right]} \quad (7.5)$$

A suggested iterative solution procedure is to initially set $Q = \text{EOQ}$; then solve for a corresponding k value in Equation 7.4, and then use this value in Equation 7.3 to find a new Q value, and so forth. In certain cases, Equation 7.4 will have the square root of a negative number. In this case, k is set at its lowest allowable value, as previously discussed in Section 6.7.5, and then Q is obtained using this k value in Equation 7.3. Because of the convex nature of the functions involved, convergence to the true simultaneous solution pair (Q and k) is ensured. For further details concerning convexity, when the solution converges, and so on, refer to Chapter 4 of Hadley and Whitin (1963) as well as to Problem 7.3.

Using functions and equations from Appendix III, it is faster to set up columns for an iteration table. (See the table in the example below.) It is possible, however, simply to use two cells, one for k and one for Q . These are defined in terms of each other— k depends on Q , and Q depends on k . Spreadsheet programs will give a “circular references” warning, but it can be safely ignored. The spreadsheet will perform the number of iterations specified and converge on a solution. In all of our tests, this happened almost instantaneously. See Problem 7.9.

From Equation 7.3, we can see that the simultaneous Q is always larger than the EOQ. This makes sense in view of Equation 7.2. The EOQ is the Q value that minimizes the sum of the first two cost components. However, at that value, the last term in Equation 7.2 is still decreasing as Q increases. Thus, it pays to go somewhat above the EOQ.

Once we have ascertained that the simultaneous Q is larger than the EOQ, it follows from Equation 7.4 that the simultaneous k is smaller than or equal to the sequential k . (They may be equal because of the boundary condition of a lowest allowable k .) Again, this is intuitively appealing—the larger Q value makes stockout opportunities less frequent; hence we can afford to carry less safety stock.

7.3.1.1 Numerical Illustration

For illustrative purposes, suppose that a firm was considering the use of a B_1 shortage cost rule for certain items with a B_1 value of \$32 per shortage occasion. One of these items has the following characteristics:

- $D = 700$ containers/year
- $v = \$12$ /container
- $\hat{x}_L = 100$ containers
- $\sigma_L = 30$ containers

Using the values of $A = \$3.20$ and $r = 0.24$ /year, we have that

$$EOQ = \sqrt{\frac{2 \times 3.20 \times 700}{12 \times 0.24}} \approx 39.4$$

With the iterative procedure, we obtain

Iteration Number	1	2	3	4	5	6	7	8	9
Q	39.4	53.3	59.1	61.8	63.0	63.5	63.8	64.0	64.0
k	1.39	1.15	1.06	1.02	1.00	0.99	0.98	0.98	0.98

The order quantity (Q) is

$$Q = 64 \text{ containers}$$

and the reorder point (s) is

$$s = \hat{x}_L + k\sigma_L = 100 + 0.98(30) \approx 130 \text{ containers}$$

The total cost from Equation 7.2 is

$$(3.20)(700/64) + \left[\frac{64}{2} + (0.98)(30) \right] (12)(0.24) + \frac{700(32)}{64}(0.1635) = \$269.07$$

Robinson (1993) shows that under very general conditions the total relevant cost for the simultaneous solution case is $(Q + k\sigma_L)rv$. In this case, it is $[64 + (0.98)(30)](0.24)(12) = \269 .

7.3.2 Cost Penalties

The percentage penalty in costs (excluding those of the control system) resulting from using the simpler sequential procedure in lieu of the sophisticated simultaneous approach is given by

$$\text{PCP} = \frac{[\text{ETRC}(\text{sequential parameter values}) - \text{ETRC}(\text{simultaneous parameter values})]}{\text{ETRC}(\text{simultaneous parameter values})} \quad (7.6)$$

where PCP is the percentage cost penalty and the ETRC (·) values are expected total relevant costs (excluding system control costs) per unit time using the parameter values in the parentheses. Figure 7.1 shows PCP as a function of two dimensionless parameters in Equation 7.4, EOQ/σ_L and B_1/A (note again the appearance of the EOQ). The curves have been terminated to the left where the simultaneous solution first gives $k = 0$.

7.3.2.1 Numerical Illustration of the Cost Penalty

For the item shown above

$$B_1/A = 10$$

and

$$\text{EOQ}/\sigma_L = 1.3$$

Figure 7.1 shows the percentage cost penalty for the previous example is

$$\text{PCP} \approx 4.21\%$$

Now the sequential parameter values are quite easy to obtain. By the method of Section 6.7.5 $Q = \text{EOQ} = 39$ and k is selected to satisfy Equation 7.3:

$$\begin{aligned} k &= \sqrt{2 \ln \left[\frac{1}{2\sqrt{2\pi}} \left(\frac{B_1}{A} \right) \left(\frac{\text{EOQ}}{\sigma_L} \right) \right]} \\ &= \sqrt{2 \ln \left[\frac{1}{2\sqrt{2\pi}} (10)(1.3) \right]} \\ &= 1.39 \end{aligned}$$

Using Equation 7.2, we have

$$\text{ETRC}(1.39, 30) = \$280.41$$

The absolute savings is $\$280.41 - \$269.07 = \$11.33/\text{year}$. A savings of \$11.33 per year for an item that has an annual usage of \$8,400 seems very small. If, however, the 4.2% savings applies to many other A items, the total saved could be impressive. Therefore, it is worthwhile checking some other items for similar savings.

7.3.3 Further Comments

We have seen that the simultaneous approach requires an iterative solution of two complicated equations in two unknowns. (The situation is even more complicated for (R, S) systems. See

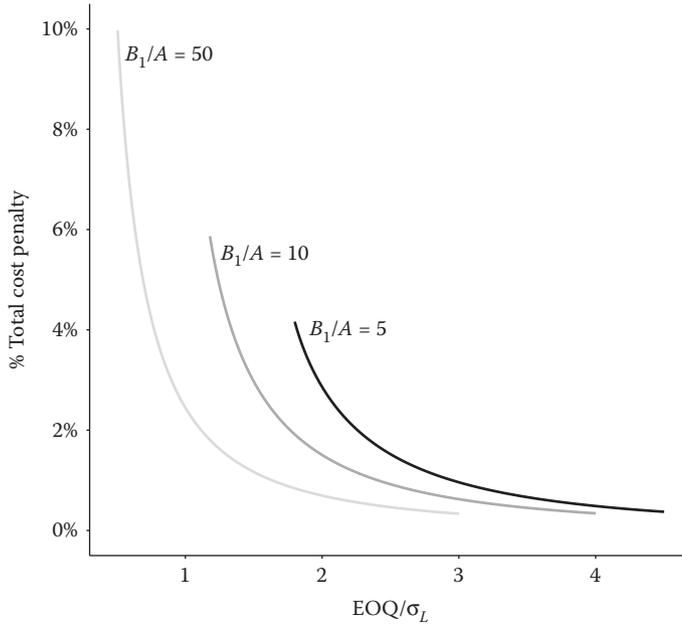


Figure 7.1 Percent cost penalty associated with using the sequential approach for an (s, Q) system for the case of a fixed cost per stockout occasion, B_1 .

Donaldson (1984) for a treatment of this case.) This amount of complexity should not be taken lightly when considering the possible use of a simultaneous solution. Even though the solution is very fast once the formulas are built into a spreadsheet, as noted above, we find that the complexity of the development of the formulas, and of the formulas themselves, generates resistance among some managers. A somewhat simpler approach is presented by Platt et al. (1997), who developed an accurate approximation for both s and Q when a fill rate is used. Their approximation can be built on a spreadsheet very easily, and is similar in development to the power approximation to be discussed in Section 7.5.1. When *the simultaneous approach is used*, we would suggest the use of a spreadsheet or another computerized routine. However, let us carefully examine the results of Figure 7.1. Notice that the percentage cost penalty, for a fixed value of B_1/A , decreases as EOQ/σ_L increases, that is, as the variability of demand decreases relative to the EOQ. This is to be expected because the sequential approach computes the order quantity ignoring the variability. Furthermore, the percentage penalties are quite low as long as EOQ/σ_L does not become too small.* Now, let us look more closely at the factor EOQ/σ_L .

$$\frac{EOQ}{\sigma_L} = \sqrt{\frac{2AD}{vr\sigma_L^2}} = \sqrt{\frac{2ADv}{r(\sigma_L v)^2}} \tag{7.7}$$

* Furthermore, Brown (1977) has suggested a pragmatic approach that eliminates the larger penalties. Specifically, if $EOQ < \sigma_L$, simply use $Q = \sigma_L$ and then find k sequentially.

As discussed in Chapter 3, some investigations have found evidence that, *on average*, $\sigma_L v$ tends to increase with Dv in the following manner:

$$\sigma_L v \approx c_1 (Dv)^{c_2}$$

where c_2 is greater than 0.5. Substituting this in Equation 7.7 yields

$$\frac{\text{EOQ}}{\sigma_L} \approx \sqrt{\frac{2ADv}{rc_1^2 (Dv)^{2c_2}}} \propto (Dv)^{0.5-c_2}$$

Because c_2 is larger than 0.5, this says that, on the average, EOQ/σ_L decreases as Dv increases; that is, low values of EOQ/σ_L are more likely with A items than with B or C items. From above, this implies that the larger percentage errors in using the sequential approach are more likely to occur with A items. This, together with the fact that the absolute costs associated with A items are high, indicates that the sophisticated simultaneous procedures may be justified for some A items. A corollary is that the absolute savings of a simultaneous approach are not likely to be justified for B items; hence, we suggested the use of the simpler sequential approach in Chapter 6.

In an interesting study, Gallego (1998) shows that using the following formula for the order size in a B_3 cost situation in conjunction with the optimal s yields a cost no more than 6.07% higher than the optimal. Therefore, a sequential approach may be used. His formula for the order quantity is

$$\text{EOQ}(1 + \sigma_L^2 v(r + B_3)/(2AD))^{0.25}$$

One final comment: If Q is small, as is common when pallet-sized quantities are delivered frequently, an order may not restore the inventory position to be greater than s . In such a case, consider an (s, nQ) system in which an order of a multiple of Q insures that the inventory position is greater than s at all times. See Zheng and Chen (1992) for a treatment of these policies.

7.4 Decision Rules for (s, S) Systems

We turn now to the commonly used (s, S) system. Once again we shall introduce a sequential approach followed by a simultaneous approach. Recall from Chapter 6 that an (s, S) system assumes continuous review. Whenever the inventory position drops to the reorder point s or lower, an order is placed to raise the position to the order-up-to level S . When all transactions are of unit size, then every order is of size $(S - s)$ and is placed when the inventory position is exactly at the level s . We can thus think of an order quantity

$$Q = S - s \tag{7.8}$$

and then the assumptions of an (s, Q) system (see Section 6.7.1 of Chapter 6) hold. Bartakke (1981) has reported on the use of such a system under Poisson (unit-sized) demand at Sperry-Univac. It is *non-unit-sized transactions that complicate the analysis*. Furthermore, in contrast to an (R, S) system where each cycle of fixed length R starts with the inventory position at the level S , here, we have a cycle of random length (how long it takes for the inventory position to drop from S to s or lower). Schneider (1979) shows that using an (s, Q) system in place of an (s, S) system can severely reduce the obtained service level.

We now discuss the sequential and simultaneous approaches for selecting s and S under conditions of non-unit-sized transactions. This is applied to one type of shortage-costing measure (cost

of B_1 per stockout occasion) and normally distributed forecast errors. Other references on (s, S) systems include Ward (1978); Archibald (1981); Williams (1982); Sahin (1982), and Federguen and Zipkin (1984). In addition, the related work of Tijms and Groenevelt (1984) will be discussed in Section 7.5.2.

7.4.1 Simple Sequential Determination of s and S

In this first approach, we choose to neglect the undershoots (how far below s the inventory position is located when an order is placed). In addition, S and s are computed in a sequential fashion as follows. The Q expression of Equation 7.8 is set equal to the economic order quantity as in Chapter 4. Then, given this value of Q , we find s by the procedure of Section 6.7.5 of Chapter 6. Finally, from Equation 7.8, the S value is given by

$$S = s + Q \quad (7.9)$$

7.4.1.1 Numerical Illustration

Consider an item XMF-014 that is a 1,000-sheet box of 8-by-10-inch rapid-process x-ray film. For illustrative purposes, suppose that this firm was considering using an (s, S) control system for this item. Relevant given characteristics of the item include

$$D = 1,400 \text{ boxes/year}$$

$$v = \$5.90/\text{box}$$

$$A = \$3.20$$

$$r = 0.24 \text{ \$/\$/year}$$

$$B_1 = \$150$$

$$\hat{x}_L = 270 \text{ boxes}$$

$$\sigma_L = 51.3 \text{ boxes}$$

The use of the above-outlined procedure leads to

$$Q = 80 \text{ boxes}$$

$$s = 389 \text{ boxes}$$

$$S = 469 \text{ boxes}$$

7.4.2 Simultaneous Selection of s and S Using the Undershoot Distribution

In the second approach, we select s and S simultaneously and also attempt to take account of the nonzero undershoots. A stockout occurs if the sum of the undershoot plus the total lead time demand exceeds s . Thus, the variable in which we are interested is

$$x' = z + x$$

where z is the undershoot and x is the total lead time demand.

We know the distribution of x (assumed to be normal). The distribution of z is quite complex, depending in general on the distance $S - s$ and the probability distribution of transaction sizes.

However, when $S - s$ is considerably larger than the average transaction size, we can make use of a result developed by Karlin (1958), namely,

$$p_z(z_0) = \frac{1}{E(t)} \sum_{t_0=z_0+1}^{\infty} p_t(t_0) \tag{7.10}$$

where

- $p_z(z_0)$ = probability that the undershoot is of size z_0
- $p_t(t_0)$ = probability that a demand transaction is of size t_0
- $E(t)$ = average transaction size

Equation 7.10 can be used to compute the mean and variance of the undershoot variable z . The results are*

$$E(z) = \frac{1}{2} \left[\frac{E(t^2)}{E(t)} - 1 \right]$$

and

$$\text{var}(z) = \frac{1}{12} \left\{ \frac{4E(t^3)}{E(t)} - 3 \left[\frac{E(t^2)}{E(t)} \right]^2 - 1 \right\}$$

The two variables z and x can be assumed to be independent.† Therefore

$$E(x') = E(z) + E(x) = \frac{1}{2} \left[\frac{E(t^2)}{E(t)} - 1 \right] + \hat{x}_L \tag{7.11}$$

and

$$\text{var}(x') = \text{var}(z) + \text{var}(x) = \frac{1}{12} \left\{ \frac{4E(t^3)}{E(t)} - 3 \left[\frac{E(t^2)}{E(t)} \right]^2 - 1 \right\} + \sigma_L^2 \tag{7.12}$$

For convenience, we assume that x' has a normal distribution‡ with the mean and variance given above. With this assumption, the method, whose derivation closely parallels that shown Appendix of this chapter, is as follows:

Step 1 Select k and Q to simultaneously satisfy the following two equations:

$$Q = \text{EOQ} \sqrt{1 + \frac{B_1}{A} p_{u \geq}(k) - E(z)}, \text{ rounded to an integer} \tag{7.13}$$

* Baganha et al. (1999) develop a simple algorithm to compute the undershoot. And Baganha et al. (1996) show that for most common demand distributions, the mean and variance given here are quite close to the actual values if $Q > 2E(t)$. See also Hill (1988) who examines the cost of ignoring the undershoot and develops an approximation for it.

† Janssen et al. (1998) show that it is better to measure both the undershoot and the lead time demand, and then use this data in setting reorder points.

‡ x is normally distributed but z is certainly not. Equation 7.10 reveals that $p_z(z_0)$ is monotonically decreasing with z_0 . Hence, the assumption of $(z + x)$ having a normal distribution is only an approximation.

and

$$k = \sqrt{2 \ln \left[\frac{1}{2\sqrt{2\pi}} \left(\frac{B_1}{A} \right) \frac{(\text{EOQ})^2}{(Q + E(z))\sigma_{x'}} \right]} \tag{7.14}$$

where

$p_{u \geq}(k)$ = probability that a unit normal variable takes on a value greater than or equal to k

EOQ = $\sqrt{2AD/vr}$ and $\sigma_{x'} = \sqrt{\text{var}(x')}$. Note the similarity of Equations 7.4 and 7.14.

Step 2

$$s = E(x') + k\sigma_{x'}, \text{ rounded to the nearest integer}$$

Step 3

$$S = s + Q$$

7.4.2.1 Numerical Illustration

We use the same item, XMF-014. Now, of course, we need a probability distribution of transaction sizes (or, alternatively, the first three moments $E(t)$, $E(t^2)$, and $E(t^3)$). Suppose for illustrative purposes that the distribution is

t_0	1	2	3	6	12	24	36	72
$p_t(t_0)$	0.25	0.05	0.05	0.1	0.25	0.15	0.1	0.05

$$E(t) = \sum_{t_0} t_0 p_t(t_0) = 14.9$$

$$E(t^2) = \sum_{t_0} t_0^2 p_t(t_0) = 515.7$$

$$E(t^3) = \sum_{t_0} t_0^3 p_t(t_0) = 25,857.2$$

From Equations 7.11 and 7.12

$$\begin{aligned} E(x') &= \frac{1}{2} \left(\frac{515.7}{14.9} - 1 \right) + 270 \\ &= 16.81 + 270 \\ &= 286.81 \end{aligned}$$

$$\begin{aligned} \text{var}(x') &= \frac{1}{12} \left\{ \frac{4(25,857.2)}{14.9} - 3 \left(\frac{515.7}{14.9} \right)^2 - 1 \right\} + (51.3)^2 \\ &= 278.90 + 2,631.69 \\ &= 2,910.59 \end{aligned}$$

Step 1 Equations 7.13 and 7.14 give that k and Q must satisfy

$$Q = 79.5 \sqrt{1 + 46.875 p_{u \geq}(k)} - 16.81$$

and

$$k = \sqrt{2 \ln \left[\frac{1}{2\sqrt{2\pi}} (46.875) \frac{(79.5)^2}{(Q + 16.81)(53.95)} \right]}$$

The solution is

$$Q = 87 \text{ boxes}$$

$$k = 2.17$$

Step 2

$$\begin{aligned} s &= 286.81 + 2.17\sqrt{2,910.59} \\ &= 404 \text{ boxes} \end{aligned}$$

Step 3

$$S = 404 + 87 = 491 \text{ boxes}$$

7.4.3 Comparison of the Methods

It is interesting to compare the results of the two methods for the numerical illustration of item XMF-014. Archibald (1976) has developed a search procedure that finds the optimal (s, S) pair for an arbitrary discrete distribution of transaction sizes and also gives the exact expected total relevant costs for any (s, S) pair used (see also Archibald and Silver 1978). Utilizing his procedure, we found the results shown in Table 7.1. Observe that the simultaneous method, which takes account of the undershoot, leads to a solution negligibly close to the optimum. Note that, even with the significant departure from unit-sized transactions shown in the numerical illustration, the sequential approach of Chapter 6 still gives a cost penalty of only 3.5%. Based on results of this type, we suggest the use of the simpler procedures of Chapter 6 unless the demand is erratic in nature (defined in Chapter 3 as when $\hat{\sigma}_1$ exceeds the demand level a , where both are measured over a unit time period). In addition, when a savings of 3%–4% represents a significant amount of money, one should use the simultaneous method or one of the more advanced procedures referenced earlier. Nevertheless, it is often worthwhile to check a small sample of items to see if the potential savings are greater.

Table 7.1 Illustrative Comparison of the Two Methods of Finding Values of s and S , Item XMF-014

<i>Method</i>	<i>Description</i>	<i>s</i>	<i>S</i>	<i>ETRC</i> (\$/Year)	<i>Percent above</i> <i>Best Value</i>
1	Sequential, ignoring undershoot	389	469	350.82	3.5
2	Simultaneous and using undershoot	404	491	339.01	0.0
	Optimal solution	406	493	339.00	–

7.5 Decision Rules for (R, s, S) Systems

As mentioned in Chapter 6, it has been shown that, under quite general conditions, the system that minimizes the total of review, replenishment, carrying, and shortage costs will be a member of the (R, s, S) family. However, the determination of the exact best values of the three parameters is extremely difficult. The problem results partly from the fact that undershoots of the reorder point s are present even if all transactions are unit-sized. This is because we only observe the inventory level every R units of time. Because of the complexity of exact analyses, we again advocate and illustrate the use of heuristic approaches. Some are simple enough that they could also be used for the control of B items. Hence, in Section 7.5.2, we demonstrate the use of a service constraint that would normally not be employed for A items. It is important to note, however, that Zheng and Federgruen (1991) have developed a fast algorithm for finding the optimal s and S for a given R and for discrete demand distributions. This algorithm is best developed using a programming language, or macros with a spreadsheet, rather than a simple spreadsheet. For applications in companies that can afford the programming time, this algorithm would be an excellent choice.

7.5.1 Decision Rule for a Specified Fractional Charge (B_3) per Unit Short at the End of Each Period

We now present a heuristic method, known as the revised power approximation. The original power approximation was suggested by Ehrhardt (1979) and the revision was proposed by Ehrhardt and Mosier (1984). An interesting alternative approach has been developed by Naddor (1975).

The revised power approximation determines values of the two control parameters $Q = S - s$ and s . We first show the approximation and then briefly discuss its nature and its derivation.

7.5.1.1 Revised Power Approximation Decision Rules

Step 1 Compute

$$Q_p = 1.30 \hat{x}_R^{0.494} \left(\frac{A}{vr} \right)^{0.506} \left(1 + \frac{\sigma_{R+L}^2}{\hat{x}_R^2} \right)^{0.116} \quad (7.15)$$

and

$$s_p = 0.973 \hat{x}_{R+L} + \sigma_{R+L} \left(\frac{0.183}{z} + 1.063 - 2.192z \right) \quad (7.16)$$

where

$$z = \sqrt{\frac{Q_p r}{\sigma_{R+L} B_3}} \quad (7.17)$$

$$\hat{x}_R = DR$$

$$\hat{x}_{R+L} = D(R + L)$$

with B_3 in \$/\$ short at the end of a review interval; r in \$/\$/review interval; D in units/year; and R and L in years.

Step 2 If $Q_p/\hat{x}_R > 1.5$, then let*

$$s = s_p \quad (7.18)$$

$$S = s_p + Q_p \quad (7.19)$$

Otherwise, go to Step 3.

Step 3 Compute

$$S_0 = \hat{x}_{R+L} + k\sigma_{R+L} \quad (7.20)$$

where k satisfies

$$p_{u \geq}(k) = \frac{r}{B_3 + r} \quad (7.21)$$

Then

$$s = \text{minimum}\{s_p, S_0\} \quad (7.22)$$

$$S = \text{minimum}\{s_p + Q_p, S_0\} \quad (7.23)$$

7.5.1.2 Numerical Illustration

Consider a particular item having $R = 1/2$ month and $L = 1$ month. Other relevant characteristics are $D = 1,200$ units/year, $A = \$25$, $v = \$2/\text{unit}$, $r = 0.24$ \$/\$/year, $B_3 = 0.2$ \$/\$ short/period, and $\sigma_{R+L} = 60$ units.

Step 1

$$\begin{aligned} \hat{x}_R &= RD = \frac{1}{24}(1,200) = 50 \text{ units} \\ \hat{x}_{R+L} &= (R + L)D = 150 \text{ units} \\ r &= \frac{0.24}{24} = 0.01 \text{ $/$/review interval} \end{aligned}$$

Then, from Equation 7.15, we have

$$\begin{aligned} Q_p &= 1.30(50)^{0.494}[25/(2.00)(0.01)]^{0.506}[1 + (60)^2/(50)^2]^{0.116} \\ &\approx 367 \text{ units} \end{aligned}$$

Next, Equation 7.17 gives

$$z = \sqrt{\frac{(367)(0.01)}{(60)(0.20)}} = 0.553$$

Then, the use of Equation 7.16 leads to

$$\begin{aligned} s_p &= 0.973(150) + 60 \left[\frac{0.183}{0.553} + 1.063 - 2.192(0.553) \right] \\ &\approx 157 \text{ units} \end{aligned}$$

* If demands are integer valued, then s and S are rounded to the nearest integer. There may also be a management-specified minimum value of s .

Step 2

$$Q_p/\hat{x}_R = 367/50 > 1.5$$

Therefore, from Equations 7.18 and 7.19, we have

$$s = 157 \text{ units}$$

$$S = 157 + 367 = 524 \text{ units}$$

7.5.1.3 Discussion

The rules were derived in the following fashion. Roberts (1962) developed analytic forms for Q and s that hold for large values of A and B_3 . Somewhat more general forms of the same nature were assumed containing a number of parameter values; for example,

$$Q_p = a\hat{x}_R^{1-b} \left(\frac{A}{vr}\right)^b \left(1 + \frac{\sigma_{R+L}^2}{\hat{x}_R^2}\right)^c \quad (7.24)$$

where a , b , and c are parameters.

Optimal values of $Q = S - s$ and s were found (by a laborious procedure not practical for routine operational use) for a wide range of representative problems, and the values of the parameters (e.g., a , b , and c in Equation 7.24) were determined by a regression fit to these optimal results. In the case of Q_p , this led to Equation 7.15. Equations 7.16 and 7.17 were found in much the same way.

Step 3 is necessary because, when Q_p/\hat{x}_R is small enough, Roberts' limiting policy is no longer appropriate; in effect, we are reduced to an order every review interval, that is, an (R, S) system. Equation 7.21 can be derived in the same fashion as Equation 6.23 of Section 6.7.7 of Chapter 6, except using the more accurate representation for the expected on-hand stock footnoted in Section 6.7.3.

Note that as $\sigma_{R+L} \rightarrow 0$, Equation 7.15 gives a result close to the basic economic order quantity. Extensive tests by Ehrhardt (1979) and Ehrhardt and Mosier (1984) have shown that the approximation performs extremely well in most circumstances.

There has been much research on (R, s, S) policies, including that of Zheng and Federgruen (1991). Porteus (1985) tests many methods for finding near-optimal (s, S) values with B_3 costing, including an older version of the power approximation. He also examines whether using the B_2 method described in Section 6.7.6 works well in the B_3 case. (Not surprisingly, it does not.) The paper provides an algorithm that works very well on a wide range of parameter values. See also Ehrhardt et al. (1981) for the case of autocorrelated demand, which might be observed by a warehouse that supplies retailers who themselves order using (s, S) policies. For a similar approximation for random lead times, see Ehrhardt (1984). See also Naddor (1978).

7.5.2 Decision Rule for a Specified Fraction (P_2) of Demand Satisfied Directly from Shelf

For the case of a fill rate constraint, we present a heuristic procedure (due to Tijms and Groenevelt 1984), which, incidentally, can be adapted for use within the (s, S) context of Section 7.4, as well as for the P_1 and TBS service measures. Schneider (1978, 1981) used a somewhat different approach to derive the same results. We show the decision rule, followed by a numerical illustration, and

then provide a brief discussion concerning the rule. The underlying assumptions are basically the same as those presented in Section 6.9.3 for (R, S) systems except, of course, now an order is placed at a review instant only if the inventory position is at or below s . We show illustrative details only for the case of normally distributed demand. This is appropriate to use as long as $(CV)_{R+L} \leq 0.5$ where $(CV)_{R+L} = \sigma_{R+L}/\hat{x}_{R+L}$ is the coefficient of variation of demand over $R + L$. When CV exceeds 0.5, a gamma distribution provides better results (see Tijms and Groenevelt 1984; and Section 6.7.14 of Chapter 6) because, with such a high value of CV , a normal distribution would lead to a significant probability of negative demands.

7.5.2.1 Decision Rule

Select s to satisfy

$$\sigma_{R+L}^2 J_u \left(\frac{s - \hat{x}_{R+L}}{\sigma_{R+L}} \right) - \sigma_L^2 J_u \left(\frac{s - \hat{x}_L}{\sigma_L} \right) = 2(1 - P_2) \hat{x}_R \left[S - s + \frac{\sigma_R^2 + \hat{x}_R^2}{2\hat{x}_R} \right] \quad (7.25)$$

where

$S - s$ = is assumed predetermined (e.g., by the EOQ)

\hat{x}_t = expected demand in a period of duration t

σ_t = standard deviation of errors of forecasts of total demand over a period of duration t

$J_u(k) = \int_k^\infty (u_0 - k)^2 f_u(u_0) du_0$ = another special function of the unit normal distribution

Note that, as shown by Hadley and Whitin (1963),

$$J_u(k) = (1 + k^2) p_{u \geq}(k) - k f_u(k) \quad (7.26)$$

so that $J_u(k)$ can be evaluated easily. We have developed a column in Table II.1 of Appendix II for $J_u(k)$ using the rational approximation described in Appendix III.

Equation 7.25 requires, in general, a trial-and-error-type solution. However, in the event that the desired service level is high enough ($P_2 \geq 0.9$), the demand pattern is relatively smooth and R is not too small compared with L ; then the second term on the left side of Equation 7.25 can be ignored. Thus, if we set

$$s = \hat{x}_{R+L} + k\sigma_{R+L} \quad (7.27)$$

the decision rule reduces to selecting k so as to satisfy

$$\sigma_{R+L}^2 J_u(k) = 2(1 - P_2) \hat{x}_R \left[S - s + \frac{\sigma_R^2 + \hat{x}_R^2}{2\hat{x}_R} \right] \quad (7.28)$$

If a table of $J_u(k)$ versus k is available, then finding the appropriate k poses no difficulty. Alternatively, one can use rational functions to find k from $J_u(k)$ as suggested by Schneider (1981) and presented in Appendix III. These are easily implemented on a spreadsheet. Using Appendix III, finding $J_u(k)$ from k is also easy on a spreadsheet.

7.5.2.2 Numerical Illustration

We illustrate a situation where the simpler Equation 7.28 can be used. Suppose a firm is controlling a particular item on a periodic basis with $R = 2$ weeks and the lead time $L = 1$ week. Assume

that $S - s$ has been preestablished as 100 units. The forecast update period is 2 weeks and current estimates are $\hat{x}_1 = 50$ units and $\sigma_1 = 15$ units, where the subscript, 1, indicates one forecast interval rather than 1 week. Suppose that $P_2 = 0.95$. First, we evaluate

$$\begin{aligned} \hat{x}_R &= \hat{x}_1 = 50 \text{ units} \\ \hat{x}_{R+L} &= \hat{x}_{3/2} = 75 \text{ units} \\ \sigma_R &= \sigma_1 = 15 \text{ units} \\ \sigma_{R+L} &= \sigma_{3/2} = \sqrt{3/2}\sigma_1 = 18.4 \text{ units} \\ (CV)_{R+L} &= \frac{\sigma_{R+L}}{\hat{x}_{R+L}} = \frac{18.4}{75} = 0.25 < 0.5 \end{aligned}$$

Thus, an assumption of normally distributed demand is reasonable. From Equation 7.28, we have

$$(18.4)^2 J_u(k) = 2(1 - 0.95)50 \left[100 + \frac{(15)^2 + (50)^2}{2(50)} \right]$$

or

$$J_u(k) = 1.885$$

Using Table II.1 or Appendix III, we find that

$$k \approx -0.98$$

Thus, from Equation 7.27, the reorder point is given by

$$\begin{aligned} s &= 75 - 0.98(18.4) \\ &\approx 57 \text{ units} \end{aligned}$$

Finally,

$$S = s + (S - s) = 57 + 100 = 157 \text{ units}$$

Incidentally, the use of the more complicated Equation 7.25 produces the same results in this example.

7.5.2.3 Discussion

To derive the decision rule, use, as in Chapter 6,

$$1 - P_2 = \frac{\text{Expected shortage per replenishment cycle}}{\text{Expected replenishment size}} \tag{7.29}$$

As in Section 7.4.2, the key random variable is the total of the undershoot z (of the reorder point) and the demand x in a lead time. The denominator of Equation 7.29 leads to the term in square brackets in Equation 7.25. The numerator in Equation 7.29 is developed based on

$$\int_s^\infty (x' - s) f_{x'}(x'_0) dx'_0$$

where $x' = z + x$. The result, after lengthy calculus, turns out to be the left side of Equation 7.25, divided by $2\hat{x}_R$.

For other research in this area, see Chen and Krass (1995) who discuss a “marginal service level” which is the minimum fraction of demand met from the shelf in *any period*. Thus, it is analogous to the fill rate (P_2) but it applies in any period rather than as a long-term average. They also describe analogies for the P_1 case. They show that (s, S) policies are optimal and provide a fast algorithm for finding the optimal values of s and S . Schneider and Ringuest (1990) present a power approximation for a *ready rate* service level. Also, for non-normal distributions, see Section 6.7.14 of Chapter 6. See also Schneider and Rinks (1989).

7.6 Coping with Nonstationary Demand

As discussed in Chapter 5, the average demand rate may vary appreciably with time. This should have an impact on the choice of control parameters such as s and Q for a continuous review system (or analogous parameters for periodic review systems). In some cases, nonstationary demand can be handled directly by the forecasting process from which we get the point forecast $\hat{x}_{t,t+L}$ and standard deviation of forecast errors $\sigma_{t,t+L}$ for demand over the next L periods.* Harrison (1967) presents expressions for the standard deviation of forecast errors over the lead time for both simple exponential smoothing and exponential smoothing with trend. It is worth emphasizing here that the degree of nonstationarity in the demand can have a significant impact on the necessary safety stock. As an example, consider a setting where demand is being forecasted with a simple exponential smoothing model. Furthermore, let σ represent the standard deviation of the one-period ahead forecast error and let α be the ideal smoothing constant for the forecasting process.† For a given safety factor k , the level of safety stock is

$$k\sigma_{t,t+L} = k\sigma\sqrt{L}\sqrt{1 + \alpha(L - 1) + \frac{\alpha^2(L - 1)(2L - 1)}{6}} \tag{7.30}$$

Note that if demand is stationary (e.g., the average level does not change over time), the ideal forecasting model would be to not smooth forecasts ($\alpha = 0$), and the level of safety stock becomes $k\sigma\sqrt{R + L}$; that is, the required level of safety stock is proportional to the square root of the protection interval $R + L$. When the underlying level of demand varies over time, the ideal smoothing constant α is greater than zero, and more safety stock is needed for the same protection level. This relationship, shown in Figure 7.2, is important to remember since often forecast errors are reported for a fixed time lag, such as one-period ahead, rather than over the lead time. For exponential smoothing with trend (Harrison 1967), where σ is the one-step ahead standard deviation of forecast errors and the ideal smoothing constants are α and β , the required level of safety stock is

$$k\sigma_{t,t+L} = k\sigma\sqrt{\sum_{j=0}^{L-1} [1 + j\alpha + j(j + 1)\beta/2]} \tag{7.31}$$

* The ideas presented here can also be applied to periodic review systems where $R + L$ would replace L as the required protection period.

† Graves (1999) analyzes a base-stock model for this setting.

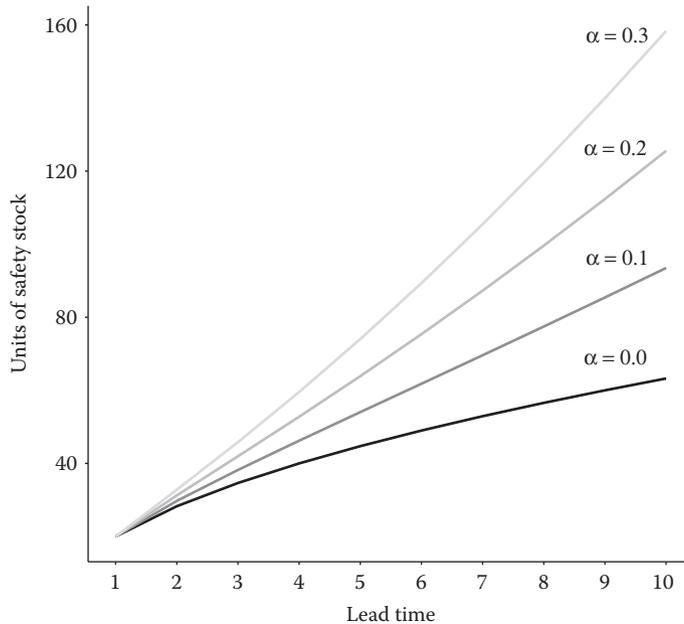


Figure 7.2 Necessary safety stock as a function of the protection period $R + L$ and the ideal smoothing constant α .

The above expressions can be used when demand is best forecasted with simple exponential smoothing or exponential smoothing with trend and one-period ahead forecast errors are measured and reported. An exact analysis of general time-varying and probabilistic demand is far too complicated for routine use in practice. Therefore, again we adopt heuristic approaches. If the average demand rate changes with time, then the values of the control parameters should also change with time. From a pragmatic standpoint, frequent changes in R , the review interval, for a periodic review system are awkward to implement. Thus, changes in R , if made at all, should be very infrequent. One appealing approach to adjusting s and Q (or s and S) with time is to compute these values in a particular period t using demand (forecast) information over the immediately following interval of duration L or $R + L$ as appropriate, but still using an underlying steady-state model. This is equivalent, in the terminology of Chapter 5, to using a rolling horizon of length $R + L$. Silver (1978) uses much the same idea but allows the demand rate to vary within the current horizon. This is accomplished by using safety stock considerations to decide when to place an order, followed by the deterministic Silver–Meal heuristic to select the size of the then current replenishment. Askin (1981) also addresses this problem. In a similar vein, Vargas (1994) develops a stochastic version of the Wagner–Whitin model described in Chapter 5. See also Vargas and Metters (1996).

One paper of interest is by Bollapragada and Morton (1993), who develop a myopic heuristic that is very fast and very accurate. Their heuristic involves precomputing the (s, S) values for various values of mean demand, possibly using methods of this chapter. Then they approximate the nonstationary problem by a stationary one—essentially, averaging the nonstationary demand over an estimate of the optimal time between orders (obtained from the stationary problem) and reading the corresponding (s, S) values from the precomputed table. Tests show that the error of using the heuristic averages only 1.7%. This heuristic is easy to implement, but should be avoided

if demand is expected to decline rapidly as a product becomes obsolete. Chapters 8 and 9 address this latter case.

Additional related work on this problem includes Eilon and Elmaleh (1970), Lasserre et al. (1985), LeFrancois and Cherkez (1987), Popovic (1987), Bookbinder and Tan (1988), Zipkin (1989), Wang and Rao (1992), Zinn et al. (1992), Charnes et al. (1995), Marmorstein and Zinn (1993), Cheng and Sethi (1999), Metters (1997), Ernst and Powell (1995), Morton and Pentico (1995), Chen and Krass (1995), and Anupindi et al. (1996).

A complexity to be faced under general, time-varying probabilistic demand, even in some of the simple procedures mentioned above, is the estimation of $\sigma_{t,t+R+L}$, the standard deviation of forecast errors for a forecast, made in period t , of demand over the next $R + L$ periods. In the exponential smoothing models discussed above, even though the level (and perhaps trend) of the demand are varying over time, the standard deviation of forecast errors is not. In general, and unlike in earlier discussions, we cannot make the assumption that $\sigma_{t,t+R+L} = \sigma_{R+L}$. Where the average demand rate changes with time, it is possible and perhaps likely that σ will also vary. A pragmatic approach to the estimation $\sigma_{t,t+R+L}$ is to use an aggregate relationship, first mentioned in conjunction with estimating σ 's of new items in Section 3.8.1 of Chapter 3, of the form

$$\sigma_{t,t+R+L} = c_1 \hat{x}_{t,t+R+L}^{c_2} \quad (7.32)$$

where $\hat{x}_{t,t+R+L}$ is the forecast, made in period t , of total demand over the next $R + L$ periods, and c_1 and c_2 are constants estimated for a population of items (as discussed in Chapter 4). Another approach, taken by Bollapragada and Morton (1993), is to assume a constant ratio of the standard deviation to the mean (i.e., $c_2 = 1$).

7.7 Comments on Multiple Sources of Supply and Expediting

Many firms are trying to reduce their number of suppliers because of the cost of managing multiple relationships. It takes time and effort to maintain good relationships with many different vendors. The buying firm must verify the quality of delivered parts, and often is required to work with the vendor to reduce the cost of those parts. Perhaps the most important factor in an industry that competes on time-to-market is the new product development cost. If the firm has multiple suppliers for a given part, the purchasing and engineering staffs must work with many vendors to insure that parts for the new products are of high quality and are ready on time. Nevertheless, from an inventory management perspective, costs can decrease if there are more suppliers. Most research in inventory theory disregards factors such as quality, time to market, and the cost of managing relationships. Unfortunately, in our discussions with managers, even at large and respected automotive firms, the inventory implications of sourcing decisions are often ignored. In this section, we comment briefly on research that should be a useful addition to decisions that are usually made on more qualitative grounds.

If multiple suppliers are used, the inventory equations developed in this chapter and in Chapter 6 must be modified. In general, the dual-sourcing models discussed here seek to balance the use of the two supply modes to get as much of the cost advantage from the slow source while keeping safety stock levels close to what could be achieved based on only the fast supply mode. There are two related challenges with dual-sourcing models. First, the resulting control policy should not be so complicated that it cannot be implemented in practice. Second, the analysis for dual-sourcing

systems is typically much more complicated, so one can expect greater effort to be required to identify good values for control parameters. Moinzadeh and Nahmias (1988) developed an approximate model for an inventory system with two possible suppliers, one having a shorter lead time. They defined a procedure for finding approximately optimal values of s and Q for both suppliers. (See also Moinzadeh and Schmidt 1991.) Kelle and Silver (1990) discussed order splitting among two or more vendors so that the average and variance of replenishment lead time are reduced. Lau and Zhao (1993) investigated a similar problem and discovered that the major advantage of order splitting is the reduction of cycle inventory cost. This result contrasts with earlier work that focused on reduction in safety stock. Guo and Ganeshan (1995) provided simple decision rules to find the optimal number of suppliers for particular lead time distributions. See also Sculli and Wu (1981), Ramasesh et al. (1991), Lau and Zhao (1994), Chiang and Benton (1994), Ernst et al. (2007), Fong and Gempesaw (1996), Parlar and Perry (1996), Zhang (1997), Chiang and Chiang (1996), and Feng and Shi (2012).

Veeraraghavan and Scheller-Wolf (2008) developed an approach for a relatively simple “dual index” policy for a periodic review system where orders for the two supply modes are determined based on two target inventory levels. At each ordering instance, two orders are placed: (1) an order with the fast lead time supply source to bring the total of on-hand and on-order inventory that will arrive within the fast lead time up to a specified target inventory level, and (2) an order with the slow lead time supplier to bring on-hand plus on-order inventory up to another (higher) target inventory level. While some computational effort is involved in finding the two control parameters, the authors demonstrate on a broad set of numerical experiments that the best dual-index policy is quite close to optimal. Their model can also accommodate capacity constraints on the supply sources.

Motivated by a setting where supply from China and Mexico was used to meet U.S. demand, Allon and Van Mieghem (2010) developed a “base-surge” approach to managing two supply sources where a constant “base” volume is sourced from the long lead time source and the fast lead time source is used to react to variations in demand (“surge”). For stationary demand, the authors developed an expression for the near-optimal long-term volume to be allocated to the slow source. This prescription may serve as a reasonable estimate even when demand is nonstationary. See also Boute and Van Mieghem (2014), Jain et al. (2010, 2011), and Sheopuri et al. (2010).

A related problem is when to place an emergency order if a replenishment order is late or cannot meet a surge in demand. White and White (1992) investigated the problem of setting the inventory policy for this case, and found that the cost of emergency ordering may be more than offset by reductions in inventory carrying costs. Dhakar et al. (1994) investigated an $(S - 1, S)$ inventory system for high-cost repairable items that has three possible modes of replenishment: normal repair orders, emergency repair orders, and expediting outstanding orders. They use a simulation procedure to find S and an expediting threshold. Kalpakam and Sapna (1995) addressed a lost sales inventory system in which orders are placed at two levels $s (> 0)$ and zero, with ordering quantities $Q > s$ and s , respectively. Finally, in an interesting paper, Song and Zipkin (1996) discussed a system in which replenishment lead times evolve over time as the system changes. They found that a longer lead time does not necessarily imply more inventory. A related work, Ben-Daya and Raouf (1994), addressed an inventory system in which lead time is a decision variable, which can be reduced at a cost. There is a significant amount of research on the inventory problem when the quantity supplied does not necessarily exactly match the quantity ordered. Costa and Silver (1996) have dealt with the situation where supply is uncertain, but different levels of supply can be reserved ahead of time. See also Lee and Moinzadeh (1987, 1989), Gerchak et al. (1988), Yano and Lee (1995), and Wang and Gerchak (1996).

7.8 Summary

In this chapter, we have dealt with the most important class of items, namely, A items. Procedures, more elaborate than those in earlier chapters, have been suggested. Where analytic models cannot incorporate all important factors and still be solvable by deductive reasoning, it may very well be worthwhile in the case of A items to resort to simulation methods that are much broader in applicability but more expensive to use. (See Section 2.7 of Chapter 2.)

In the next chapter, we turn to the opposite end of the spectrum, namely, C items. However, in later chapters, where we deal with coordination among items, A and B items will again play a prominent role.

Problems

- 7.1 Consider an (s, Q) system of control for a single item and with normally distributed forecast errors.
- Find an expression that k must satisfy given a Q value if we wish to have an expected number of stockout occasions per year equal to N .
 - Derive the two simultaneous equations that Q and k must satisfy if we wish to minimize the total expected costs of replenishment and carrying inventory subject to the expected number of stockout occasions per year being N .
 - For the following example, develop Q , k , and the associated total relevant costs per year for two strategies:
 - $Q = \text{EOQ}$ and corresponding k
 - Simultaneous best (Q, k) pair

Item characteristics:

$$\begin{array}{ll}
 A = \$5 & v = \$2/\text{unit} \\
 r = 0.16\$/\$/\text{year} & D = 1,000 \text{ units/year} \\
 \sigma_L = 80 \text{ units} & N = 0.5/\text{year}
 \end{array}$$

- 7.2 Ceiling Drug is concerned about the inventory control of an important item. Currently, they are using an (s, Q) system where Q is the EOQ and the safety factor k is selected based on a B_2v shortage-costing method. Relevant parameter values are estimated to be

$$\begin{array}{ll}
 D = 2,500 \text{ units/year} & B_2 = 0.6 \\
 v = \$10/\text{unit} & \hat{x}_L = 500 \text{ units} \\
 A = \$5 & \sigma_L = 100 \text{ units} \\
 r = 0.25\$/\$/\text{year} &
 \end{array}$$

- What are the Q and s values currently in use?
- Determine the simultaneous best values of Q and s .
- What is the percent penalty (in the total of replenishment, carrying, and shortage costs) of using the simpler sequential approach?

- 7.3 Consider a situation where the lowest allowable k value is 0. The B_1 shortage-costing method is appropriate.
- On a plot of k versus Q/σ_L sketch each of Equations 7.3 and 7.4. (Hint: For each, evaluate $dk/d(Q/\sigma_L)$. Also look at the behavior of each for limiting values of k and Q/σ_L .)
 - Attempt to ascertain under precisely what conditions there will be no simultaneous solution of the two equations.
 - Verify, using Equation 7.2, that under such circumstances the cost-minimizing solution is to set $k = 0$ and

$$Q/\sigma_L = \text{EOQ}/\sigma_L \sqrt{1 + 0.5B_1/A}$$

- 7.4 In relation to an (s, Q) inventory control system, an operations analyst has made the following observation:

Suppose we determine Q from the EOQ; then find the safety factor from the “fraction demand satisfied routinely from shelf” criterion. Suppose that through explicit cost considerations we estimate r to be a particular value r_1 . Now if we deliberately set r at a lower value r_2 , when determining Q and s , we can reduce actual total relevant costs.

- Could the analyst be telling the truth? Attempt to develop an analytic proof.
- Illustrate with an item having the following characteristics:

$$\begin{aligned} Dv &= \$4000/\text{year} & \sigma_L v &= \$500 \\ A &= \$100 & P_2 &= 0.975 \\ r_1 &= 0.2\$/\$/\text{year} \end{aligned}$$

- 7.5 Company X has an A item which it controls by an (s, S) system. The item has only a few customers. According to the (s, S) policy, if a large order drives the inventory position to or below s , enough is ordered to raise the position to S . Discuss the implicit cost trade-off that the company is making in still raising the position to S even after a large demand has just occurred.
- 7.6 a. Consider an item for which an (R, s, S) system is used. The review interval is 1 week. Other characteristics include

$$\begin{aligned} L &= 2 \text{ weeks} & r &= 0.26\$/\$/\text{year} \\ D &= 800 \text{ units/year} & B_3 &= 0.30\$/\$ \text{ short/week} \\ A &= \$20 & \sigma_{R+L} &= 14.2 \text{ units} \\ v &= \$1.50/\text{unit} \end{aligned}$$

Using the revised power approximation, find the appropriate values of s and S . (Assume that there are exactly 52 weeks per year.)

- Repeat for another item with the same characteristics except $v = \$150$ /unit.
- 7.7 An oil distribution depot supplies customers by road tanker. The depot itself is supplied by a unit train with 24 cars, each with a capacity of 75 tons. At the depot, three different products are stored in eight 1,000-ton tanks. At present, three tanks are used for product A, three tanks for product B, and two tanks for product C. Analysis of demand records shows that the average weekly demand for product A is 600 tons with a standard deviation 400

tons, average weekly demand for product B is 500 tons with a standard deviation 300 tons, and average weekly demand for product C is 400 tons with a standard deviation 200 tons. Any seasonal effects can be ignored. The stock level of each product is reported by the depot to the refinery every Monday morning. The distribution office then has to decide whether to replenish the depot the following Friday and, if so, how many cars of the train should be filled with each product. The cost of operating the train is not reduced if any cars are empty and is approximately \$3,000 per trip. If a product runs out at the depot, then the road tankers have to be diverted from that depot to another depot at considerably higher cost.

- a. How does the limited capacity tankage at the depot affect the applicability of standard inventory control models?
 - b. Are there any other features of the problem that make standard inventory control models difficult to apply?
 - c. *Without doing any detailed numerical calculations*, indicate the type of inventory control guidelines that you could propose for the distribution office.
 - d. *Briefly* indicate what other (if any) types of data you would need. Also, what type of solution procedure (for finding precise values of control parameters within the general control guidelines of part c) do you think would be needed?
- 7.8 A particular inventory control manager places orders for very expensive items. A report filed by an outside consultant included the following quotes. Comment on each.
- a. “He tried to order only enough items to meet actual sales for the 1-month period, one lead time period ahead.” Comment in particular the choice of a 1-month period.
 - b. “To meet unexpected demand he usually ordered 10%–20% more than the number of units he was able to identify through purchase orders or commitments.”
- 7.9 Develop a spreadsheet approach for Equations 7.3 and 7.4. First develop a table for iteration. Then use fewer cells by using circular references.
- 7.10 Repeat Problem 7.2 using $B_1 = \$100$ in place of $B_2 = 0.6$.
- 7.11 Consider the item in Problem 7.2 and assume that the sequential approach is used. Suppose that, through improved forecasting, σ_L can be reduced to 70 units. Up to what extra amount per year should the company be willing to pay to achieve this improved forecasting?
- 7.12 Air Alberta has been using an (s, Q) control system for a fast-moving spare part for one of its types of aircraft. The properties of this part are

$$\begin{aligned}
 D &= 1,920 \text{ units/year} & A &= \$15 \\
 v &= \$20.00/\text{unit} & r &= 0.2\$/\$/\text{year} \\
 \hat{x}_L &= 150 \text{ units} & \sigma_L &= 60 \text{ units}
 \end{aligned}$$

Whenever a shortage is impending, an expediting action is taken, avoiding the shortage. The cost of such an expedite is roughly \$200, independent of the magnitude of the impending shortage. The company uses an EOQ as the order quantity.

- a. Determine the EOQ and s .
- b. What are the expected total relevant costs (replenishing, carrying, and expediting) per year?
- c. A management science graduate, working in Air Alberta’s operations research department, says that increasing Q to 5% above the EOQ will reduce the expected total relevant costs per year? Can he be right? Discuss.

- 7.13 The two equations to solve simultaneously for the case of B_2 shortage cost for the (s, Q) system are

$$Q = \sqrt{\frac{2AD + 2DB_2v\sigma_L G_u(k)}{rv}}$$

and $p_{u \geq}(k) = \frac{Qr}{DB_2}$

For the following item, compute the optimal s and Q .

$$\begin{aligned} L &= 2 \text{ weeks} & r &= 0.26\$/\$/\text{year} \\ D &= 800 \text{ units/year} & B_2 &= 0.30 \\ A &= \$20 & \sigma_L &= 14.2 \text{ units} \\ v &= \$4.50/\text{unit} \end{aligned}$$

- 7.14 An arrangement, sometimes made between a supplier and a customer, is what is called a “standing” order. Under this arrangement, the supplier ships a fixed quantity Q_f at regular intervals (R time units apart) throughout the year. Because of the customer commitment for regular shipments, the supplier offers an attractive unit cost, v_f . However, because demand is variable, special auxiliary orders must be placed from time to time and/or one or more of the standard shipments may have to be postponed for R units of time. Suppose a particular steel products distributor, Salem Inc., is considering a standing order arrangement for a specific product with a supplier, Jericho Steel. The demand rate for this product averages 12,000 units/year. For a standing order arrangement of 1,000 units each month, Jericho offers a unit price (including shipment cost) of \$2.00/unit. Without a standing order arrangement, the unit price (including shipment cost) is \$2.20/unit. The Salem inventory manager estimates that the fixed cost associated with any standing order is \$25, whereas the fixed cost of any other order is \$35. Salem uses a carrying charge of 0.26 $\$/\$/\text{year}$. If a standing order arrangement is adopted, auxiliary orders can only be placed so as to arrive with standing orders (the lead time for such an auxiliary order is 1 month; in fact, this is the lead time of any nonstandard order). The decision to postpone a standard order also must be made 1 month before its scheduled arrival. Assume that demands in non-overlapping periods are independent, normally distributed variables with a standard deviation of 250 units for a 1 month period. Salem wishes to ensure a maximum probability of 0.03 of a stockout on any replenishment.
- If the standing order policy was adopted, indicate how you would assist Salem in deciding on
 - When to place an auxiliary order
 - When to postpone a standard order
 - Indicate how you might use a model to ascertain whether Salem should even adopt the standing order arrangement.

Appendix 7A: Simultaneous Solutions for Two Control Parameters

7A.1 (s, Q) System with B_1 Cost Penalty

We let

$$s = \hat{x}_L + k\sigma_L$$

From Equation 7.2, the expected total relevant costs per unit time under use of a safety factor k and an order quantity Q are given by

$$\text{ETRC}(k, Q) = AD/Q + (Q/2 + k\sigma_L)vr + \frac{DB_1}{Q}p_{u \geq}(k)$$

To reduce the number of separate parameters that have to be considered, we normalize by multiplying both sides by the constant term $2/vr\sigma_L$. We refer to the result as the normalized total relevant costs, $\text{NTRC}(k, Q)$. Thus,

$$\text{NTRC}(k, Q) = \frac{2AD}{vrQ\sigma_L} + \frac{Q}{\sigma_L} + 2k + \frac{2AD}{vrQ\sigma_L} \frac{B_1}{A} p_{u \geq}(k)$$

Recall that

$$\text{EOQ} = \sqrt{\frac{2AD}{vr}}$$

Therefore,

$$\text{NTRC}(k, Q) = \left(\frac{\text{EOQ}}{\sigma_L}\right)^2 \frac{\sigma_L}{Q} \left[1 + \frac{B_1}{A} p_{u \geq}(k)\right] + \frac{Q}{\sigma_L} + 2k$$

A necessary condition (unless we are at a boundary) for the minimization of a function of two variables is that the partial derivative with respect to each variable be set to zero.

$$\frac{\partial \text{NTRC}(k, Q)}{\partial Q} = -\left(\frac{\text{EOQ}}{\sigma_L}\right)^2 \frac{\sigma_L}{Q^2} \left[1 + \frac{B_1}{A} p_{u \geq}(k)\right] + \frac{1}{\sigma_L} = 0$$

or

$$\begin{aligned} \frac{Q}{\sigma_L} &= \frac{\text{EOQ}}{\sigma_L} \sqrt{\left[1 + \frac{B_1}{A} p_{u \geq}(k)\right]} \\ \frac{\partial \text{NTRC}(k, Q)}{\partial k} &= 2 - \left(\frac{\text{EOQ}}{\sigma_L}\right)^2 \frac{\sigma_L}{Q} \frac{B_1}{A} f_u(k) = 0 \end{aligned}$$

or

$$f_u(k) = 2 \frac{A}{B_1} \frac{Q}{\sigma_L} \left(\frac{\sigma_L}{\text{EOQ}}\right)^2$$

that is,

$$\frac{1}{\sqrt{2\pi}} \exp(-k^2/2) = 2 \frac{A}{B_1} \frac{Q}{\sigma_L} \left(\frac{\sigma_L}{\text{EOQ}}\right)^2$$

This simplifies to Equation 7.4.

Note: We have used the fact, proved in Appendix II, that

$$\frac{dp_{u \geq}(k)}{dk} = -f_u(k)$$

References

- Allon, G. and J. A. Van Mieghem. 2010. Global dual sourcing: Tailored base-surge allocation to near-and offshore production. *Management Science* 56(1), 110–124.
- Anupindi, R., T. E. Morton, and D. Pentico. 1996. The nonstationary stochastic lead-time inventory problem: Near-myopic bounds, heuristics, and testing. *Management Science* 42(1), 124–129.
- Archibald, B. 1976. *Continuous Review (s, S) Policies for Discrete, Compound Poisson, Demand Processes*. PhD thesis. University of Waterloo, Waterloo, Ontario, Canada.
- Archibald, B. and E. A. Silver. 1978. (s, S) Policies under continuous review and discrete compound Poisson demands. *Management Science* 24(9), 899–909.
- Archibald, R. 1981. Continuous review (s, S) policies with lost sales. *Management Science* 27(10), 1171–1177.
- Askin, R. G. 1981. A procedure for production lot sizing with probabilistic dynamic demand. *AIIE Transactions* 13(2), 132–137.
- Baganha, M. P., D. F. Pyke, and G. Ferrer. 1996. The undershoot of the reorder point: Tests of an approximation. *International Journal of Production Economics* 45, 311–320.
- Baganha, M. P., D. F. Pyke, and G. Ferrer. 1999. The residual life of the renewal process: A simple algorithm. *Naval Research Logistics* 46, 435–443.
- Bartakke, M. N. 1981. A method of spare parts inventory planning. *OMEGA* 9(1), 51–58.
- Ben-Daya, M. and A. Raouf. 1994. Inventory models involving lead time as a decision variable. *Journal of the Operational Research Society* 45(5), 579–582.
- Bollapragada, S. and T. Morton. 1993. *The Non-Stationary (s,S) Inventory Problem: Near-Myopic Heuristics, Computational Testing*. PhD thesis. Carnegie-Mellon. Pittsburgh, Pennsylvania, United States.
- Bookbinder, J. H. and J. Y. Tan. 1988. Strategies for the probabilistic lot-sizing problem with service-level constraints. *Management Science* 34(9), 1096–1108.
- Boute, R. N. and J. A. Van Mieghem. 2014. Global dual sourcing and order smoothing: The impact of capacity and lead times. *Management Science*. 61(9), 2080–2099.
- Brown, R. G. 1977. *Materials Management Systems*. New York: Wiley-Interscience.
- Charnes, J. M., H. Marmorstein, and W. Zinn. 1995. Safety stock determination with serially correlated demand in a periodic-review inventory system. *Journal of the Operational Research Society* 46(8), 1006–1013.
- Chen, F. Y. and D. Krass. 2001. Inventory models with minimal service level constraints. *European Journal of Operational Research* 134(1), 120–140.
- Cheng, F. and S. P. Sethi. 1999. Optimality of state-dependent (s, S) policies in inventory models with Markov-modulated demand and lost sales. *Production and Operations Management* 8(2), 183–192.
- Chiang, C. and W. Benton. 1994. Sole sourcing versus dual sourcing under stochastic demands and lead times. *Naval Research Logistics* 41(5), 609–624.
- Chiang, C. and W. Chiang. 1996. Reducing inventory costs by order splitting in the sole sourcing environment. *Journal of the Operational Research Society* 47, 446–456.
- Costa, D. and E. A. Silver. 1996. Exact and approximate algorithms for the multi-period procurement problem where dedicated supplier capacity can be reserved. *Operations-Research-Spektrum* 18(4), 197–207.
- Dhakar, T. S., C. P. Schmidt, and D. M. Miller. 1994. Base stock level determination for high cost low demand critical repairable spares. *Computers and Operations Research* 21(4), 411–420.
- Donaldson, W. A. 1984. An equation for the optimal value of T, the inventory replenishment review period when demand is normal. *Journal of the Operational Research Society* 35(2), 137–139.

- Ehrhardt, R. 1979. The power approximation for computing (s, S) inventory policies. *Management Science* 25(8), 777–786.
- Ehrhardt, R. 1984. (s, S) Policies for a dynamic inventory model with stochastic lead times. *Operations Research* 32(1), 121–132.
- Ehrhardt, R. and C. Mosier. 1984. Revision of the power approximation for computing (s, S) policies. *Management Science* 30(5), 618–622.
- Ehrhardt, R. A., C. R. Schultz, and H. M. Wagner. 1981. (s, S) Policies for a wholesale inventory system. In L. B. Schwarz (Ed.), *Multi-Level Production/Inventory Control Systems: Theory and Practice*, Volume 16, pp. 145–161. Amsterdam: North-Holland.
- Eilon, S. and J. Elmaleh. 1970. Adaption limits in inventory control. *Management Science* 16(18), B533–B548.
- Ernst, R., B. Kamrad, and K. Ord. 2007. Delivery performance in vendor selection decisions. *European Journal of Operational Research* 176(1), 534–541.
- Ernst, R. and S. G. Powell. 1995. Optimal inventory policies under service-sensitive demand. *European Journal of Operational Research* 87, 316–327.
- Federgruen, A. and P. Zipkin. 1984. An efficient algorithm for computing optimal (s, S) policies. *Operations Research* 32(6), 1268–1285.
- Feng, Q. and R. Shi. 2012. Sourcing from multiple suppliers for price-dependent demands. *Production and Operations Management* 21(3), 547–563.
- Fong, D. K. H. and V. M. Gempesaw. 1996. Evaluating re-order levels and probabilities of stockout during a lead time for some stock-control models. *Journal of the Operational Research Society* 47, 321–328.
- Gallego, G. 1998. New bounds and heuristics for (Q, r) policies. *Management Science* 44(2), 219–233.
- Gerchak, Y., R. Vickson, and M. Parlar. 1988. Periodic review production models with variable yield and uncertain demand. *IIE Transactions* 20(2), 144–156.
- Graves, S. C. 1999. A single-item inventory model for a nonstationary demand process. *Manufacturing & Service Operations Management* 1(1), 50–61.
- Guo, Y. and R. Ganeshan. 1995. Are more suppliers better? *Journal of the Operational Research Society* 46, 892–895.
- Hadley, G. and T. M. Whitin. 1963. *Analysis of Inventory Systems*. Englewood Cliffs, NJ: Prentice-Hall, Inc.
- Harrison, P. J. 1967. Exponential smoothing and short-term sales forecasting. *Management Science* 13(11), 821–842.
- Hill, R. M. 1988. The effect of transfer-pack size on order quantity. *Journal of the Operational Research Society* 39(3), 305–309.
- Hill, R. M. 1992. Numerical analysis of a continuous-review lost-sales inventory model where two orders may be outstanding. *European Journal of Operational Research* 62, 11–26.
- Jain, A., H. Groenevelt, and N. Rudi. 2010. Continuous review inventory model with dynamic choice of two freight modes with fixed costs. *Manufacturing & Service Operations Management* 12(1), 120–139.
- Jain, A., H. Groenevelt, and N. Rudi. 2011. Periodic review inventory management with contingent use of two freight modes with fixed costs. *Naval Research Logistics* 58(4), 400–409.
- Janssen, F., R. Heuts, and A. de Kok. 1998. On the (R, s, Q) inventory model when demand is modelled as a compound Bernoulli process. *European Journal of Operational Research* 104(3), 423–436.
- Jardine, A. K., D. Lin, and D. Banjevic. 2006. A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing* 20(7), 1483–1510.
- Kalpakam, S. and K. P. Sapna. 1995. A two reorder level inventory system with renewal demands. *European Journal of Operational Research* 84, 402–415.
- Karlin, S. 1958. The application of renewal theory to the study of inventory policies. In K. Arrow, S. Karlin, and H. Scarf (Eds.), *Studies in the Mathematical Theory of Inventory and Production*, Chapter 15. Stanford, CA: Stanford University Press.
- Kelle, P. and E. A. Silver. 1990. Safety stock reduction by order splitting. *Naval Research Logistics* 37, 725–743.
- Lasserre, J. B., C. Bes, and F. Roubellat. 1985. The stochastic discrete dynamic lot size problem: An open-loop solution. *Operations Research* 33(3), 684–689.
- Lau, H. and L. Zhao. 1994. Dual sourcing cost-optimization with unrestricted lead-time distributions and order-split proportions. *IIE Transactions* 26(5), 66–75.

- Lau, H.-S. and L.-G. Zhao. 1993. Optimal ordering policies with two suppliers when lead times and demands are all stochastic. *European Journal of Operational Research* 68, 120–133.
- Lee, H. L. and K. Moynzadeh. 1987. Two-parameter approximations for multi-echelon repairable inventory models with batch ordering policy. *IIE Transactions* 19(2), 140–149.
- Lee, H. L. and K. Moynzadeh. 1989. A repairable item inventory system with diagnostic and repair service. *European Journal of Operational Research* 40, 210–221.
- LeFrancois, P. and C. Cherkez. 1987. Adaptive limits for pc-based inventory control. *International Journal of Production Research* 25(9), 1325–1337.
- Marmorstein, H. and W. Zinn. 1993. A conditional effect of autocorrelated demand on safety stock determination. *European Journal of Operational Research* 68(1), 139–142.
- Metters, R. 1997. Production planning with stochastic seasonal demand and capacitated production. *IIE Transactions* 29(11), 1017–1029.
- Moynzadeh, K. and S. Nahmias. 1988. A continuous review model for an inventory system with two supply modes. *Management Science* 34(6), 761–773.
- Moynzadeh, K. and C. P. Schmidt. 1991. An $(S - 1, S)$ inventory system with emergency orders. *Operations Research* 39(2), 308–321.
- Moon, I. and S. Choi. 1994. The distribution free continuous review inventory system with a service level constraint. *Computers and Industrial Engineering* 27(1–4), 209–212.
- Moon, I. and G. Gallego. 1994. Distribution free procedures for some inventory models. *Journal of Operational Research Society* 45(6), 651–658.
- Morton, T. E. and D. W. Pentico. 1995. The finite horizon nonstationary stochastic inventory problem: Near myopic bounds, heuristics, testing. *Management Science* 41(2), 334–343.
- Naddor, E. 1975. Optimal heuristic decisions for the s, S inventory policy. *Management Science* 21(9), 1071–1072.
- Naddor, E. 1978. Sensitivity to distributions in inventory systems. *Management Science* 24(16), 1769–1772.
- Parlar, M. and D. Perry. 1996. Inventory models of future supply uncertainty with single and multiple suppliers. *Naval Research Logistics* 43, 191–210.
- Platt, D. E., L. W. Robinson, and R. B. Freund. 1997. Tractable (Q, R) heuristic models for constrained service levels. *Management Science* 43(7), 951–965.
- Popovic, J. B. 1987. Decision making on stock levels in cases of uncertain demand rate. *European Journal of Operational Research* 32, 276–290.
- Porteus, E. L. 1985. Investing in reducing setups in the EOQ model. *Management Science* 31(8), 998–1010.
- Ramasesh, R. V., J. K. Ord, J. C. Hayya, and A. Pan. 1991. Sole versus dual sourcing in stochastic lead-time (s, Q) inventory models. *Management Science* 37(4), 428–443.
- Roberts, D. 1962. Approximations to optimal policies in a dynamic inventory model. In K. Arrow, S. Karlin, and H. Scarf (Eds.), *Studies in Applied Probability and Management Science*, pp. 207–229. Stanford, CA: Stanford University Press.
- Robinson, L. 1993. The cost of following the optimal inventory policy. *IIE Transactions* 25(5), 105–108.
- Sahin, I. 1982. On the objective function behavior in (s, S) inventory models. *Operations Research* 30(4), 709–724.
- Schneider, H. 1978. Methods for determining the re-order point of an (s, S) ordering policy when a service level is specified. *Journal of the Operational Research Society* 29(12), 1181–1193.
- Schneider, H. 1979. The service level in inventory control systems. *Engineering and Process Economics* 4, 341–348.
- Schneider, H. 1981. Effect of service-levels on order-points or order-levels in inventory models. *International Journal of Production Research* 19(6), 615–631.
- Schneider, H. and J. L. Ringuest. 1990. Power approximation for computing (s, S) policies using service level. *Management Science* 36(7), 822–834.
- Schneider, H. and D. Rinks. 1989. Optimal policy surfaces for a multi-item inventory problem. *European Journal of Operational Research* 39, 180–191.
- Sculli, D. and S. Y. Wu. 1981. Stock control with two suppliers and normal lead times. *Journal of the Operational Research Society* 32(11), 1003–1009.

- Sheopuri, A., G. Janakiraman, and S. Seshadri. 2010. New policies for the stochastic inventory control problem with two supply sources. *Operations Research* 58(3), 734–745.
- Silver, E. A. 1978. Inventory control under a probabilistic time-varying, demand pattern. *AIIE Transactions* 10(4), 371–379.
- Silver, E. A. and T. G. Wilson. 1972. Cost penalties of simplified procedures for selecting reorder points and order quantities. In *Fifteenth Annual International Conference of the American Production and Inventory Control Society*, Toronto, Canada, pp. 219–234.
- Song, J. and P. H. Zipkin. 1996. Evaluation of base-stock policies in multiechelon inventory systems with state-dependent demands. Part II: State-dependent depot policies. *Naval Research Logistics* 43, 381–396.
- Tijms, H. and H. Groenevelt. 1984. Simple approximations for the reorder point in periodic and continuous review (s, S) inventory systems with service level constraints. *European Journal of Operational Research* 17, 175–190.
- Vargas, V. A. 1994. *The Stochastic Version of the Wagner-Whitin Production Lot-Size Model*. PhD thesis. Emory University. Atlanta, Georgia.
- Vargas, V. A. and R. Metters. 1996. Adapting lot-sizing techniques to stochastic demand through production scheduling policy. *IIE Transactions* 28, 141–148.
- Veeraraghavan, S. and A. Scheller-Wolf. 2008. Now or later: A simple policy for effective dual sourcing in capacitated systems. *Operations Research* 56(4), 850–864.
- Wang, M. C. and S. S. Rao. 1992. Estimating reorder points and other management science applications by bootstrap procedure. *European Journal of Operational Research* 56, 332–342.
- Wang, Y. and Y. Gerchak. 1996. Periodic review production models with variable capacity, random yield, and uncertain demand. *Management Science* 42(1), 130–137.
- Ward, J. B. 1978. Determining reorder points when demand is lumpy. *Management Science* 24(6), 623–632.
- White, C. R. and B. R. White. 1992. Emergency inventory ordering assuming Poisson demands. *Production Planning and Control* 3(2), 168–174.
- Williams, B. W. 1982. Vehicle scheduling: Proximity priority searching. *Journal of the Operational Research Society* 33(10), 961–966.
- Yano, C. A. 1985. New algorithms for (Q, r) systems with complete backordering using a fill-rate criterion. *Naval Research Logistics Quarterly* 32(4), 675–688.
- Yano, C. A. and H. L. Lee. 1995. Lot sizing with random yields: A review. *Operations Research* 43(2), 311–334.
- Zhang, A. X. 1997. Demand fulfillment rates in an assemble to-order system with multiple products and dependent demands. *Production and Operations Management* 6(3), 309–324.
- Zheng, Y. and F. Chen. 1992. Inventory policies with quantized ordering. *Naval Research Logistics* 39, 285–305.
- Zheng, Y. S. and A. Federgruen. 1991. Finding optimal (s, S) policies is about as simple as evaluating a single policy. *Operations Research* 39(4), 654–665.
- Zinn, W., H. Marmorstein, and J. Charnes. 1992. The effect of autocorrelated demand on customer service. *Journal of Business Logistics* 13(1), 173–192.
- Zipkin, P. 1989. Critical number policies for inventory models with periodic data. *Management Science* 35(1), 71–80.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Chapter 8

Managing Slow-Moving and Low-Value (Class C) Inventories

Slow-moving and low-value items often fall in the C category or so-called “cats and dogs,” although as noted in Chapter 7, slow-moving items may be important enough to be included in the A or B category. Collectively, these items usually represent an appreciable percentage of the total number of distinct SKUs, but a small fraction of the total dollar investment in the inventory of a typical company. Most of these items taken singly are relatively unimportant, but, because of their large numbers, appropriate simple control procedures must be used. In this chapter, we discuss such methods.

In Section 8.1, decision rules are presented for important (e.g., high-value or critical) slow-moving items using (s, Q) systems. The special case of intermittent demand for high-unit-value items is discussed in Section 8.2. Section 8.3 spells out in greater detail the nature of C items and the primary objectives in designing methods for controlling them. In Section 8.4, we propose control procedures for low-value items having relatively steady demand. Specifically, we consider, in order, the type of inventory records to use, the selection of the reorder quantity (or reorder interval), and the choice of the reorder point (or order-up-to level). Next, in Section 8.5, we treat the case of items with significantly declining usage patterns, that is, approaching the termination of usage. Section 8.6 is concerned with the important issue of removing totally (or relatively) inactive items from the organization’s inventory. Finally, in Section 8.7, we address the related question of whether demands for a particular item should be purchased (made) to order or met from stock.

8.1 Order-Point, Order-Quantity (s, Q) Systems for Slow-Moving A Items

In Chapter 6, for the case of B items, we argued for the almost exclusive use of the normal distribution to represent the distribution of forecast errors over a lead time. For A items, where the

potential benefits of using a more accurate representation are higher, we suggest, based on extensive tests (see Tyworth and O'Neill 1997), still using the normal distribution if the average (or forecast) demand in a lead time is high enough (at least 10 units). As mentioned in Chapter 6, we advocate the use of explicit (or implicit) costing of shortages. Therefore, decision rules such as those in Sections 6.7.5 through 6.7.8 of Chapter 6 should be used. If the average, \hat{x}_L , is below 10 units, then a discrete distribution, such as the Poisson, is more appropriate. (See the Appendix of this chapter for a discussion of some properties of the distribution and a derivation of expected units short per cycle. See Appendix III for details on how to find expected units short per cycle using common spreadsheet functions.)

The Poisson distribution has but a single parameter, namely, the average demand (in this case, \hat{x}_L). Once \hat{x}_L is specified, a value of the standard deviation of forecast errors, σ_L , follows from the Poisson relation

$$\sigma_L = \sqrt{\hat{x}_L} \quad (8.1)$$

Therefore, the Poisson is appropriate to use only when the observed σ_L (for the item under consideration) is quite close to $\sqrt{\hat{x}_L}$. An operational definition of “quite close” is within 10% of $\sqrt{\hat{x}_L}$. When this is not the case, somewhat more complicated discrete distributions such as the negative binomial (see Taylor 1961; Porteus 1985; Aggarwal and Moinzadeh 1994) or a compound Poisson (see, e.g., Adelson 1966; Feeney and Sherbrooke 1966; Mitchell et al. 1983) should be considered.

The discrete nature of the Poisson distribution (in contrast, the normal distribution deals with a continuous variable) is both a blessing and a curse. For slow-moving items, it is important to be able to deal with discrete units. On the other hand, discrete mathematics create problems in implementation, as we will soon see. In most of our discussions in Chapter 6, there were two control parameters per item:

- s and Q in an (s, Q) system
- R and S in an (R, S) system

Our approach was first to determine one of the parameters—namely, Q or R ; then (sequentially) to find the best value of the second parameter, s or S , conditional on the value of Q or R , respectively. The derivation of the value of Q or R ignored the effect that this parameter has on the service level or shortage costs per unit time. To illustrate, consider the (s, Q) control system with a fixed cost B_1 per stockout occasion. From Section 6A.1 of the Appendix of Chapter 6 and Section 6.7.5, the total relevant costs per year, ETRC, are

$$AD/Q + (Q/2)vr + (SS)vr + \frac{DB_1}{Q} \times (\text{Probability of a stockout in a cycle}) \quad (8.2)$$

where SS is the safety stock.

The approach, advocated in Chapter 6, was first to pick the Q value that minimizes the first two terms, ignoring the last two terms; then to choose the best value of safety stock (or $k\sigma_L$ in the case of normal demand). But, clearly, the value of Q influences the shortage costs (the last term). The larger the Q is, the smaller these costs become. The exact approach would be to simultaneously determine

* In an (R, S) system \hat{x}_L , the average demand in a lead time would be replaced by \hat{x}_{R+L} , the average demand in a review interval plus a lead time.

SS and Q to minimize the ETRC expression.* As we saw in Chapter 7, this exact approach is more involved from both a conceptual and a computational standpoint. Furthermore, the order quantity is no longer the simple, well-understood economic order quantity, and hence is likely to be more difficult to implement. Fortunately, as we saw, the percentage penalty in the total of replenishment, carrying, and shortage costs using the simpler sequential approach tends to be quite small.† Thus, in most situations, we were justified in advocating its use for B items. However, even small percentage savings may be attractive for A items. In the following sections, we present first the sequential approach, and then comment on, or present in detail, the simultaneous approach.

We now present rules for two different shortage-costing measures for slow-moving items. In Sections 8.1.1 and 8.1.2, we choose Q and s sequentially. In Section 8.1.3, we present simultaneous selection of s and Q .

8.1.1 B_2 Cost Measure for Very-Slow-Moving, Expensive Items ($Q = 1$)

As mentioned above, the discrete nature of the Poisson distribution causes problems in obtaining an operational decision rule. For one important situation, namely, where the replenishment quantity $Q = 1$, an analysis leads to a decision rule quite easy to implement.‡ First, let us see under what conditions $Q = 1$ makes sense, for any distribution, including the Poisson. From Equation 4.3 of Chapter 4, the total relevant costs, if an order quantity of size Q is used, are

$$TRC(Q) = Qvr/2 + AD/Q$$

We are indifferent between $Q = 1$ and $Q = 2$ where

$$TRC(1) = TRC(2)$$

that is,

$$vr/2 + AD = vr + AD/2$$

or

$$D = vr/A \tag{8.3}$$

For any lower value of D , we prefer the use of $Q = 1$. Again note that this rule ignores shortage costs and thus may lead to some errors. (See Problem 8.25.) To illustrate the rule, suppose that $A/r \approx 13.11$. Therefore, we have

$$\begin{aligned} \text{Use } Q = 1 \text{ if} \\ D < 0.0763v \end{aligned}$$

* Natarajan and Goyal (1994) discuss many methods of setting safety stock and illustrate the interdependence of s and Q .

† As noted in Chapter 4, Zheng (1992) shows that the fractional cost penalty of using the EOQ is always less than 1/8, and is usually much less. See also Axsäter (1996).

‡ When $Q = 1$, an (s, Q) policy is identical to an $(S - 1, S)$ policy because an order is placed to replace units sold every time a demand occurs. See Moinzadeh (1989), Schultz (1990), and Moinzadeh and Schmidt (1991) for additional work on this problem. The latter paper uses information about pipeline orders, which is available in many cases due to supply chain integration and other technological developments.

8.1.1.1 Assumptions behind the Derivation of the Decision Rule

The assumptions include:

1. Continuous-review, order-point, order-quantity system with $Q = 1$. The remaining decision variable is the order point, s , or the order-up-to level, $S = s + 1$.
2. Poisson demand. (The derivation in Section 8A.2 of the Appendix of this chapter does not depend on Poisson demand, however, so that similar decision rules apply for any discrete demand distribution.)
3. The replenishment lead time is a constant L time periods. (The results still hold if L has a probability distribution as long as we use its mean value, $E(L)$. However, the derivation is considerably more complicated in this more general case.)
4. There is complete backordering of demands when out of stock. Karush (1957) and Mitchell (1962) have investigated the case of complete lost sales; the results differ very little from the complete backordering situation. (See Rabinowitz et al. 1995 and Ouyang et al. 1996 for the case of partial backorders.)
5. There is a fixed cost, B_2v , per unit backordered (i.e., a fixed fraction, B_2 , of the unit value is charged for each unit backordered).

8.1.1.2 Decision Rule

As shown in Section 8A.2 of the Appendix of this chapter, we are indifferent between reorder points s and $s + 1$ when

$$\frac{p_{p_0}(s + 1|\hat{x}_L)}{p_{p_0\leq}(s|\hat{x}_L)} = \frac{r}{DB_2} \tag{8.4}$$

where*

- $p_{p_0}(s + 1|\hat{x}_L)$ = probability that a Poisson variable with mean \hat{x}_L takes on the value $s + 1$
- $p_{p_0\leq}(s|\hat{x}_L)$ = probability that a Poisson variable with mean \hat{x}_L takes on a value less than or equal to s

and, as earlier,

- r = carrying charge, in \$/\$/unit time
- D = demand rate, in units/unit time
- B_2 = fraction of unit value charged per unit short
- \hat{x}_L = average (forecast) demand in a replenishment lead time

Equation 8.4 is developed by equating the total relevant cost function evaluated at s with the same function evaluated at $s + 1$, and manipulating the resulting equation. (The left-hand side of this equation is sometimes called the “reversed hazard rate,” the hazard rate being the ratio of the probability density function to the complement of the cumulative density function.) Equation 8.4 can be implemented relatively quickly on a spreadsheet.

* $p_{p_0}(x_0|\hat{x}_L) = \frac{(\hat{x}_L)^{x_0} \exp(-\hat{x}_L)}{x_0!} x_0 = 0, 1, 2, \dots$

8.1.1.3 Numerical Illustration

Consider the case of a high-value item with

- $v = \$350/\text{unit}$
- $D = 25 \text{ units/year}$
- $r = 0.24 \text{ \$/\$/year}$
- $L = 3.5 \text{ weeks} = 3.5/52 \text{ year}$
- $A = \$3.20$

and B_2 is estimated to be $1/5$ (i.e., $350/5$ or $\$70$ per unit backordered).

First, we can see from the sentence after Equation 8.3 that $Q = 1$. Now

$$\frac{r}{DB_2} = \frac{0.24}{25(1/5)} = 0.048/\text{unit}$$

and

$$\hat{x}_L = DL = 25 \left(\frac{3.5}{52} \right) \approx 1.7 \text{ units}$$

The use of Equation 8.4 shows that the best order point is 4. That is, we order one-for-one each time a demand drops the inventory position to 4. Another way of saying this is that the inventory position should always be maintained at the level of 5.

8.1.1.4 Sensitivity to Value of B_2

A precise value of B_2 is not easy to ascertain. Fortunately, there is a range of B_2 values for which each s value holds. To illustrate, consider the item in the above numerical illustration. Assume that D , r , and \hat{x}_L are all known at the indicated values. Repeated use of Equation 8.4 identifies the following regions where each s value shown is best for the given value of \hat{x}_L .

s Value	Range of r/DB_2	Corresponding Range of B_2
3	0.064–0.163	0.059–0.150
4	0.021–0.064	0.150–0.456
5	0.006–0.021	0.456–1.636

These results are striking. Note the wide ranges of B_2 for which the same value of s is used. For example, as long as B_2 is anywhere between 0.150 and 0.456, the best s value is 4. Again, we see a situation where the result is insensitive to the value of one of the parameters.

8.1.1.5 More General Case of $Q > 1$ and a B_2 Cost Structure

This is still the case in which a fraction of the unit cost is charged per unit short and Q and s are determined sequentially. The solution discussed above is only appropriate for $Q = 1$. When the use of Equation 8.4 indicates that the best replenishment quantity is 2 or larger, we can no longer obtain as convenient a solution, at least for the case of B_2 (a fraction of the unit value charged per

unit short). An exact analysis (see Problem 8.25) leads to indifference between s and $s + 1$ where the following condition holds:

$$\frac{\sum_{j=1}^Q p_{p_0}(s + j|\hat{x}_L)}{p_{p_0 \leq}(s|\hat{x}_L)} = \frac{Qr}{DB_2} \tag{8.5}$$

where

$p_{p_0}(x_0|\hat{x}_L)$ is the probability that a Poisson variable with mean \hat{x}_L takes on the value x_0

$p_{p_0 \leq}(x_0|\hat{x}_L)$ is the probability that the same variable takes on a value less than or equal to x_0

This equation is more challenging to develop on a spreadsheet. The use of another computerized routine is more appropriate.

8.1.2 Case of $Q \geq 1$ and a B_1 Cost Structure

In some cases, it may be more appropriate to assume a fixed cost B_1 per (impending) stockout occasion as opposed to a cost per unit short. For such a case, we have developed indifference curves useful for selecting the appropriate reorder point. As shown in Section 8A.2 of the Appendix of this chapter, we are indifferent between s and $s + 1$ as reorder points when

$$\frac{p_{p_0}(s + 1|\hat{x}_L)}{p_{p_0 \leq}(s|\hat{x}_L)} = \frac{Qvr}{DB_1} \tag{8.6}$$

where all the variables have been defined above. The indifference curves of Figure 8.1 are developed by repeated calculation of Equation 8.6: for a given value of s and \hat{x}_L , Equation 8.6 is solved for Qvr/DB_1 . For a given value of s , this is repeated for a number of values of \hat{x}_L to produce a curve representing indifference between s and $s + 1$ as a function of \hat{x}_L and Qvr/DB_1 . These curves define regions on a graph of Qvr/DB_1 versus \hat{x}_L where we prefer one s value to others.

For a given item, values of Q , v , r , D , and B_1 are used to evaluate the quantity Qvr/DB_1 . The point in Figure 8.1 corresponding to this quantity and the appropriate \hat{x}_L indicates the best value of the order point, s , to use.

8.1.3 Simultaneous Determination of s and Q for Slow-Moving Items

As noted above, the sequential approach of first finding Q and then finding s can lead to higher costs. For the case of Poisson demand and various shortage costs, there are efficient methods for finding s and Q simultaneously. In a groundbreaking study, Federgruen and Zheng (1992) provide a very fast algorithm for both B_2 and B_3 shortage costs. (See also Gallego 1996.) An alternative is to build the total cost equation on a spreadsheet and search (perhaps using a data table) for the optimal values. Here, we simply present two total cost equations and refer to the relevant literature for the algorithm. It will be clear from the cost equations that a spreadsheet will not be as effective as a routine written with a macro in a spreadsheet or in another computer programming language.

For the case of B_3 shortage costs, we have

$$\text{ETRC}(s, Q) = AD/Q + \frac{1}{Q} \sum_{y=s+1}^{s+Q} \left\{ rv \sum_{i=0}^y (y - i)p_{p_0}(i|\hat{x}_L) + B_3v \sum_{i=y+1}^{\infty} (i - y)p_{p_0}(i|\hat{x}_L) \right\} \tag{8.7}$$

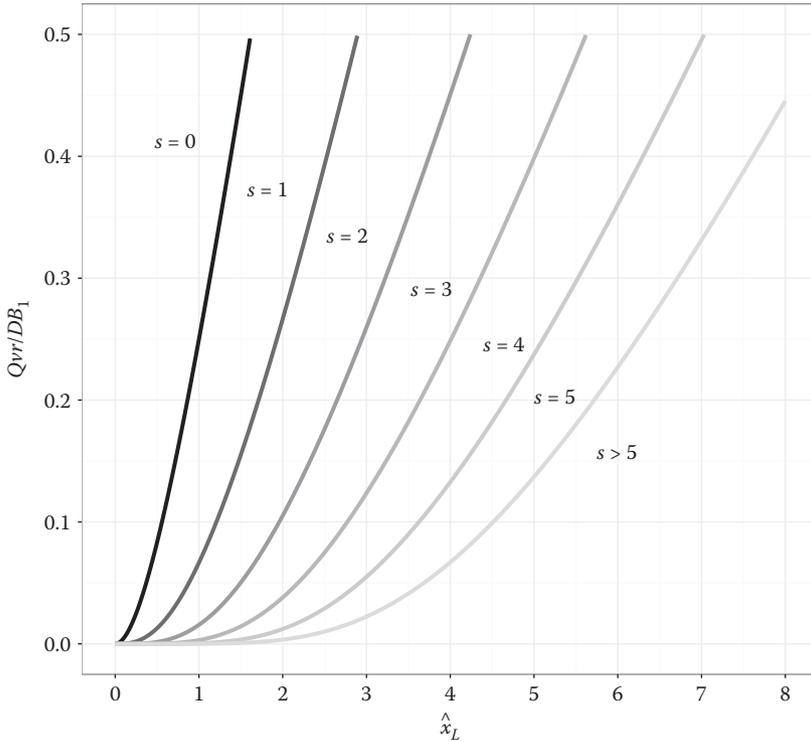


Figure 8.1 Indifference curves for the reorder point under the B_1 cost measure.

For the case of B_2 shortage costs, we have

$$ETRC(s, Q) = AD/Q + \frac{1}{Q} \sum_{y=s+1}^{s+Q} \left\{ rv \sum_{i=0}^y (y-i)p_{p_0}(i|\hat{x}_L) + B_2 Dv \sum_{i=y}^{\infty} p_{p_0}(i|\hat{x}_L) \right\} \quad (8.8)$$

See Federgruen and Zheng (1992) for details. We have found that in spite of the speed of the algorithm, and the fast computers available, many managers resist these more complex equations and algorithms. Thus, we suggest that a test be done on a small sample of items to see if the simultaneous solution is worth the added effort.

8.2 Controlling the Inventories of Intermittent Demand Items

In Section 3.8.1, we argued for a special type of forecasting for this kind of item that specifically separates out as two variables (1) the time (n) between demand transactions and (2) the sizes (z) of the transactions. In particular, we advocated estimating \hat{n} (the average number of periods between transactions), \hat{z} (the average size of a transaction), and $MSE(z)$. Now, in selecting values of control variables, such as s , we must take account of the intermittent nature of the demands; specifically, there may now be a significant chance of no demand at all during a lead time. Problem 8.29 deals with the case of an (s, Q) system. When intermittent demand is known well ahead of time, it is

more appropriate to use a decision rule based on a deterministic, time-varying demand pattern (see Chapter 5). Ekanayake and Norman (1977) discuss this type of issue in a case study context. Scala et al. (2013) describe using a base-stock model with intermittent demand, and Teunter et al. (2011) provide an adjustment for obsolescence in an intermittent demand setting. Bachman et al. (2016) present an approach for managing highly variable items that has been effectively applied for the U.S. Defense Logistics Agency. See also Croston (1972) and Shenstone and Hyndman (2005). Related works include Silver (1970), Wilcox (1970), Cawdery (1976), Ward (1978), Williams (1982, 1984), Alscher et al. (1986), Etienne (1987), Schultz (1987), Watson (1987), Yilmaz (1992), Hollier et al. (1995), Johnston and Boylan (1996), Syntetos and Boylan (2001), and Snyder et al. (2012).

8.3 Nature of C Items

The categorization of items for control purposes was first discussed in Chapter 2. As a reminder, we repeat some of the important points here.

The primary factor that indicates that an item should be placed in the C category is a low dollar usage (Dv value). The exact cut-off value between the B and C categories should be selected so that somewhere on the order of 30%–50% of all the items are classified in the C category. A low Dv value alone is not sufficient to dictate that an item should be placed in the C category, however. The real requirement is that the annual total of replenishment, carrying, and shortage costs be quite low under any reasonable control strategy. Normally, Dv is a useful surrogate for this total. However, a low-dollar-usage item may have potentially severe shortage penalties associated with it. Some typical examples come to mind:

1. A slow-moving product that rounds out a service line provided to an important customer. Shortages on this item can cause severe reductions in the usage of several faster-moving items.
2. A product that is the “pride and joy” of the president because he or she was instrumental in its development. In this case, there is a high *implicit* cost associated with a shortage.
3. An inexpensive product that has a high *explicit* shortage cost. The authors vividly remember the case of an electronics manufacturer whose limited manufacturing area was clogged with five expensive pieces of custom-made equipment, each the size of two large refrigerators, because the company had once again run out of the plastic control knobs worth 7 cents each.

For a true C item, the low total of replenishment, carrying, and shortage costs implies that, regardless of the type of control system used, we cannot achieve a sizable *absolute* savings in these costs. Therefore, the guiding principle should be to use simple procedures that keep the control costs per SKU quite low. That is, we wish to keep the labor and paperwork per item to a minimum. Equipment that permits electronic capture of data at the point of sale clearly reduces control costs. However, in our opinion, some organizations are overusing this capability. If the actual cost of using the electronic system (including maintenance of the software and updating the inventory records) is quite low, such a system could be a useful tool for C items.* Finally, we should be careful in defining C items because an item’s classification can change over time. The classification should be reviewed periodically so that important items are not relegated to C item status.

* In spite of the widespread use of automated data collection systems and even the availability of spreadsheets for computing inventory policies, a remarkable number of firms still operate relatively manually.

8.4 Control of C Items Having Steady Demand

In this section, we deal with items having a relatively low importance and a demand rate that is not changing appreciably with time. Moreover, we assume that it makes sense to actually stock the C items under consideration. In Section 8.7, we address the more fundamental question of stocking an item versus making or buying it to each customer order.

8.4.1 Inventory Records

As a consequence of the primary objective mentioned above, in most cases, it may be most appropriate not to maintain any inventory record of a C item. It may be more desirable simply to rely on an administrative mechanism for reordering, such as placing an order when the last box in a bin is opened. If an inventory record is maintained, it should certainly not require that each transaction be recorded (except for the possible use of electronic point-of-sale data capture). Instead, a rather long review interval, such as 6 months, should be considered. Of course, for demand estimation and order control purposes, a record should be kept of when orders were placed and received.

Stated another way, we are saying that there are two choices for selecting a review interval for a C item:

1. Periodic review with a relatively long interval.
2. Continuous review but with a mechanism for triggering orders that requires neither a physical stock count nor the manual updating (in a ledger) of the stock status. Examples are (i) a two-bin system of control, and (ii) electronic capture of usage data.

8.4.2 Selecting the Reorder Quantity (or Reorder Interval)

Stated bluntly, the frills of the economic order quantity, Silver–Meal heuristic, and so on are completely unwarranted for Class C items. Instead, a simple offshoot of the basic EOQ developed in Chapter 4 should be used. Specifically, one of at most a few possible time supplies should be assigned to each Class C item. The time supply to use should be based as much on shelf life and obsolescence as on setup cost, unit cost, or carrying charge. It is usually reasonable to use a single value of the ratio A/r for all C items. (One convenient way of estimating this value is through the exchange curves of Chapter 4.) To illustrate, consider the following situation. The value of r is found to be 0.24\$/\$/year and a reasonable average value of A is \$3.20. Furthermore, for C items, management has decided that only three possible time supplies—6, 12, and 18 months—are worth considering. Using the approach suggested in Section 4.4 of Chapter 4, we develop the particularly simple decision rule shown in Table 8.1. D would not be estimated through a forecasting system but rather through knowledge of the starting and ending inventories over a convenient time period. For example, the procedure could be as follows in a two-bin (s, Q) system. Consider two consecutive orders A and B. Let A be *received* at time t_A and B be *placed* at time t_B . Let I_A be the inventory level just *before* A is received (a typically low level, hence easy to count). Then we have

$$\text{Demand rate} \approx \frac{I_A + Q - s}{t_B - t_A} \quad (8.9)$$

Of course, if pilferage losses are significant, we would have an overly high estimate of the demand rate. Needless to say, rather than taking account of these losses, the preferable approach is to reduce the losses themselves.

Table 8.1 Suggested Reorder Time Supplies for a Sample Firm's C Items with Steady Demand

For Annual Dollar Movement (Dv) in This Range (\$/Year)	Use This Number of Months of Supply
$53 \leq Dv$	6
$18 \leq Dv < 53$	12
$Dv < 18$	18

8.4.3 Selecting the Reorder Point (or Order-Up-to Level)

Any of the criteria discussed in Chapter 6 could be chosen for selecting the safety factor. (And with spreadsheets, the safety factors can be found easily.) However, we advocate the use of a specific criterion: select the safety factor to provide a specified expected time, TBS, between stockout occasions. We have found this criterion to be particularly appealing to management in most C item contexts. It appears to be a method that managers are comfortable with for expressing their risk aversion. Thinking in terms of an average time between stockouts is apparently more straightforward than dealing with probabilities or fractions. Quite often, many C items are involved in a single customer order. Therefore, to assure a reasonable chance of satisfying the complete customer order, a very high level of service must be used for each C item. Thus, large values of TBS (e.g., 5–100 years) are not unreasonable when we recognize the small added expense of carrying a high safety stock. The decision rule for the TBS criterion (see Section 6.7.11 of Chapter 6) is to select the safety factor k to satisfy*

$$p_{u \geq}(k) = \frac{Q}{D(\text{TBS})} \quad (8.10)$$

where

$$p_{u \geq}(k) = \Pr\{\text{Unit normal variable (mean 0, standard deviation 1) takes on a value of } k \text{ or larger}\}$$

is a widely tabulated function (see Table II.1 in Appendix II) and is a built-in function in many spreadsheets. (See Appendix III.) Then we set the reorder point[†]

$$s = \hat{x}_L + k\sigma_L \quad (8.11)$$

where[‡]

$$\hat{x}_L = DL \quad (8.12)$$

* In Equation 8.10, a negative k results if $Q/D(\text{TBS}) > 0.5$. This is very unlikely to occur if TBS is large, as suggested earlier. In any event, management should specify a minimum allowable value of k .

† If a periodic review, order-up-to-level (R, S) system is to be used, then throughout this section, the decision rules should be modified as follows (see Section 6.9):

1. Replace L by $R + L$
2. Replace Q/D by R
3. Replace s by S

‡ Note that \hat{x}_L is not an output of the forecast system (as mentioned earlier, we are advocating not forecasting C items). D is estimated as discussed in the preceding subsection.

is the expected demand (in units) in the replenishment lead time of length L (years), and σ_L is the estimate of the standard deviation (in units) of forecast errors over L . (We return shortly to a discussion of how σ_L is estimated.)

The appropriate question at this point is: “Who needs all this aggravation, especially for C items?” Equation 8.10 holds for general values of Q/D and TBS. In the context of any company, only a few values of each of these two parameters would be specified. For these few combinations, the equations could be used once and for all to develop a simple table. We illustrate this in Table 8.2, when demand is normally distributed for a sample firm, using the earlier selected Q/D values of 6, 12, and 18 months, and also management-specified values of 10, 20, and 50 years for TBS.

Once k is selected, the reorder point follows from Equation 8.11. However, for this purpose, an estimate of σ_L is required. One point is certain: σ_L should not be developed from observing forecast errors of the specific C item under consideration. Instead, we suggest the use of the Poisson distribution for developing an estimate of σ_L because it typically applies to slow-moving items. This approach provides quick estimates of σ_L even though it may not be exact.

For the Poisson distribution (see Equation 8A.1), we have the simple relation, which again will only approximate the true standard deviation, and therefore safety stock,* if demand is not Poisson

$$\sigma_L = \sqrt{\hat{x}_L} = \sqrt{DL} \text{ units} \tag{8.13}$$

Finally, we note that it may be a good idea to keep a record of stockouts to see if the specified TBS is being met.

8.4.3.1 Numerical Illustration

Consider an item that is a slow-moving with $D = 48$ units/year, $L = 2$ months, an order quantity time supply of 12 months, and a specified TBS of 20 years.

$$\hat{x}_L = DL = 48 \text{ units/year} \times 1/6 \text{ year} = 8 \text{ units}$$

From Table 8.2, we read that the safety factor k should be set at 1.64. Suppose we make the Poisson assumption. Then Equation 8.13 gives

$$\sigma_L \approx \sqrt{8} = 2.83 \text{ units}$$

From Equation 8.11 we have that the reorder point is

$$s = 8 + 1.64(2.83) = 12.6, \text{ say } 13 \text{ units}$$

8.4.4 Two-Bin System Revisited

In Section 6.4 of Chapter 6, we described a particularly simple physical form of an (s, Q) system—the two-bin system. To review, in a two-bin system, the on-hand inventory is segregated into two distinct sections (generically called bins). The capacity of one of the bins (called the reserve bin) is set equal to the reorder point. Demands are satisfied from the other bin until its stock is depleted.

* Alternatively, the estimating technique in Section 3.8.1 of Chapter 3 may be used.

Table 8.2 Table to Select Safety Factor, k , for a Sample Firm

TBS Value (Years)	Order Quantity Expressed as a Time Supply in Months ($12Q/D$)		
	6	12	18
10	1.64	1.23	1.04
20	1.96	1.64	1.44
50	2.33	2.05	1.88

Opening the reserve bin is a signal to place a replenishment order. When the order arrives, the reserve bin is refilled and sealed. The remainder of the order is placed in the other bin.

The authors have seen some rather clever forms of the bins concept. Examples include: (i) the use of “baggies” in the case of small electronic components, and (ii) color coding the end of an appropriate bar in steel stacked in a distribution warehouse.

To facilitate ordering, a bin reserve tag should be attached to the reserve bin so that when this bin is opened, the clerk will have the tag to remind him/her to report that the bin was opened. In fact, in many cases, the tag itself can be used as an order form. At the very least, dates and order quantities can be recorded on it (permitting an estimate of D when it is needed).

Obviously, for satisfactory performance of this system, it is imperative that a tag be promptly submitted to purchasing (or production scheduling) whenever a reserve bin is opened. Supervisors, assemblers, sales personnel, and others must be motivated to follow this procedure while they meet their own objectives in the rush of daily activities. One approach to maintaining this level of discipline is known as *plan for every part (PFEP)*, where the movement of each part is recorded in one database. See Harris (2004) for more detail on PFEP.

8.4.5 Simple Form of the (R, S) System

Smith (1984) reports on the implementation of a simple form of the (R, S) system, known as the semiautomatic stock control system, that is particularly effective for control of C (as well as B) items in a retail environment. Here is how it operates for any particular item, illustrated for a weekly review interval and a lead time less than R .

Periodically (likely less often than every review period, for example, each quarter-year), management specifies a desired value of S and at *only* that time the on-hand stock is counted and an order is placed to raise the inventory position to S . That is, the on-hand plus the order quantity must equal S . Then, at each regular review instant (here each week), the computer simply orders enough to replace sales since the last review. Therefore, the only record that must be maintained is the total sales in the current review interval. Problem 8.3 deals with how a desired increase or decrease in S would be implemented.

8.4.5.1 Numerical Illustration

Consider an item where the desired value of S is 35 units and a physical stock count reveals that the on-hand inventory is 26 units (with nothing on order). Then an order for 9 units would be placed. Suppose that the sales prior to the next review (ordering time) were 12 units. As a result, the next order would simply be for 12 units.

8.4.6 Grouping of Items

In some cases, it may be advantageous to group C items for control purposes. In particular, if a group of items (including some from the A and B categories) are provided by the same supplier, or produced in-house on the same equipment, then coordinated control (to be discussed in Chapter 10) may very well be in order to reduce replenishment costs. That is, when one item in the group needs reordering, several others should likely be included in the order. This is particularly appealing for a C item. By including it in an order initiated by another item, we avoid incurring a full setup cost (A) for its replenishment.

Coordination does not rule out the use of a two-bin system of control for the individual C item. Instead, reserve tags of opened reserve bins are held centrally between designated periodic times at which that group is ordered. Of course, the reserve stock must be appropriately scaled up to provide protection over a length of time equal to the sum of the periodic ordering interval plus the replenishment lead time.

8.5 Control of Items with Declining Demand Patterns

In this section, we deal with items nearing the end of their usage lifetimes (the end of the product life cycle that was discussed in Section 2.5 of Chapter 2). These ideas, however, are applicable beyond just C items. In particular, high-unit-value service parts (i.e., B and A items) eventually enter a period of declining requirements when the parent population is phased out.*

8.5.1 Establishing the Timing and Sizes of Replenishments under Deterministic Demand

Brousseau (1982) studied the special, but important, case where the (instantaneous) demand rate is dropping linearly with time. The demand pattern therefore is

$$x_t = a - bt \quad (8.14)$$

where t is time, and a and b are constants. For example, let $a = 10$, $b = 2$, and let t be measured in weeks. Then the demand rate at the end of week 1 is projected to be $10 - 2(1) = 8$, and will be zero at the end of week 5. The total remaining requirements are $a^2/2b$ (or 25 in the example). The optimal strategy turns out to be a function of only a single parameter, M , which is given by

$$M = \frac{Ab^2}{\sqrt{a^3}} \quad (8.15)$$

Note that M gets larger as a/b decreases—that is, as we get closer to the end of the demand pattern.

Brousseau ascertains as a function of M (1) how many replenishments to make and (2) the timing of the replenishments (as fractions of the horizon a/b). These results are shown in the form of a simple table. More importantly, however, he shows that a very simple procedure, based again on the EOQ, produces negligible cost penalties. Specifically, use the basic EOQ, based on the

* The case of increasing demand was addressed in Chapters 5 and 7.

current demand rate and expressed as a time supply, as long as $M < 0.075$, and use a single last replenishment equal to the total remaining requirements ($a^2/2b$) when $M \geq 0.075$. The cutoff value of 0.075 turns out to be approximately the value of M where we are indifferent between 1 and 2 replenishments *in the optimal solution*. (See Problem 8.12.)

Smith (1977) and Brown (1977) found similar results for other types of declining demand patterns (including a continuous analog of the geometric pattern described in Section 3.8.4 of Chapter 3)—namely, that use of the EOQ, until it implies an order quantity almost as large as the total remaining requirements, produces excellent results.

We operationalize these results by suggesting the use of the EOQ as long as the total remaining requirements are at least 1.3 EOQ. If not, simply use one last replenishment to cover the remaining requirements. It can be shown that the 1.3 factor is equivalent to Brosseau's cutoff M value for the case of linearly decreasing demand.

For further research on this topic, see Chakravarty and Martin (1989), who addressed the case of exponentially decreasing demand and price reductions. Wee (1995) developed an exact solution to the problem of finding a replenishment policy for an item that faces declining demand, and that deteriorates over time. See also Kicks and Donaldson (1980), Hariga (1994), Ting and Chung (1994), Hariga (1995), Bose et al. (1995), and Chapter 9.

8.5.2 Sizing of the Final Replenishment under Probabilistic Demand

When demand is declining, but probabilistic in nature, the choice of the size of one last replenishment involves a trade-off between (1) the costs of insufficient inventory (reordering, expediting, shortages, etc.) if the total remaining demand exceeds the inventory position after the order is placed, and (2) the costs of overacquisition if the total remaining demand is less than the inventory position. Again, the first category of costs is typically very difficult to establish. So management may resort to the use of a desired service level, which, in turn, is likely to be largely dictated by industry norms (e.g., meeting 95% of all requests for spare parts during a 10-year period after the last sale of a new unit). We present a decision rule based on satisfying a certain fraction of the remaining demand. The derivation is outlined in Section 8A.3 of the Appendix of this chapter. Fortuin (1980) treated other service measures. See also Cattani and Souza (2003) and Pourakbar et al. (2012, 2014).

8.5.2.1 Decision Rule

Select the order-up-to-level* S where

$$S = \hat{y} + k\sigma_y \quad (8.16)$$

with \hat{y} being the forecasted total remaining demand,[†] and σ_y the standard deviation of the total remaining demand. (One way to estimate σ_y would be to use the procedure of Brown (1981),

* $S = (\text{current inventory position}) + (\text{order quantity})$.

† For the deterministic linear model of Equation 8.14,

$$\hat{y} = \frac{a^2}{2b} \quad (8.17)$$

For the geometrically decaying pattern of Equation 3.77 of Chapter 3, namely

$$x_t = f^t x_0 = f^t a$$

referenced in Section 3.8.2 of Chapter 3.) The safety factor k satisfies

$$G_u(k) = \frac{\hat{y}(1 - P_2)}{\sigma_y} \tag{8.19}$$

where $G_u(k)$ is again the special function of the unit normal distribution tabulated in Table II.1 of Appendix II, and given for spreadsheets in Appendix III.

8.5.2.2 Numerical Illustration

An automotive manufacturer is concerned with the size of the last replenishment of a spare part for a particular make of automobile that was discontinued 3 years ago. A fit to recent historical monthly usage indicates that a geometrically decaying pattern is appropriate with

$$\begin{aligned} \hat{y} &= 83.1 \text{ units} \\ \sigma_y &= 16.8 \text{ units} \end{aligned}$$

Suppose that it is management’s decision to select an S value for the last replenishments such that management expects 95% of the remaining demand for the spare part to be met by S . This does not mean that the other 5% would be lost; instead, cost-incurring actions, such as parts conversion, may be taken to satisfy the remaining demand, or the customer (or service department at the dealer) would seek a third-party/aftermarket part.

By using Equation 8.19 , we obtain

$$G_u(k) = \frac{83.1}{16.8}(1 - 0.95) = 0.2473$$

Then, from Table II.1 of Appendix II,

$$k \approx 0.35$$

Hence, from Equation 8.16

$$\begin{aligned} S &= 83.1 + 0.35(16.8) \\ &\approx 89 \text{ units} \end{aligned}$$

8.6 Reducing Excess Inventories

When studying the inventories of many companies, it is common to find a significant percentage of the stocked items that have had absolutely no sales (or internal usage) in the last one or more

we have, from Equation 3.79 of Chapter 3,

$$\hat{y} = \frac{fa}{1 - f} \tag{8.18}$$

years. In some industries, the percentage of stocked items that have had no usage in the previous 52 weeks can be as high as 47 %. We refer to these SKUs as dead items.*

Of course, any remaining stock of a dead item is, by definition, excess stock. In addition, there are often many slow-moving items for which stock levels are excessive. The reasons for excess inventories can be grouped into two categories. First are errors associated with replenishments, that is, having purchased or produced too much of an item. These include production overruns, unjustified quantity purchases, errors in the transmission of an order request (e.g., 1,000 units instead of a desired 100 units), and inaccurate inventory records (e.g., location errors and “hoarding” of parts). The second class of reasons relates to overestimating the demand rate. Included here are inaccurate forecasts, deliberate changes in sales/marketing efforts, technological obsolescence (such as through an engineering change), and customer cancellations.

Whatever the causes of excess stocks, it is important to be able to (1) identify the associated items and (2) decide on what remedial action to take. Ideally, it would be desirable to be able to anticipate the decrease in usage of an item and take appropriate action to avoid being caught with a large surplus stock. It should be noted that this general problem area is increasing in importance as the rate of technological change increases, causing the life cycle of the typical product to shorten.

8.6.1 Review of the Distribution by Value

Again, we make use of a simple, but powerful, tool of analysis, the distribution by value (DBV) list. Recall from an earlier discussion (Table 2.1 in Chapter 2) if that the DBV is a listing of items in decreasing order of annual dollar usage (Dv value). At the bottom of the table listing will be all of the items that had no sales (or internal usage) over the period from which D was estimated (normally the most recent full year). Furthermore, moving up the table, we immediately encounter the items that have had very low usage.

To establish the DBV, values of D_i and v_i are required for each item i . If, at the same time, a reading on the inventory level I_i of item i can be obtained, then a very useful additional table can be developed as follows.

For each item, the expected time at which the current stock level will be depleted, also called the *coverage* or *runout time*, is computed by the formula

$$CO = \frac{12I}{D} \quad (8.20)$$

where

CO = coverage, in months

I = on-hand inventory, in units

D = expected usage rate, in units/year

The items are shown in Table 8.3 in decreasing order of coverage. Also shown is the inventory (in dollars) of each item and the cumulative percentage of the total inventory as one proceeds down the table. In Table 8.3 (involving a population of 200 items with a total inventory valued at \$7,863), we can see that 4.2% of the total inventory is tied up in stock of zero-movers and

* Occasionally, items are kept in inventory to be used as parts for other products. These items never sell as is, but they are vital to the overall service performance. Clearly, these items should be stocked.

Table 8.3 Items Listed by Coverage

Rank Order	Cumulative Percent of Items	Item Identification	I (Units)	D (Units/Year)	v (\$/Unit)	CO (Months)	I _v (\$)	Cumulative Inventory	
								In Dollars	As Percent of Total
1	0.5	–	150	0	1.80	Infinite	270	270	3.4 ^a
2	1.0	–	60	0	0.50	Infinite	30	300	3.8
3	1.5	–	10	0	1.10	Infinite	11	311	4.0
4	2.0	–	5	0	3.20	Infinite	16	327	4.2
5	2.5	–	53	2	1.00	318	53	380	4.8
6	3.0	–	40	3	0.90	160	36	416	5.3
7	3.5	–	64	12	2.00	64	128	544	6.9
8	4.0	–	180	37	0.35	58.4	63	607	7.7
		·							
		·							
		·						7,861	100.0
199	99.5	–	2	1,000	1.00	0.024	2.00	7,863	100.0
200	100.0	–	0	463	0.20	0	0	7,863	100.0

^a 3.4 = (270/7,863) × 100.

some 6.9% is included in items having a coverage of 5 years (60 months) or more. Such information would be of use to management in deciding how serious the overstock situation is and what remedial action to take.*

Individual item usage values (D_i 's) are needed to construct the distribution by value table. Where these are not readily available, a practical alternative is to simply tour the storage facilities, applying a so-called dust test: any stock that looks overly dusty (or its equivalent) is a candidate for identification as a dead or very-slow-moving item.

8.6.2 Rule for the Disposal Decision

By disposing of some stock, we now achieve benefits of salvage revenue (possibly negative) and reduced inventory carrying costs. On the other hand, we will have to replenish at an earlier time than if no stock disposal was made. With regard to inventory carrying costs, it should be recognized that often the most costly aspect of excess inventory is that it is taking up significant space in a storage area of limited capacity. In addition, there is often a nuisance cost associated with stock that has physically deteriorated. Examples include the presence of rats and vermin, excessive rusting, and so forth. The authors recall the case of manufactured, large, metal piping that was stored in an open yard. One particular item, left standing for a number of years, had slowly sunk into the mud with each spring thaw until it stabilized in a semisubmerged state. High-cost excavation equipment was required to remove the dead stock in order to free up the space it was occupying. In this subsection, we present two approaches to this disposal decision, one that assumes deterministic demand, and one that assumes probabilistic demand.

8.6.2.1 Deterministic Demand

The derivation, presented in Section 8A.4 of the Appendix of this chapter, is based on a deterministic level demand pattern. Alternative decision rules could be developed (1) for a declining, but still deterministic, demand pattern, or (2) for probabilistic demand; the latter case closely parallels the selection of S for the terminal replenishment (see Section 8.5.2).

Let W be the amount of stock to dispose. Then W is given by

$$W = I - \text{EOQ} - \frac{D(v - g)}{vr} \quad (8.21)$$

where I is the current inventory level of the item,

$$\text{EOQ} = \sqrt{\frac{2AD}{vr}}$$

as earlier, and g is the unit salvage value (g can be negative if there is a net cost per unit associated with disposing of stock). Note that when $v = g$, that is, the salvage value is equal to the full unit variable cost, Equation 8.21 says that $I - W$ (which is the amount to leave in stock) is equal to the EOQ. That is, when we get our full money back, we dispose of stock so that on-hand inventory will just equal the EOQ. (See also Problem 8.10.)

* We worked with one firm in which nearly 1,000 of the 3,600 items in the firm's inventory sold so infrequently that management was aggressively working to eliminate them altogether.

8.6.2.2 Numerical Illustration

The manager of the hardware division of a retail outlet has discovered that the stock level of a particular item is 1,200 units, not 200 units as he had expected. Thus, he feels that some of the stock should be disposed of by means of a sale at $2/3$ of unit cost (i.e., $g = (2/3)v$). He estimates the following values for other characteristics of the item:

$$\begin{aligned} D &= 400 \text{ units/year} \\ A &= \$20 \\ v &= \$3.00/\text{unit} \\ r &= 0.25 \text{ \$/\$/year} \end{aligned}$$

Equation 8.21 implies that the disposal amount should be

$$W = 1,200 - \sqrt{\frac{2(20)(400)}{(3.00)(0.25)}} - \frac{400(3.00 - 2.00)}{(3.00)(0.25)} = 521 \text{ units}$$

8.6.2.3 Probabilistic Demand

The second approach due to Rosenfield (1989) focuses on the number of units to be saved (not disposed of) and initially assumes that demands arrive according to a Poisson process and no further replenishments are to be made. If there is one unit demanded per arrival, a simple formula finds the optimal number of units to be saved. Find the largest integer less than

$$\log((V_c + a/i)/(V_u + a/i))/\log(D/D + i) \tag{8.22}$$

where

- V_c = the salvage value at the present time as a percentage of the current value
- V_u = the ultimate salvage value as a percentage of the current value
- i = discount rate
- a = carrying charge excluding capital holding costs, as a percentage of the current value
- D = average demand per year.

8.6.2.4 Numerical Illustration

Consider an item with the following characteristics:

$$V_c = 0.5, V_u = 0.7, i = 0.12, a = 0.09, D = 20$$

Find the optimal number of units to save in inventory, if there is always one unit demanded per customer arrival, and demands arrive according to a Poisson process. The use of Equation 8.22 gives

$$\begin{aligned} &\log((V_c + a/i)/(V_u + a/i))/\log(D/D + i) \\ &= \log((0.5 + 0.09/0.12)/(0.7 + 0.09/0.12))/\log(20/20 + 0.12) = 24.81 \end{aligned}$$

So the firm should save 24 units. If current inventory is 50 units, they should dispose of 26 units.

Rosenfield also develops formulas for several more complex cases, including items that are perishable. For example, if the batch size distribution is geometric with parameter q , a simple formula is given:

Find the largest integer less than

$$\frac{\log((V_c + a/i)/(V_u + a/i))}{\log(qD/(D + i) + (1 - q))}$$

For a geometric distribution, the probability of a batch size of y is $q(1 - q)^{y-1}$, and the mean batch size is $1/q$. The variance of the batch size is $(1 - q)/q^2$.

8.6.3 Options for Disposing of Excess Stock

Possible courses of action for disposing of excess stock include the following:

1. *Use for other purposes.* For example, in the case of components or raw materials, consider their use, possibly requiring some rework, in different final products.
2. *Shipment of the material to another location.* Where an item is stocked at more than one location, it may have negligible demand at one location but still experience a significant demand elsewhere. The transfer of stock not only frees up space at the “first” location, but it also helps avoid the costly setup (or acquisition) costs associated with satisfying the likely infrequent demands at the other location(s).
3. *Use of stock for promotional purposes.* Excess stock may be converted to a revenue generator by providing free samples of it when customers purchase other products. For example, an appliance store with an overstock of an inexpensive table lamp might offer it free as a bonus to each customer buying a television set. The extreme case of this action is donating the stock to some charitable cause.
4. *Mark-downs or special sales.* The unit price is reduced sufficiently to generate appreciable demand for the product. An example would be the reduced price of wrapping paper or decorations immediately following the Christmas season.
5. *Returns to suppliers at a unit value likely lower than the initial acquisition cost.* A good example is that of a college bookstore returning extra copies of a book to the publishing company’s distributor.
6. *Auctions.* These should not be overlooked as a possible means of salvaging some value from excess stock.
7. *Disposal for scrap value.* The best course of action may be to sell the material directly to a scrap dealer, a common transaction when there is a significant metallic component of the item.

Internal uses of the inventory (numbers 1 and 2) often provide better return than trying to move the inventory on the external market. Whatever course of action is selected, the excess stock will not disappear overnight. Therefore, it is important to lay out an approximate timetable for disposal, and then to ensure that there is appropriate follow-up. Also note that whenever inventory is disposed of and generates revenue that is less than the acquisition cost, there is an effect on the firm’s income statement. Operating profit may decrease because of the loss taken on the inventory.

Thus, income taxes may decrease, and cash flow may improve. These effects should be examined closely. (See Smith 1987.)

8.7 Stocking versus Not Stocking an Item

Many items with extremely low sales may be minor variations of faster-moving products. In such a case, it can be attractive to eliminate the slow-moving special versions. Of course, decisions of this type, as well as those where there is no substitute product, go beyond the area of production planning and inventory management. Marketing considerations, including customer relations, are obviously relevant. For example, the slow mover may be carried as a service item for an important customer. Furthermore, the appropriate course of action may not be to discontinue selling the item, but rather to make or buy it to order as opposed to stocking it, an issue addressed in the following material.

Given that we have decided to satisfy customer demands for a given SKU, the question we now face is: *Should we make a special purchase from the supplier (or production run) to satisfy each customer-demand transaction or should we purchase (or produce) to stock?* It should be emphasized that this question is not restricted to C items. As we will see, a low demand rate tends to favor nonstocking, but there are a number of other factors that also influence the decision.

Earlier in the book, we have seen that, given that an item is stocked, its total relevant costs per year are quite insensitive to the precise settings of the control variables (e.g., the reorder point and the order quantity). Interestingly enough, more substantial savings can often be achieved through the appropriate answer to the question of whether or not the item should even be stocked.

8.7.1 Relevant Factors

For simplicity, our discussion is in terms of a purchased item, but the concepts are readily adaptable to in-house production. There are a number of factors that can influence the decision to stock or not stock the item. These include:

1. The system cost (file maintenance, forecasting, etc.) per unit time of stocking an item.
2. The unit variable cost of the item both when it is bought for stock and when it is purchased to meet each demand transaction. A more favorable price may be achieved by the regular larger buys associated with stocking. In addition, a premium per unit may be necessary if the nonstocking purchases are made from a competitor.
3. The cost of a temporary backorder associated with each demand when the item is not stocked.
4. The fixed setup cost associated with a replenishment in each context. An account should be taken of possible coordination with other items, because setup costs may be reduced.
5. The carrying charge (including the effects of obsolescence) which, together with the unit variable cost, determines the cost of carrying each unit of inventory per unit time.
6. The frequency and magnitude of demand transactions.
7. The replenishment lead time.

Often, it is best to avoid stocking an item at a given location if there is an internal supplier (a central warehouse, for example) that can hold inventory for several locations. Of course, it is important to examine the replenishment lead time and ascertain that customers are willing to wait

the additional transportation time. If not, the cost of a temporary shortage should be included in the analysis. In what follows, we assume that the replenishment lead time is quite short.

8.7.2 Simple Decision Rule

A general model to handle all of the above factors would be of limited value to the typical practitioner because of its complexity. Therefore, in this subsection, we present a simple rule (based on work by Popp 1965), valid under specific assumptions. Even when these assumptions are violated, the rule should still be a useful guideline. In any event, in the next subsection, some extensions will be discussed.*

8.7.2.1 Assumptions

1. The unit variable cost is the same under stocking and nonstocking.
2. The fixed setup cost is the same under stocking and nonstocking.
3. In deriving the decision rule to decide on whether or not to stock the item, we allow the order quantity to be a noninteger. Of course, if the item was actually stocked and demands were in integer units, we would use an integer value for the order quantity.
4. The replenishment lead time is negligible; consequently, there is no backordering cost.

Based on comparing the cost of not stocking with the cost of stocking, we establish the following decision rule†:

8.7.2.2 The Decision Rule

Do not stock the item if either of the following two conditions holds (otherwise stock the item):

$$c_s > A/E(i) \quad (8.23)$$

or

$$E(t)vr > \frac{E(i)}{2A} \left[\frac{A}{E(i)} - c_s \right]^2 \quad (8.24)$$

where

c_s = system cost, in dollars per unit time, of having the item stocked

A = fixed setup cost, in dollars, associated with a replenishment

$E(i)$ = expected (or average) interval (or time) between demand transactions

$E(t)$ = expected (or average) size of a demand transaction in units

v = unit variable cost of the item, in \$/unit

r = carrying charge, in \$/\$/unit time

If the decision is to stock the item, then the economic order quantity should be used.

$$EOQ = \sqrt{\frac{2AD}{vr}} \quad (8.25)$$

* Valadares Tavares and Tadeu Almeida (1983) have presented an even simpler decision rule under somewhat more restrictive assumptions (Poisson demand and a choice solely between stocking 1 unit versus no stocking).

† The derivation of this rule is given in Section 8A.5 of the Appendix of this chapter.

where

$$D = E(t)/E(i) \tag{8.26}$$

is the demand rate, in units per unit time.

The behavior of the rule as a function of the factors involved makes intuitive sense. As the setup cost A goes up, we tend to stock the item. Likewise, as the expected time between transactions decreases, we prefer to stock the item. As v or r increases, we tend not to stock because it becomes too expensive to carry the item in inventory. As c_s increases, we are less likely to prefer stocking the item. Finally, the larger the expected transaction size $E(t)$, the less likely we are to stock the item because the individual special orders satisfy a relatively large demand.

8.7.2.3 Numerical Illustration

A firm has been purchasing a specialty item PDF-088 from a competitor each time that a customer demand has been encountered. The manager of purchasing feels that it would be attractive to make less frequent, larger purchases. Relevant parameter values have been estimated as follows:

$$\begin{aligned} v &= \$4.70/\text{roll} \\ E(t) &= 1.4 \text{ rolls} \\ E(i) &= 10 \text{ weeks or } 10/52 \text{ year} \\ A &= \$2.50 \\ r &= 0.24 \text{ \$/\$/year} \\ c_s &= 0.20 \text{ \$/year} \end{aligned}$$

We have

$$\begin{aligned} E(t)vr &= (1.4)(4.70)(0.24) = 1.58 \text{ \$/year} \\ A/E(i) &= 2.50/10/52 = 13.00 \text{ \$/year} \end{aligned}$$

The use of Equations 8.22 and 8.23 reveals that the item should indeed be purchased for stock, with the best order quantity being 6 rolls.

8.7.3 Some Extensions

It is rather straightforward (see Problem 8.6) to generalize the decision rule of the previous subsection to allow the unit variable cost and the fixed setup cost to each depend on whether or not the item is purchased for stock. In place of Equations 8.23 and 8.24, we can develop the following:

Do not stock the item if either of the following conditions holds (otherwise stock the item):

$$c_s > \frac{A_{ns}}{E(i)} + \frac{E(t)}{E(i)}(v_{ns} - v_s) \tag{8.27}$$

or

$$E(t)v_s r > \frac{E(i)}{2A_s} \left[\frac{A_{ns}}{E(i)} + \frac{E(t)}{E(i)}(v_{ns} - v_s) - c_s \right]^2 \tag{8.28}$$

where

- v_s = unit variable cost, in \$/unit, if the item is stocked
- v_{ns} = unit variable cost, in \$/unit, if the item is not stocked
- A_s = fixed setup cost, in dollars, if the item is stocked
- A_{ns} = fixed setup cost, in dollars, if the item is not stocked (the cost of backordering and extra paperwork would be included in A_{ns})

and all the other variables are as defined after Equations 8.22 and 8.23.

Croston (1974) developed a decision rule for the case of a periodic review order-up-to-level (R, S) system. His model assumes a negligible replenishment lead time, at most one demand transaction in each review interval, and a normal distribution of transaction sizes. Shorrock (1978) developed a tabular aid for the stock/no stock decision.

It should be emphasized that the two quantities, $E(t)$ and $E(i)$, needed in Equations 8.23 and 8.24 or Equations 8.27 and 8.28 would be estimated from rather sparse data because the items we are considering are slow-moving. The statistical fluctuations of such limited data could cause an item, whose underlying parameters had really not changed, to pass the stocking test 1 year and fail it in the next. Johnson (1962) proposed the useful idea of two threshold values, one to discontinue stocking an item, the other to institute stocking. The intermediate area helps prevent an item from flipping back and forth between stocking and nonstocking.

8.8 Summary

In this chapter, we have discussed procedures for controlling the inventories of the large number of items with low annual dollar usage as well as important items that are slow-moving. For low-value items, simplicity of control was stressed because of the low annual total of replenishment, carrying, and shortage costs associated with such items. In other words, if the total of replenishment, carrying, and shortage costs for an item is just a few dollars per year, we must use a control system for that item that costs only pennies per year.

A final point worth mentioning is that with C items, in particular, the best method of control may be ascertained from the warehouse or production supervisor who knows, for example, that a lot size of 10 units fits conveniently into a storage bin or into the usual production schedule. Using a decision rule that provides a different quantity, with a trivial annual difference in costs, is hardly worth the risk of alienating the supervisor.

Problems

- 8.1 Consider an item with the following characteristics:

$$v = \$0.40/\text{unit}$$

$$D = 50 \text{ units/year}$$

$$L = 3 \text{ months}$$

In establishing a value of σ , it is reasonable to assume that demand for the item is adequately approximated by a Poisson process. For an (s, Q) control system

- a. Find the appropriate Q value using Table 8.1.
- b. Find the s value assuming a desired TBS of 20 years and using Table 8.2.

- 8.2 Suppose for the item in Problem 8.1 we instead decided to use an (R, S) system of control. What should be the values of R and S ?
- 8.3 Suppose that in a retail outlet an item was being controlled by the semiautomatic stock control system described in Section 8.4.5 and that an S value of 35 units had been in use.
- Owing to changing conditions, management now feels that it is appropriate to increase S to 40 units. Exactly what, if anything, has to be done to ensure that the system uses $S = 40$ from now on?
 - Repeat part (a) for a decrease of S from 35 to 28 units.
- 8.4 The Fly-by-Night airlines has been phasing out one of the older types of aircraft that has been a mainstay of its illustrious fleet. Consequently, it has been observed that the associated usage of a glue used for maintenance of this type of aircraft has been “dropping off” linearly with time. The current level and trend of usage are estimated to be 110 liters/year and minus 45 liters/year/year. Other characteristics of the glue include $A = \$25$, $v = \$4.50/\text{liter}$, and $r = 0.3 \text{ } \$/\$/\text{year}$.
- If it is now time to replenish (i.e., there is no inventory left), what order quantity should be used?
 - If the usage rate continues to drop off at the same linear rate, what should be the size of the next replenishment?
- 8.5 Dr. Hunsen Bunnydew of Puppet Laboratories, Inc. had asked his assistant Test Tube to order in some beakers. Poor Test Tube made a mistake in placing the order, and the lab now has a 20-year supply! Hunsen has informed Test Tube that he must dispose of an appropriate number of beakers. Test Tube’s friend, Zongo, has arranged for another medical laboratory to purchase any quantity that Puppet offers, but at $\$0.50/\text{beaker}$, which is $1/3$ of what Test Tube paid for each of them. Using the following information, come to Test Tube’s aid by suggesting an appropriate amount to sell off.
- $$A = \$15$$
- $$r = 0.2 \text{ } \$/\$/\text{year}$$
- $$D = 500 \text{ beakers/year}$$
- 8.6 Develop Equations 8.27 and 8.28.
- 8.7 For the specialty item discussed in Section 8.7.2, how large would c_i have to be to make it unattractive to stock the item?
- 8.8 A C item is often purchased from the same supplier as a B or A item. Suppose that such is the case and, based on an EOQ analysis for the B (or A) item involved, an order is placed for the B (or A) item every 2 months. Assume that demand for the C item is essentially constant with time at a rate of 18 units/year. Assume that the item has a unit value of $\$3/\text{unit}$. The *additional* fixed cost of including the C item in an order of the B (or A) item is $\$1.20$. The carrying charge is $0.24\$/\$/\text{year}$. It is reasonable to restrict attention to ordering the C item in a quantity that will last for 2 months, 4 months, 6 months, and so on (i.e., an integer multiple of the time supply of the other item). Which of these time supplies is preferred?
- 8.9 Suppose you have been hired as an inventory control consultant by a client having thousands of inventoried items. The client feels that there is merit in each of the (R, S) and two-bin (s, Q) systems of control and is willing to have some C items under each of the two types of control. Using whatever symbols you feel are appropriate, develop as simple a decision rule as possible for deciding on which type of control to use for each specific item.

- 8.10 How are the disposal decision of Section 8.6.2 and the special opportunity to buy decision of Section 4.9.5 related? In particular, compare Equations 8.21 and 4.25.
- 8.11 For the item used in the numerical illustration of Rosenfield's approach in Section 8.6.2:
- What happens if the current salvage value is 0.8 instead of 0.5? Explain the result.
 - What happens if the current salvage value is 0.7 instead of 0.5? Explain the result.
 - Now suppose that the batch size is geometric with parameter $q = 0.4$ and the current salvage value is 0.5. How many units should be saved?
- 8.12 For the context studied by Brosseau, namely, a deterministic, linearly decreasing demand pattern, $x_t = a - bt$, use a dynamic programming formulation to show that the optimal strategy does indeed only depend upon the single parameter $M = Ab^2/vra^3$.
- 8.13 Suppose a company is using the following rule for controlling the stock of C items. Whenever the inventory position drops to 2 months or lower, order a 6 month supply. For each of the following items, what is the implied value of TBS? Assume Poisson demand.
- Item 372: $D = 60$ units/year, $L = 1.5$ months
 - Item 373: $D = 60$ units/year, $L = 3$ months
 - Item 374: $D = 30$ units/year, $L = 1.5$ months
- 8.14 Under deterministic level demand and an EOQ strategy, how does the coverage of an item vary throughout a cycle?
- 8.15 Consider an item with a linearly decreasing demand pattern. Suppose that the current level is 1,000 units/year and the trend is -400 /year.
- What are the remaining all-time requirements of the item?
 - Suppose the fixed cost per replenishment is \$50, the unit variable cost is \$0.80/unit, and the carrying charge is 0.20 \$/\$/year. Use Brosseau's method to establish the timing and sizes of all remaining replenishments. (Assume the current inventory level is 0.)
- 8.16 Consider a company with $A = \$4.00$, $r = 0.30$ \$/\$/year, and cost of stocking $c_s = \$0.50$ /year. There are two items (X and Y) having the same annual demand rate, $D = 10$ units/year and unit value $v = \$6.00$ /unit. However, the transactions for item X tend to be larger than those for item Y , namely, the average transaction size for X is 5 units while that for Y is only 2 units.
- Show, using Section 8.7, that X should not be stocked and Y should be stocked.
 - Why are the decisions different for the two items?
- 8.17 A company, attempting to remove slow movers from its line of stocked items, proposes the following types of rules for deciding on whether or not to stock an item. If an item is currently being stocked and no demands have been observed in the last T years, then the item should not be stocked any longer. On the other hand, if the item is not currently being stocked, we should begin stocking it if a total demand of at least m units is observed in a 1-year period. The quantities T and m are control variables. There is a fixed cost c_1 associated with activating the stocking of an item (preparation of a new form, etc.). Suppose that it is reasonable to assume that demand for an item is Poisson distributed with rate D (where D is not really known with certainty).
- For a given value of D , what is the probability that a currently stocked item will be removed from the files precisely T years after a particular demand transaction? Develop a graph of T values, as a function of D , that ensure a probability of 0.90.
 - Consider an item that has been removed from routine stocking at a time t . For a given value of D , express the probability of activating the item in the very next year (following t) as a function of m .

- c. Without developing explicit decision rules, how do you believe that each of T and m will change as c_1 increases?
- d. In reality, D is changing with time. How might one attempt to recognize this in selecting values of T and m ?

8.18 Under the assumption that lead time demand has a Poisson distribution, we have that

$$s = DL + k\sqrt{DL}$$

For fixed values of Q/D and TBS, indicate how you would develop a graphical aid giving s as a function of DL . Illustrate for $Q/D = 0.5$ year and TBS = 10 years. What is the s value for $DL = 3.7$ units?

- 8.19 For an item that has demand rate $D = 36$ units per year, $L = 1$ month, and an order quantity of 12 units:
 - a. Find the reorder point if TBS is specified as 10 years.
 - b. How does the result change if TBS = 2 years?
 - c. How does the result change if TBS = 20 years?
- 8.20 If the inventory of an item is currently at 2,000 units because of a clerk who bought too much, how much should be disposed of if the following data apply to the item?

$$g = 0.5v$$

$$D = 600 \text{ units/year}$$

$$A = \$25$$

$$v = \$4/\text{unit}$$

$$r = 0.2 \text{ \$/\$/year}$$

8.21 For the item described below, use Rosenfield's approach in Section 8.6.2 to find the number of units to be saved.

$$V_c = 0.2, V_u = 0.5, i = 0.14, a = 0.07, D = 30$$

Assume that there is always one unit demanded per customer arrival, and that demands arrive according to a Poisson process.

- 8.22 In an (s, Q) system, it can be shown that the inventory position has a uniform distribution with probability $1/Q$ at each of the integers $s + 1, s + 2, \dots, s + Q - 1, s + Q$.
 - a. Using this result, show, for Poisson demand and the B_2 shortage cost measure, that

$$ETRC(s) = vr \sum_{x_0=0}^s (s - x_0) p_{p_0}(x_0|\hat{x}_L) + B_2 v D \sum_{j=1}^Q \frac{1}{Q} \sum_{x_0=s+j}^{\infty} p_{p_0}(x_0|\hat{x}_L)$$

b. Show that indifference between s and $s + 1$ exists when

$$\frac{\sum_{j=1}^Q p_{p_0}(s + j|\hat{x}_L)}{p_{p_0 \leq}(s|\hat{x}_L)} = \frac{Qr}{DB_2}$$

(Note: $p_{p_0}(x_0|\hat{x}_L)$ is the p.m.f. of a Poisson variable with mean \hat{x}_L .)

- 8.23 Build a spreadsheet to quickly solve for the indifference point using Equation 8.4.
- 8.24 An item in a warehouse seems to sell quite slowly—only 20 units per year. However, it is an expensive item, valued at \$750. Assume that the variance of the lead time demand is roughly equal to the mean, so a Poisson distribution adequately describes the lead time demand. The following data may be relevant:

$$A = \$7 \quad r = 0.22\$/\$/\text{year} \quad B_2 = 2$$

$$L = 1 \text{ week (assume a 50 week year)}$$

- a. What (s, Q) policy should be used?
 - b. Now assume that $A = \$16$. What is the new (s, Q) policy?
- 8.25 Gregg Olson stocks an important but slow-moving item in his variety store. The relevant data are the following. Demand is 8 per year on average, and the variance of lead time demand is roughly equal to the mean of lead time demand.
- $v = \$400/\text{unit}$
 - $A = \$5$
 - $r = 0.25/\text{year}$
 - $B_2 = 0.6$
 - $L = 2 \text{ weeks} = 1/26 \text{ year}$
- a. What (s, Q) policy should Gregg use?
 - b. If the methods of Chapter 6 are used, what (s, Q) policy would Gregg use?
- 8.26 Consider an (s, Q) system with Poisson demand, complete backordering, and $Q = 1$.
- a. Develop an expression for the expected number of units on backorder at any random point in time.
 - b. Using a cost of B_2v per unit backordered per unit time, develop the equation for indifference between s and $s + 1$.
- 8.27 Because of the industrious nature of the faculty members involved, the Department of Management Sciences at a well-known university consumes a large amount of $8\frac{1}{2} \times 11$ lined paper. One of the secretaries, with the aid of a graduate student, has developed the following procedure: Once a week (at 5 p.m. on Friday), she reviews the stock level. If there are 20 or fewer pads, she orders 100 additional pads, which always arrive late on the next Tuesday. If there are more than 20 pads on the shelf, she does nothing until at least the next Friday. Assume that the faculty's needs follow a Poisson distribution with the following rates (in pads/day):

Day	Sun	Mon	Tue	Wed	Thu	Fri	Sat
Mean	0.5	2	2	2	2	2	0.5

If a professor needs a pad of paper and none are available, he rushes over to the university bookstore (or to an outlet in the local community) and buys the pad.

- a. Indicate how one might arrive at the cost per unit short.
- b. Make a crude estimate of each of (i) the safety stock, (ii) the probability of stockout per replenishment cycle, and (iii) the expected number of pads short per year.

- c. Discuss how one could more accurately estimate the quantities in part 8.27b.
- 8.28 A maker of venetian blinds builds a wide variety of products. Demand tends to be erratic and is difficult to forecast. Each order of a particular blind requires a set of components with some of the components (e.g., cords) being common to many finished items. The company's policy is to satisfy any customer order within 1 week. The actual time dedicated to the satisfaction of an order, including some design work and retrieving the required set of components from inventory, varies between 1 and 3 days. Inventory records are not updated on a perpetual basis but rather the inventories of items are physically reviewed periodically (with R varying from 2 weeks to 5 weeks, depending upon the item). An order-up-to level is used for each component. Make some realistic suggestions for improving the inventory management situation in this company.
- 8.29 The Acme Company stocks an item where the demand (x) during the lead time has the following probability distribution:

$$\Pr(x = 0) = 0.2$$

is normal with mean 40 units and standard deviation 8 units. In other words, nonzero demands are normally distributed.

- If an (s, Q) system of control is used with $Q = 100$ units and the desired fraction of demand satisfied from shelf is 0.98, what value of s should be used?
- What is the associated value of the safety stock?

Appendix 8A: Poisson Distribution and Some Derivations

8A.1 Poisson Distribution

An extensive reference is the book by Haight (1967).

8A.1.1 Probability Mass Function

A Poisson variable is an example of a discrete random variable. Its probability mass function (p.m.f.) is given by

$$p_x(x_0) = \Pr\{x = x_0\} = \frac{a^{x_0} e^{-a}}{x_0!} \quad x_0 = 0, 1, 2, \dots \quad (8A.1)$$

where a is the single parameter of the distribution. The graph rises to a peak, then drops off except for very low values of a where the peak is at $x_0 = 0$.

8A.1.2 Moments

One can show (see, e.g., Rohatgi 1976) that

- $E(x) = \sum_{x_0=0}^{\infty} x_0 p_x(x_0) = a$
that is, the parameter a is the mean of the distribution.
- $\sigma_x^2 = \sum_{x_0=0}^{\infty} [x_0 - E(x)]^2 p_x(x_0) = a$

Hence, the standard deviation σ_x is given by

$$\sigma_x = \sqrt{a}$$

8A.1.3 Expression for Expected Units Short per Cycle with Poisson Demand

For a Poisson variable x with mean a , the probability mass function is defined above in Equation 8A.1. Define the cumulative distribution function as

$$P_x(s) = \sum_{x_0=0}^s p_x(x_0)$$

Expected units short during a replenishment cycle can be calculated as follows:

$$\begin{aligned} \text{ESPRC} &= \sum_{x_0=s+1}^{\infty} (x_0 - s)p_x(x_0) = \sum_{x_0=s+1}^{\infty} (x_0 - s) \frac{a^{x_0} e^{-a}}{x_0!} \\ &= \sum_{x_0=s+1}^{\infty} x_0 \frac{a^{x_0} e^{-a}}{x_0!} - \sum_{x_0=s+1}^{\infty} s \frac{a^{x_0} e^{-a}}{x_0!} \\ &= \sum_{x_0=s+1}^{\infty} \frac{(x_0)(a)a^{x_0-1} e^{-a}}{(x_0)(x_0 - 1)!} - s[1 - P_x(s)] \\ &= a \sum_{x_0=s+1}^{\infty} \frac{a^{x_0-1} e^{-a}}{(x_0 - 1)!} - s[1 - P_x(s)] \end{aligned}$$

By changing the starting point for the index for the summation and adjusting the inner term, the first term of the previous expression can be rewritten as

$$a \sum_{x_0=s+1}^{\infty} \frac{a^{x_0-1} e^{-a}}{(x_0 - 1)!} = a \sum_{x_0=s}^{\infty} \frac{a^{x_0} e^{-a}}{(x_0)!} = a[1 - P_x(s - 1)]$$

leading to the following expression for ESPRC using only the reorder point s , the parameter a , and the cumulative distribution function.

$$\text{ESPRC} = a[1 - P_x(s - 1)] - s[1 - P_x(s)] \tag{8A.2}$$

As shown in Appendix III, we can use standard spreadsheet functions to calculate the ESPRC.

8A.2 Indifference Curves for Poisson Demand

8A.2.1 Case of $Q = 1$ and a B_2 Penalty

Suppose an order-up-to-level S is used, or equivalently, an order for one unit is placed when the inventory position drops to the level $S - 1$. In this system, the inventory position is effectively always at the level S . All outstanding orders at a time t must arrive by time $t + L$ and no order

placed after t can arrive by $t + L$. Therefore, the net (on-hand minus backorders) stock at time $t + L$ must be equal to the inventory position at time t minus any demand in t to $t + L$; that is,

$$(\text{Net stock at time } t + L) = S - (\text{demand in } L)$$

or

$$p_{NS}(n_0) = \Pr\{x = S - n_0\} \tag{8A.3}$$

where

$p_{NS}(n_0)$ = probability that the net stock at a random point in time takes on the value n_0
 x = total demand in the replenishment lead time

Furthermore, with Poisson demand, the probability that a particular demand requires backordering is equal to the probability that the net stock is zero or less; that is,

$$\Pr\{\text{a demand is not satisfied}\} = p_{NS \leq}(0)$$

Using Equation 8A.3, we have

$$\Pr\{\text{a demand is not satisfied}\} = p_{x \geq}(S)$$

The expected on-hand inventory (\bar{I}) at the end of the lead time is the expected *positive* net stock at that time,

$$\begin{aligned} \bar{I} &= \sum_{n_0=0}^S n_0 p_{NS}(n_0) \\ &= \sum_{n_0=0}^S n_0 p_x(S - n_0) \end{aligned}$$

where

$p_x(x_0)$ = probability that total time demand is x_0

Substituting, $j = S - n_0$, we have

$$\bar{I} = \sum_{j=0}^S (S - j) p_x(j)$$

The expected shortage costs per unit time (C_s) are

$$\begin{aligned} C_s &= (\text{cost per shortage}) \times (\text{expected demand per unit time}) \\ &\quad \times \Pr\{\text{a demand is not satisfied}\} \\ &= B_2 v D p_{x \geq}(S) \end{aligned}$$

Expected total relevant costs per unit time, as a function of S , are

$$\begin{aligned} \text{ETRC}(S) &= \bar{I}vr + C_s \\ &= vr \sum_{j=0}^S (S-j)p_x(j) + B_2vDp_{x \geq}(S) \end{aligned}$$

Notice the similarity with Equation 8.8 which is for $Q \geq 1$. Notice also that the ordering cost term in Equation 8.8 is not included here, because it drops out in the indifference curve solution. These indifference curves are obtained by equating $\text{ETRC}(S)$ to $\text{ETRC}(S+1)$ for a given value of S . In general, this gives, after simplification,

$$\frac{p_x(S)}{p_{x \leq}(S-1)} = \frac{r}{DB_2} \tag{8A.4}$$

However, because $Q = 1$, we have that

$$s = S - 1$$

Therefore, Equation 8A.4 can be written as

$$\frac{p_x(s+1)}{p_{x \leq}(s)} = \frac{r}{DB_2}$$

Equation 8.4 follows by recognizing that the lead time demand (x) has a Poisson distribution with mean \hat{x}_L .

8A.2.2 Case of a B_1 Cost Penalty

We assume that the cost B_1 is incurred only if the demand in the lead time exceeds the reorder point s . The expected total relevant costs per unit time associated with using a reorder point s are

$$\text{ETRC}(s) = vr \sum_{x_0=0}^s (s-x_0)p_x(x_0) + \frac{D}{Q}B_1p_{x>}(s)$$

Again, the fixed cost of ordering is independent of s . Equating $\text{ETRC}(s) = \text{ETRC}(s+1)$ for indifference between s and $s+1$ leads to

$$\frac{p_x(s+1)}{p_{x \leq}(s)} = \frac{Qvr}{DB_1}$$

Again, Equation 8.6 follows by recognizing that the lead time demand (x) has a Poisson distribution with mean \hat{x}_L .

8A.3 Decision Rule for the Size of the Last Replenishment

We are interested in the total remaining demand y for the item, where y is a random variable. We approximate its mean value $E(y)$ by the total demand \hat{y} implied by the deterministic representation

of the declining demand rate as a function of time (see Equations 8.17 and 8.18). In addition, we make the plausible assumption that y has a normal distribution (although the derivation could be modified to handle any other desired distribution) with standard deviation σ_y . Then, closely paralleling the P_2 service derivation in Section 6A.3 of the Appendix of Chapter 6, we have that the expected shortage over the lifetime

$$ES = \sigma_y G_u(k)$$

and we want

$$\frac{ES}{\hat{y}} = 1 - P_2$$

so that

$$G_u(k) = \frac{\hat{y}(1 - P_2)}{\sigma_y}$$

8A.4 Decision Rule for Disposal of Stock

Assuming a deterministic demand pattern (rate D), the situation is conceptually identical with that of a special opportunity to procure at a reduced unit cost (see Section 4.9.5 of Chapter 4). Let I be the current inventory level and W be the disposal amount. Then, as in Section 4.9.5, we look at costs out to time I/D —that is, when the current inventory would be depleted if no disposal took place. Under the option of no disposal, the costs would be

$$C_{ND} = \frac{I^2 vr}{2D}$$

If we dispose of W , then the remaining stock will run out at time $(I - W)/D$. Assuming an EOQ strategy after that, the relevant costs of the disposal decision out to time I/D are approximately

$$C_D = -gW + \left(\frac{I - W}{D}\right) \left(\frac{I - W}{2}\right) vr + \frac{W}{D} (\sqrt{2ADvr} + Dv)$$

As in Section 4.9.5, we select W to maximize $C_{ND} - C_D$. Using calculus, this leads to the decision rule of Equation 8.21.

8A.5 Decision Rule for Stocking versus Not Stocking

The demand rate per unit time D is given by

$$D = E(t)/E(i) \tag{8A.5}$$

Ignoring the minor effects of the non-unit-sized transactions, we have the same setting as in the derivation of the economic order quantity (Section 4.2 of Chapter 4). Therefore, from Equation 4.4 of Chapter 4 and Equation 8.25, the best order quantity, if the item is stocked, is

$$EOQ = \sqrt{\frac{2AE(t)}{vrE(i)}}$$

Moreover, from Equation 4.5 of Chapter 4 and including the c_s system cost, the best total relevant costs per unit time, if the item is stocked, are

$$\text{TRC}_s(\text{EOQ}) = \sqrt{2AE(t)vr/E(i)} + c_s \quad (8A.6)$$

If the item is not stocked, there is a cost A associated with each transaction which occurs, on the average, every $E(i)$ units of time. Therefore, the total relevant costs per unit time, if the item is not stocked, are

$$\text{TRC}_{ns} = A/E(i) \quad (8A.7)$$

Clearly, if c_s itself is larger than $A/E(i)$, then $\text{TRC}_{ns} < \text{TRC}_s$. This gives the condition of Equation 8.23. More generally, requiring

$$\text{TRC}_{ns} = \text{TRC}_s$$

and using Equations 8A.6 and 8A.7 leads to

$$A/E(i) < \sqrt{2AE(t)vr/E(i)} + c_s$$

or

$$2AE(t)vr/E(i) > \left[\frac{A}{E(i)} - c_s \right]^2$$

that is,

$$E(t)vr > \frac{E(i)}{2A} \left[\frac{A}{E(i)} - c_s \right]^2$$

References

- Adelson, R. 1966. The dynamic behavior of linear forecasting and scheduling rules. *Operational Research Quarterly* 17(4), 447–462.
- Aggarwal, P. and K. Moinzadeh. 1994. Order expedition in multi-echelon production/distribution systems. *IIE Transactions* 26(2), 86–96.
- Alscher, J., M. Kühn, and C. Schneeweiss. 1986. On the validity of reorder point inventory models for regular and sporadic demand. *Engineering Costs and Production Economics* 10(1), 43–55.
- Axsäter, S. 1996. Using the deterministic EOQ formula in stochastic inventory control. *Management Science* 42(6), 830–834.
- Bachman, T. C., P. J. Williams, K. M. Cheman, J. Curtis, and R. Carroll. 2016. PNG: Effective inventory control for items with highly variable demand. *Interfaces* 46(1), 18–32.
- Bose, S., A. Goswami, and K. S. Chaudhuri. 1995. An EOQ model for deteriorating items with linear time-dependent demand rate and shortages under inflation and time discounting. *Journal of the Operational Research Society* 46, 771–782.
- Brousseau, L. J. A. 1982. An inventory replenishment policy for the case of a linear decreasing trend in demand. *INFOR* 20(3), 252–257.
- Brown, R. G. 1977. *Materials Management Systems*. New York: Wiley-Interscience.
- Brown, R. G. 1981. *Confidence in All-Time Supply Forecasts*. Materials Management Systems, Inc., Norwich, Vermont.
- Cattani, K. D. and G. C. Souza. 2003. Good buy? Delaying end-of-life purchases. *European Journal of Operational Research* 146(1), 216–228.

- Cawdery, M. N. 1976. The lead time and inventory control problem. *Operational Research Quarterly* 27(4ii), 971–973.
- Chakravarty, A. K. and G. E. Martin. 1989. Discount pricing policies for inventories subject to declining demand. *Naval Research Logistics* 36, 89–102.
- Croston, J. D. 1972. Forecasting and stock control for intermittent demands. *Operational Research Quarterly* 23(3), 289–303.
- Croston, J. D. 1974. Stock levels for slow-moving items. *Operational Research Quarterly* 25(1), 123–130.
- Ekanayake, P. and J. Norman. 1977. Aspects of an inventory control study. *European Journal of Operational Research* 1, 225–229.
- Etienne, E. C. 1987. A simple and robust model for computing the service level impact of lot sizes in continuous and lumpy demand contexts. *Journal of Operations Management* 7(1), 1–9.
- Federgruen, A. and Y.-S. Zheng. 1992. An efficient algorithm for computing an optimal (r, Q) policy in continuous review stochastic inventory systems. *Operations Research* 40(4), 808–813.
- Feeney, G. J. and C. C. Sherbrooke. 1966. The $(S-1, S)$ inventory policy under compound Poisson demand. *Management Science* 12(5), 391–411.
- Fortuin, L. 1980. The all-time requirement of spare parts for service after sales—Theoretical analysis and practical results. *International Journal of Operations & Production Management* 1(1), 59–70.
- Gallego, G. 1996. Lecture Notes, Production Management.
- Haight, F. 1967. *Handbook of the Poisson Distribution*. New York: Wiley.
- Hariga, M. 1994. Two new heuristic procedures for the joint replenishment problem. *Journal of the Operational Research Society* 45(4), 463–471.
- Hariga, M. 1995. Effects of inflation and time-value of money on an inventory model with time-dependent demand rate and shortages. *European Journal of Operational Research* 81, 512–520.
- Harris, C. 2004. The plan for every part (PFEP). <http://www.lean.org>.
- Hollier, R. H., K. L. Mak, and C. L. Lam. 1995. Continuous review (s, S) policies for inventory systems incorporating a cutoff transaction size. *International Journal of Production Research* 88(10), 2855–2865.
- Johnson, J. 1962. On stock selection at spare parts stores sections. *Naval Research Logistics Quarterly* 9(1), 49–59.
- Johnston, F. R. and J. E. Boylan. 1996. Forecasting for items with intermittent demand. *Journal of the Operational Research Society* 47, 113–121.
- Karush, W. 1957. A queueing model for an inventory problem. *Operations Research* 5(5), 693–703.
- Kicks, P. and W. A. Donaldson. 1980. Irregular demand: Assessing a rough and ready lot size formula. *Journal of the Operational Research Society* 31(8), 725–732.
- Mitchell, C. R., R. A. Rappold, and W. B. Faulkner. 1983. An analysis of air force EOQ data with an application to reorder point calculation. *Management Science* 29(4), 440–446.
- Mitchell, G. 1962. Problems of controlling slow-moving engineering spares. *Operational Research Quarterly* 13, 23–39.
- Moinzadeh, K. 1989. Operating characteristics of the $(S-1, S)$ inventory system with partial backorders and constant resupply times. *Management Science* 35(4), 472–477.
- Moinzadeh, K. and C. P. Schmidt. 1991. An $(S-1, S)$ inventory system with emergency orders. *Operations Research* 39(2), 308–321.
- Natarajan, R. and S. K. Goyal. 1994. Safety stocks in JIT environments. *International Journal of Operations & Production Management* 14(10), 64–71.
- Ouyang, L., N. Yeh, and K. Wu. 1996. Mixture inventory model with backorders and lost sales for variable lead time. *Journal of the Operational Research Society* 47, 829–832.
- Popp, W. 1965. Simple and combined inventory policies, production to stock or to order. *Management Science* 11(9), 868–873.
- Porteus, E. L. 1985. Investing in reducing setups in the EOQ model. *Management Science* 31(8), 998–1010.
- Pourakbar, M., J. Frenk, and R. Dekker. 2012. End-of-life inventory decisions for consumer electronics service parts. *Production and Operations Management* 21(5), 889–906.
- Pourakbar, M., E. Van der Laan, and R. Dekker. 2014. End-of-life inventory problem with phaseout returns. *Production and Operations Management* 23(9), 1561–1576.

- Rabinowitz, G., A. Mehrez, C. Chu, and B. E. Patuwo. 1995. A partial backorder control for continuous review (r, Q) inventory system with Poisson demand and constant lead time. *Computers and Operations Research* 22(7), 689–700.
- Rohatgi, V. K. 1976. *An Introduction to Probability Theory and Mathematical Statistics*. New York: John Wiley & Sons.
- Rosenfield, D. 1989. Disposal of excess inventory. *Operations Research* 37(3), 404–409.
- Scala, N. M., J. Rajgopal, and K. L. Needy. 2013. A base stock inventory management system for intermittent spare parts. *Military Operations Research* 18(3), 63–77.
- Schultz, C. R. 1987. Forecasting and inventory control for sporadic demand under periodic review. *Journal of the Operational Research Society* 38(5), 453–458.
- Schultz, C. R. 1990. On the optimality of the $(S - 1, S)$ policy. *Naval Research Logistics* 37, 715–723.
- Shenstone, L. and R. J. Hyndman. 2005. Stochastic models underlying Croston's method for intermittent demand forecasting. *Journal of Forecasting* 24(6), 389–402.
- Shorrock, B. 1978. Some key problems in controlling component stocks. *Journal of the Operational Research Society* 29(7), 683–689.
- Silver, E. 1970. Some ideas related to the inventory control of items having erratic demand patterns. *Canadian Operations Research Society Journal* 8(2), 87–100.
- Smith, B. T. 1984. *Focus Forecasting: Computer Techniques for Inventory Control*. Essex Junction, VT: Oliver Wight Limited Publications, Inc.
- Smith, G. 1987. The inventory doctor: Obsolete inventory—Symptoms, treatment, & cure. *Production and Inventory Management Review* December 7, 32–35.
- Smith, P. 1977. Optimal production policies with decreasing demand. *European Journal of Operational Research* 1, 355–367.
- Snyder, R. D., J. K. Ord, and A. Beaumont. 2012. Forecasting the intermittent demand for slow-moving inventories: A modelling approach. *International Journal of Forecasting* 28(2), 485–496.
- Syntetos, A. A. and J. E. Boylan. 2001. On the bias of intermittent demand estimates. *International Journal of Production Economics* 71(1), 457–466.
- Taylor, C. 1961. The application of the negative binomial distribution to stock control problems. *Operational Research Quarterly* 12, 81–88.
- Teunter, R. H., A. A. Syntetos, and M. Z. Babai. 2011. Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research* 214(3), 606–615.
- Ting, P. and K. Chung. 1994. Inventory replenishment policy for deteriorating items with a linear trend in demand considering shortages. *International Journal of Operations & Production Management* 14(8), 102–110.
- Tyworth, J. E. and L. O'Neill. 1997. Robustness of the normal approximation of lead-time demand in a distribution setting. *Naval Research Logistics* 44(2), 165–186.
- Valadares Tavares, L. and L. Tadeu Almeida. 1983. A binary decision model for the stock control of very slow moving items. *Journal of the Operational Research Society* 34(3), 249–252.
- Ward, J. B. 1978. Determining reorder points when demand is lumpy. *Management Science* 24(6), 623–632.
- Watson, R. 1987. The effects of demand-forecast fluctuations on customer service and inventory cost when demand is lumpy. *Journal of the Operational Research Society* 38(1), 75–82.
- Wee, H.-M. 1995. A deterministic lot-size inventory model for deteriorating items with shortages and a declining market. *Computers and Operations Research* 22(3), 345–356.
- Wilcox, J. 1970. How to forecast “lumpy” items. *Production and Inventory Management Journal* 11(1, 1st Qtr), 51–54.
- Williams, B. W. 1982. Vehicle scheduling: Proximity priority searching. *Journal of the Operational Research Society* 33(10), 961–966.
- Williams, T. M. 1984. Special products and uncertainty in production/inventory systems. *European Journal of Operational Research* 15, 46–54.
- Yilmaz, C. 1992. Incremental order quantity for the case of very lumpy demand. *International Journal of Production Economics* 26, 367–371.
- Zheng, Y. S. 1992. On properties of stochastic inventory systems. *Management Science* 38(1), 87–103.

Chapter 9

Style Goods and Perishable Items

The developments in Chapters 4 through 8 were all based on the assumption of a demand pattern for an item continuing well into the future (possibly on a time-varying basis)—that is, “a going concern” with the opportunity to store inventory from one selling period to the next. In this chapter, we remove the ability to indefinitely store inventory, due to perishability or obsolescence. Decision situations where this type of framework is relevant include

1. The newsvendor: How many copies of a particular issue of a newspaper to stock?
2. The garment manufacturer (or retailer): What quantity of a particular style good should be produced (or bought for sale) prior to the short selling season?
3. The Christmas tree vendor: How many trees should be purchased to put on sale?
4. The cafeteria manager: How many hot meals of a particular type should be prepared prior to the arrival of customers?
5. The supermarket manager: How much meat or fresh produce should be purchased for a particular day of the week?
6. The administrator of a regional blood bank: How many donations of blood should be sought and how should they be distributed among hospitals?
7. The supplies manager in a remote region (e.g., medical and welfare work in northern Canada): What quantity of supplies should be brought in by boat prior to the long winter freeze-up?
8. The farmer: What quantity of a particular crop should be planted in a specific season?
9. The toy manufacturer: A particular product shows significant potential sales as a “fad” item. How many units should be produced on the one major production run to be made?

These are notoriously difficult decisions. Every Christmas, for instance, news reports focus on the inability to find certain extremely popular toys. At the same time, retailers are left with piles of toys that turned out to be less popular than expected. Many firms have been able to slash lead times by prepositioning raw materials ready for assembly, using air freight to expedite delivery, communicating orders electronically, and so on. Therefore, they can estimate demand closer to

the actual selling season. Unfortunately, they are still faced with considerable uncertainty about demand.

In the simplest situation, inventory cannot be carried forward at all from one period to the next. In this context, a single period is relevant for each replenishment decision, thus significantly simplifying the analysis (in part because we generally do not calculate a reorder point). This is in contrast with the earlier, more general, case where inventory is storable from one period to the next. In that context, the effects of a replenishment action can last for several periods and thus can interact with later replenishment actions. If we refer to the lifetime of an item as the length of time until it becomes obsolete or perishes, then the analysis becomes much more complicated as the lifetime increases.

In Section 9.1, we spell out in more detail the characteristics (and associated complexity) of the general style goods problem. The models presented in Sections 9.2 through 9.5 deal with progressively more realistic, but more complicated, versions of the problem. In Section 9.2, we are concerned with the case of a single product with but one replenishment (procurement or production) opportunity. This is followed in Section 9.3 by the important generalization to more than one product, thus permitting the aggregate viewpoint first emphasized in Chapter 2. Again, an exchange curve is a key by-product of the analysis. In Section 9.4, we examine the case when differentiation of the product can be delayed, a technique known as product postponement. Next, in Section 9.5, we provide some brief comments concerning the situation where there are several production periods with capacity restrictions prior to the selling season. Section 9.6 extends the discussion to the case where obsolescence is several periods in the future. Therefore, there are several opportunities to reorder before demand completely disappears. Section 9.7 deals with some other issues relevant to the control of style goods. The chapter is completed by a discussion in Section 9.8 concerning items subject to perishability.

9.1 Style Goods Problem

The main features of the style goods problem are:

1. There is a relatively short (typically no longer than 3 or 4 months) selling season with a well-defined beginning and end.
2. Buyers (at the stocking point) or producers have to commit themselves to a large extent, in terms of how much of each SKU to order or produce, prior to the start of the selling season.
3. There may be one or more opportunities for replenishment after the initial order is placed. Such replenishment actions may be taken prior to the selling season (if the forecast of demand has risen appreciably) or during the early part of the selling season itself (if actual demand to date indicates that the original forecast was considerably low). Occasionally, the season extends several periods in the future; we will discuss such a situation in Section 9.6.
4. Forecasts prior to the season include considerable uncertainty stemming from the long period of inactivity (no sales) between seasons. During this inactive period, the economic conditions or style considerations may have changed appreciably. Consequently, forecast revisions tend to be heavily influenced by firm orders or sales to date in the current season, far more than in the case of reasonably stable items (as discussed in Chapter 3).
5. When the total demand in the season exceeds the stock made available, there are associated underage costs. These may simply be lost profit if the sales are foregone. On the other hand, they may represent the added costs of expediting (e.g., a special order by air freight) or acquiring the material from a competitor at a high unit value.

6. When the total demand in the season turns out to be less than the stock made available, overage costs result. The unit salvage value of leftover inventory at the end of a season is likely to be quite low (well below the unit acquisition cost). It is quite expensive to carry stock over to the next season; moreover, there is no guarantee that such leftover stock will even sell at that time. Markdowns (i.e., reduced selling prices) or transfers (between locations) may be used to avoid leftover stock. (Incidentally, the style goods situation can be thought of as a special case of perishability where there is not physical deterioration but, instead, a marked deterioration in the economic value of the goods as of a particular point in time.)
7. Style goods products are often substitutable. Depending on how finely one defines a *style*, a stockout in one SKU does not necessarily mean a lost sale, because the demand may be satisfied by another product.
8. Sales of style goods are usually influenced by promotional activities, and space allocation in the store, among other things.

In the simplest form of the style goods problem, wherein we ignore the last two points made above as well as markdowns and transfers, the remaining decision variables are the timing and magnitude of the replenishment quantities. To illustrate the complexity of even this simplified version, we need only refer to the early work of Wadsworth (1959). He considered the case of a style good sold in a fall season with raw material acquisition required in the spring, summer production of unlimited capacity, and a limited amount of fall production (after the total season demand is known exactly) being possible but at a higher unit cost than the summer production. The complex nature of Wadsworth's solution to this problem indicates that an exact analysis of a more realistic style goods problem is out of the question. For this reason, as elsewhere in this book, for all but the simplest problems we advocate the use of heuristic methods designed to generate reasonable solutions.

9.2 Simplest Case: Unconstrained, Single-Item, Newsvendor Problem

9.2.1 Determination of the Order Quantity by Marginal Analysis

Let us consider the situation faced by the owner of a newsstand. Each day he has to decide on the number of copies of a particular paper to stock. There is an underage cost, c_u , associated with each demand that cannot be met, and an overage cost, c_o , associated with each copy that is not sold. Suppose the underage cost was exactly equal to the overage cost. Then it seems intuitively reasonable that he would want to select the number (Q) of copies to purchase such that there was a 50% chance of the total demand being below Q and a 50% chance of it being above Q . But what about the case where $c_u = 2c_o$ or $c_u = 3c_o$? It turns out that his decision rule should be to select the Q^* value that satisfies

$$p_{u<}(Q^*) = \frac{c_u}{c_u + c_o} \quad (9.1)$$

where $p_{x<}(x_0)$ is the probability that the total demand x is less than the value x_0 .

Notice that when $c_u = c_o$ Equation 9.1 gives

$$p_{u<}(Q^*) = 0.5$$

which is the intuitive result discussed above.

Table 9.1 Marginal Analysis of the News Vendor Problem

Specific Cost Element	Probability That the Specific Cost Element Is Incurred (or Avoided) by the Acquisition of the Qth Unit	Expected Value of the Specific Cost Element Associated with the Qth Unit (Product of Previous Two Columns)
Overage, c_o	$\Pr\{\text{Demand is less than } Q\} = p_{x <}(Q)$	$c_o p_{x <}(Q)$
Underage, c_u	$\Pr\{\text{Demand is greater than or equal to } Q\} = 1 - p_{x <}(Q)$	$c_u [1 - p_{x <}(Q)]$

When $c_u = 2c_o$, we have from Equation 9.1

$$p_{x <}(Q^*) = \frac{2c_o}{2c_o + c_o} = \frac{2}{3}$$

that is, he should select the order quantity such that the probability of the total demand being smaller than the order quantity is two-thirds.

Now let us reflect on how Equation 9.1 is developed. The simplest derivation is a marginal analysis, as first used in Section 6A.1 of the Appendix of Chapter 6. The argument proceeds as follows.

Consider the Q th unit purchased. It will be sold if, and only if, the total demand equals or exceeds Q ; otherwise, an overage cost will be incurred for this Q th unit. However, if the demand equals or exceeds Q , we have avoided an underage cost by having the Q th unit available. In other words, we have the analysis shown in Table 9.1.

Consider a particular Q th unit. If the expected overage cost associated with acquiring it exceeded the expected saving in underage costs, we would not want to acquire that unit. In fact, we might not want to acquire the $(Q - 1)$ st, the $(Q - 2)$ nd, and so forth. On the other hand, if the expected saving in underage costs exceeded the expected overage cost associated with acquiring the Q th unit, we would want to acquire it. However, we might also wish to acquire the $(Q + 1)$ st, the $(Q + 2)$ nd, etc. The last unit (Q^*) we would want to acquire is that one where the expected overage cost incurred exactly equaled the expected underage cost saved; that is, Q^* must satisfy

$$c_o p_{x <}(Q^*) = c_u [1 - p_{x <}(Q^*)]$$

or

$$p_{x <}(Q^*) = \frac{c_u}{c_u + c_o}$$

which is Equation 9.1.

The solution of Equation 9.1 is shown graphically in Figure 9.1 where we have plotted the cumulative distribution of demand, $p_{x <}(x_0)$. Locate the value of $c_u/(c_u+c_o)$ on the vertical axis, come across horizontally to the curve, then move down vertically to read the corresponding x_0 , which is the Q^* value. There is always a solution because the fraction lies between zero and unity, which is the range covered by the cumulative distribution.

We have just seen that the marginal cost analysis approach leads to the decision rule of Equation 9.1 in a rather straightforward fashion. However, it does not give us an expression for evaluating the expected costs when any particular Q value is used. The alternative approach of minimizing total cost does, but the analysis is more complicated.

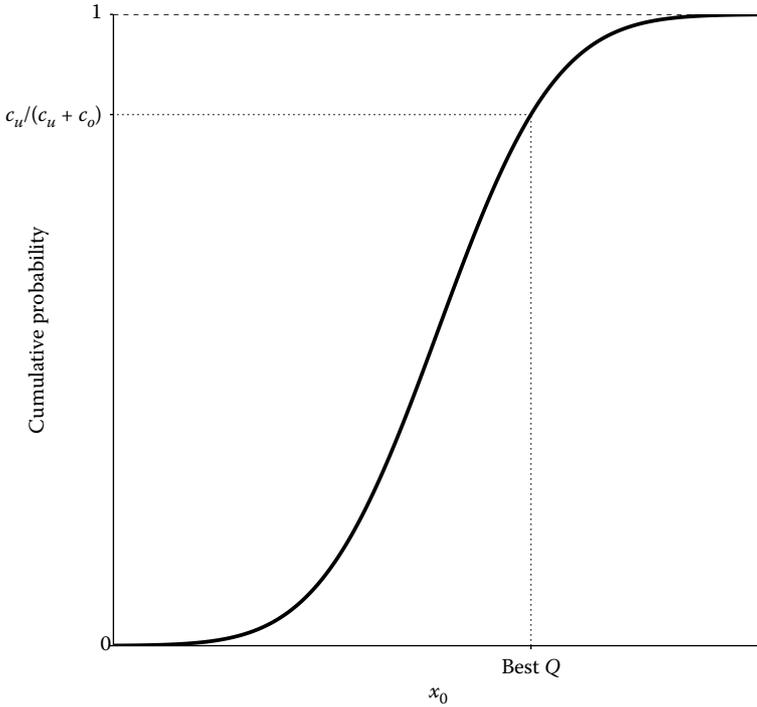


Figure 9.1 Graphical solution of the newsvendor problem.

9.2.2 An Equivalent Result Obtained through Profit Maximization

The above discussion has centered on cost minimization. Instead, let us now approach the problem from a profit maximization standpoint. Let

v = acquisition cost, in dollars/unit

p = revenue per sale (i.e., the selling price), in dollars/unit

B (or B_2v) = penalty (beyond the lost profit) for not satisfying demand, in dollars/unit (B_2 , as in Chapter 6, is the fraction of unit value that is charged per unit of demand unsatisfied)

g = salvage value, in dollars/unit

As earlier, we let

Q = quantity to be stocked, in units

and $p_{x <}(x_0)$ represent the cumulative distribution of total demand, the probability that total demand x takes on a value less than x_0 .

Then, as shown in Section 9A.1 of the Appendix of this chapter, the decision rule is to select the Q^* that satisfies

$$p_{x <}(Q^*) = \frac{p - v + B}{p - g + B} \tag{9.2}$$

This result is identical with that of Equation 9.1 when we realize that the cost of underage is given by

$$c_u = p - v + B$$

and the cost of overage is

$$c_o = v - g$$

Because the cumulative distribution function is nondecreasing as its argument increases, we can make the following intuitively appealing deductions from Equation 9.2.

1. If the sales price, cost per unit short, or salvage value increases, the best Q increases.
2. If the acquisition cost per unit increases, the best Q decreases.

As shown in Section 9A.1 of the Appendix of this chapter, the expected profit is

$$E[P(Q)] = (p - g)\hat{x} - (v - g)Q - (p - g + B)ES \quad (9.3)$$

where \hat{x} is expected demand and ES is the expected number of units short, or expected shortages. If we let EO represent expected units left over, we can also represent the expected profit as

$$\begin{aligned} E[P(Q)] &= (p - v)\hat{x} - (v - g)EO - (p - v + B)ES \\ &= (p - v)\hat{x} - c_o EO - c_u ES \end{aligned} \quad (9.4)$$

This rearrangement shows the expected profit as a “perfect information” expected profit term minus an overstocking loss and an understocking loss. The first term of Equation 9.4 is the expected profit we would receive if we knew what demand would be and chose the quantity to exactly meet that demand.

9.2.3 Case of Normally Distributed Demand

The result of Equation 9.1 or 9.2 is applicable for any distribution of total demand. Again, it is worthwhile, as in Chapter 6, to look at the special case of normally distributed demand. First, such a distribution tends to be applicable in a significant number of situations. Second, as we will see, sensitivity analysis (effects of incorrectly specifying the order quantity Q) can be performed rather easily when demand is normal, in contrast with the case of a general probability distribution.

Suppose the demand is normally distributed with a mean of \hat{x} and a standard deviation of σ_x and we define

$$k = \frac{Q - \hat{x}}{\sigma_x} \quad (9.5)$$

This is analogous to what we did in Chapter 6. Then, as shown in Section 9A.1 of the Appendix of this chapter, the decision rule of Equation 9.2 transforms to

$$\text{Select } k \text{ such that } p_{u \geq}(k) = \frac{v - g}{p - g + B} \quad (9.6)$$

where $p_{u \geq}(k)$ is the probability that a unit normal variable takes on a value of k or larger; this is a widely tabulated function (see, e.g., Table II.1 of Appendice II) first introduced in Chapter 6. Then, from Equation 9.5, set

$$Q = \hat{x} + k\sigma_x \quad (9.7)$$

Furthermore, in this case, as shown in the Appendix of this chapter, we obtain a relatively simple expression for expected profit as a function of Q , namely,

$$E[P(Q)] = (p - g)\hat{x} - (v - g)Q - (p - g + B)\sigma_x G_u\left(\frac{Q - \hat{x}}{\sigma_x}\right) \quad (9.8)$$

where $G_u(u_0)$ is a tabulated function of the unit normal variable (see Table II.1 of Appendice II) first introduced in Chapter 6. It can also be shown that (see Section A.1 of the Appendix of this chapter):

$$\frac{E[P(Q^*)]}{\sigma_x} = \frac{(p - v)\hat{x}}{\sigma_x} - (p - g + B)f_u(k^*) \quad (9.9)$$

9.2.3.1 Numerical Illustration

Sales of ski equipment and apparel depend heavily on the snow conditions. A small discount retailer (called The Ski Bum) of ski equipment near some of the large ski areas of New Hampshire and Vermont faces the difficult task of ordering a particular line of ski gloves in the month of May well before the ski season begins. The selling price for these gloves is \$50.30 per pair, while the cost to the retailer is only \$35.10. The retailer can usually buy additional gloves from competitors at their retail price of \$60. Gloves left over at the end of the short selling season are sold at a special discount price of \$25 per pair. The owner and manager of the store assumes that demand is approximately normally distributed with a mean of 900 gloves and a standard deviation of 122 gloves.* Using our notation

- $p = \$50.30/\text{pair}$
- $v = \$35.10/\text{pair}$
- $g = \$25.00/\text{pair}$
- $B = (\$60.00 - 50.30) \text{ or } \$9.70/\text{pair}$

From Equation 9.6, we want

$$p_{u \geq}(k) = \frac{35.10 - 25}{50.30 - 25 + 9.70} = 0.288$$

* One way of arriving at an estimate of the standard deviation would be to specify the probability that total demand exceeded a certain value (other than the mean); for example, $\text{prob}(x \geq 1,100) = 0.05$ leads to the standard deviation of 122 pairs. For x normally distributed,

$$\text{Pr}(x \geq 1,100) = p_{u \geq}\left(\frac{1,100 - \hat{x}}{\sigma_x}\right) = p_{u \geq}\left(\frac{200}{\sigma_x}\right)$$

But from the table in Appendix II

$$p_{u \geq}(1.645) = 0.05$$

Therefore,

$$\begin{aligned} \frac{200}{\sigma_x} &= 1.645 \\ \sigma_x &= 122 \end{aligned}$$

Table 9.2 Sensitivity of Expected Profit to Q Value Selected in Example

Q	<i>k</i>	<i>G(k)</i>	Percent Deviation from Best Q (Namely, 968)	<i>E[P(Q)]</i>	Percent of Best Possible Profit
700	-1.64	1.6611	27.7	8,607	71.0
800	-0.82	0.9360	17.4	10,690	87.5
900	0.00	0.3989	7.0	11,980	98.0
925	0.20	0.3069	4.4	12,126	99.2
930	0.25	0.2863	3.9	12,150	99.4
968	0.56	0.1799	0.0	12,220	100.0
1,000	0.82	0.1160	3.3	12,170	99.6
1,030	1.07	0.0728	6.4	12,050	98.6
1,100	1.64	0.0211	13.6	11,570	94.7
1,200	2.46	0.0023	24.0	10,640	87.1

From the table in Appendix II, we have

$$k = 0.56$$

Therefore, Equation 9.7 gives

$$Q = 900 + 0.56(122) = 968.3, \text{ say } 968 \text{ pairs}$$

The use of Equation 9.8 gives the expected profit for various values of *Q*; the results are shown in Table 9.2. It is interesting to note that by following the temptation of ordering the expected demand—namely, *Q* = 900—we would be led to a 2.0% reduction in expected profit.

Lau (1980) has shown that, for any demand distribution, to maximize the probability of achieving a target profit, *T*, one should choose *Q** such that $Q^* = T / (p - v)$. For the example above, to maximize the probability of achieving a profit of \$10,690, set *Q** = 703, whereas to maximize the probability of achieving a profit of \$12,220, set *Q** = 804. Note that these order quantities are significantly different from those that provide the same expected profit (from Table 9.2). The optimal probability that profit achieves (or exceeds) the optimal expected profit target is of course $1 - p_{x <}(Q^*)$. See also Sankarasubramanian and Kumaraswamy (1983).

9.2.4 Case of a Fixed Charge to Place the Order

Where there is a high enough fixed setup cost *A* to place any size order, the best strategy may be to not order anything at all. The decision would be based on comparing the cost of not ordering at all with the expected profit (ignoring the fixed cost as we did earlier) under use of the *Q** value:

$$\begin{aligned} &\text{If } E[P(Q^*)] > A - B\hat{x}, \text{ order } Q^* \\ &\text{If } E[P(Q^*)] < A - B\hat{x}, \text{ order nothing} \end{aligned}$$

If there is inventory on hand at the start of the period, the policy of ordering Q^* should be seen as ordering up to Q^* instead. Observe then that there is a threshold value implied in the decision rule. If A is large enough, do not order. Otherwise, order up to Q^* . Said another way, if beginning inventory is relatively high, the penalty cost associated with not ordering will be relatively low, so do not order. However, if inventory is very low, order up to Q^* . This policy is a single-period version of the (s, S) policy introduced in Chapter 7. See Porteus (1990). As discussed above, we do not advocate explicitly determining s . Rather, for any given initial inventory level, we compare the expected profit of not ordering with the expected profit of using the best order-up-to-level, given that we order.

9.2.5 Case of Discrete Demand

Strictly speaking, when we are dealing with discrete (integer) units of demand and purchase quantity, it is unlikely that there is an integer Q value that exactly satisfies Equation 9.1 or 9.2. To illustrate, suppose we have an item where

$$\frac{c_u}{c_u + c_o} = 0.68$$

and the probability distribution of demand is as shown in the second row of the following table (the third row is derived from the second row):

Number of Units, x_0	1	2	3	4	5	Total
$p_x(x_0) = \Pr(x = x_0)$	0.2	0.3	0.2	0.2	0.1	1
$p_{x <}(x_0) = \Pr(x < x_0)$	0	0.2	0.5	0.7	0.9	–

We can see from the third row that there is no Q that would give $p_{x <}(Q) = 0.68$; the nearest values are $p_{x <}(3) = 0.5$ and $p_{x <}(4) = 0.7$. As shown in Section 9A.1 of the Appendix of this chapter, when discrete units are used, the best value of Q is the smallest Q value that satisfies

$$p_{x \leq}(Q) \geq \frac{c_u}{c_u + c_o} = \frac{p - v + B}{p - g + B} \tag{9.10}$$

9.2.5.1 Numerical Illustration

Consider the case of a Mennonite farmer who raises cattle and sells the beef at the weekly farmer’s market in Kitchener-Waterloo, Ontario. He is particularly concerned about a specific expensive cut of meat. He sells it in a uniform size of 5 kg at a price of \$30. He wants to know how many 5-kg units to take to market. Demand is not known with certainty; all that he has available are frequency counts of the total 5-kg units demanded in each of the last 20 market sessions. These are chronologically as follows: 1, 3, 3, 2, 2, 5, 1, 2, 4, 4, 2, 3, 4, 1, 5, 2, 2, 3, 1, and 4. Frequency counts of how many 1’s, how many 2’s, etc. occurred in the 20 weeks are shown in the second row of Table 9.3. The third row is a set of estimates of the probabilities of the various sized demands based on the assumption that the historical occurrences are representative of the probabilities. For

Table 9.3 Demand Distribution for Farmer’s Meat Example

1. Demand in a session x_0	1	2	3	4	5	Total
2. Frequency count $N(x_0)$	4	6	4	4	2	20
3. $p_x(x_0) \approx \frac{N(x_0)}{20}$	0.2	0.3	0.2	0.2	0.1	1.0
4. $p_{x \leq}(x_0)$	0.2	0.5	0.7	0.9	1.0	–

example, in 6 out of 20 weeks, the total demand has been 2 units; therefore, the probability of a total demand of 2 units is approximately 6/20 or 0.3.

The farmer stores the processed beef in a freezer on his farm. He estimates the value of a 5-kg unit to be \$19 after he removes it from the freezer and gets it to the market. Once the beef is brought to market, he chooses not to take it home and refreeze it. Therefore, any leftover beef at the market he sells to a local butcher at a discount price of \$15 per 5-kg unit. Finally, other than the foregone profit, he feels that there is no additional cost (e.g., loss of goodwill) associated with not satisfying a demand.

According to our notation

$$p = \$30/\text{unit}$$

$$v = \$19/\text{unit}$$

$$B = 0$$

$$g = \$15/\text{unit}$$

Therefore,

$$\frac{p - v + B}{p - g + B} = \frac{30 - 19 + 0}{30 - 15 + 0} = 0.733$$

The use of the fourth row of Table 9.3, together with the decision rule of Equation 9.10, shows that he should bring four 5-kg units to market. It is interesting to note that the average or expected demand is

$$\begin{aligned} \hat{x} &= \sum_{x_0} x_0 p_x(x_0) = (1)(0.2) + 2(0.3) + 3(0.2) + 4(0.2) + 5(0.1) \\ &= 2.7 \text{ units} \end{aligned}$$

Therefore, because of the economics, the best strategy is to take (4 – 2.7) or 1.3 units above the average demand.

For other non-normal demand distributions, see Scarf (1958), Gallego and Moon (1993), and Moon and Choi (1995). In particular, if the demand distribution is not known, but managers have some sense of the mean and variance of demand, Scarf (1958) and Gallego and Moon (1993) show that using the following formula maximizes the expected profit against the worst possible distribution of demand:

$$Q^* = \hat{x} + \frac{\sigma_x}{2} \left(\sqrt{\frac{p/v - 1}{1 - g/v}} - \sqrt{\frac{1 - g/v}{p/v - 1}} \right) \tag{9.11}$$

For the numerical example given for the normal distribution, this formula yields an order quantity of 925, versus 968 when the distribution is known to be normal. We can see from Table 9.2 that the cost penalty of using 925 is quite small.

9.3 Single-Period, Constrained, Multi-Item Situation

Now we consider the situation of more than one type of SKU to be stocked for a single period's demand. It should be emphasized that the model to be developed here is based on the assumption of independent demands for the various products. However, there is a space or budget constraint on the group of items. Examples include the following:

1. Several different newspapers sharing a limited space or budget in a corner newsstand.
2. A buyer for a style goods department of a retail outlet who has a budget limitation for a group of items.
3. The provisioning of supplies or spare parts on a spacecraft, submarine, or the like, prior to a long mission without the possibility of resupply.
4. The repair kit taken by a maintenance crew on a routine visit to an operating facility.

We illustrate the solution procedure for the case of a restriction on the total dollar value of the units stocked. The procedure is quite similar if the constraint is on space, capacity, or some other resource. We discuss how one would adapt the approach for some other constrained resource. First, some notation is required. Let

- n = number of different items (SKUs) involved (numbered $1, 2, 3, \dots, n$)
- $iP_{x < (x_0)}$ = Pr{total demand for item i is less than x_0 }
- v_i = acquisition cost of item i , in \$/unit
- p_i = selling price of item i , in \$/unit
- B_i = penalty (beyond the lost profit) for not satisfying demand of item i , in \$/unit
- g_i = salvage value of item i , in \$/unit
- W = budget available for allocation among the stocks of the n items, in dollars

We are assuming the budget is constrained, and that a manager wishes to use up the total budget to maximize the total expected profit of the n items. If this is not necessarily the case, each item i 's unconstrained optimal Q_i^* should be ascertained independently using Equation 9.2. If the total $\sum Q_i^* v_i < W$, then the Q_i^* s should be used. If not, the constrained approach must be used to find the Q_i s.

As shown in Section 9A.2 of the Appendix of this chapter, a convenient way to solve this problem is through the use of a Lagrange multiplier. This approach is useful for multi-item problems that are connected through some common resource constraint. In this case, it is relatively easy to solve each single-item problem using Equation 9.2, but of course this leads to a solution that exceeds the budget constraint. The Lagrange multiplier approach works by charging a penalty cost M per unit of resource consumed beyond the budget. In this case, the unit of resource is a dollar of the budget.

The resulting decision procedure is*

Step 1 Select an initial positive value of the multiplier, M .

Step 2 Determine each Q_j ($i = 1, 2, 3, \dots, n$) to satisfy

$$i p_{x <}(Q_j) = \frac{p_i - (M + 1)v_i + B_i}{p_i - g_i + B_i} \text{ and } Q_j \geq 0 \tag{9.12}$$

$$\sum_{i=1}^n Q_i v_i$$

with W .

If $\sum Q_i v_i \approx W$, stop. The optimal solution has been obtained.

If $\sum Q_i v_i < W$, return to Step 2 with a smaller value of M .

If $\sum Q_i v_i > W$, return to Step 2 with a larger value of M .

The multiplier has an interesting economic interpretation; it is the value (in terms of increased total expected profit) of adding one more dollar to the available budget, W . We revisit this interpretation in the context of the numerical illustration below.

It is important to recognize that as different values of M are used, we are actually tracing out an exchange curve of total expected profit versus total allowed budget (investment in stock). This is illustrated in Figure 9.2, which is an exchange curve for the numerical illustration to be presented momentarily.

Other research on the multiproduct problem includes Nahmias and Schmidt (1984), Lau and Lau (1988, 1994, 1996b), Li et al. (1990, 1991), and Jönsson and Silver (1996). Erlebacher (2000) develops simple expressions for certain special cases such as when all items follow a uniform distribution or all items follow the same demand distribution. Lau and Lau (1996b) show that if demand can never be zero, or has a very long left tail, it may be impossible to solve Equation 9.11 for certain values of M . Therefore, common search methods, such as the one given above, may not work. They provide a more complex method that works for any distribution.

The above formulation assumes that the individual-item stocking decisions have independent effects on the objective function to be maximized or minimized. Smith et al. (1980) have addressed the following more complicated situation, often called the “repair kit problem.” Consider the kit of spare parts stocked by a repair crew visiting an off-site location. A particular repair job may require more than one spare part. An appropriate measure of effectiveness is now the fraction of jobs completed without stockout, as opposed to the fraction of needed individual spare parts that are provided. In addition, Mamer and Smith (1982) extend the analysis to the situation where part demands are no longer independent. See also Graves (1982), Hausman (1982), Scudder and March (1984), Mamer and Smith (1985), Mamer and Shogan (1987), and Moon et al. (2000).

* If the constraint is, instead, of a production nature such that no more than W units can be produced, then the decision procedure is modified in two ways:

1. Equation 9.12 is changed to

$$i p_{x <}(Q_j) = \frac{p_i - v_i - M + B_i}{p_i - g_i + B_i}$$

2. In Step 3, we compare $\sum_{i=1}^n Q_i$ with W .

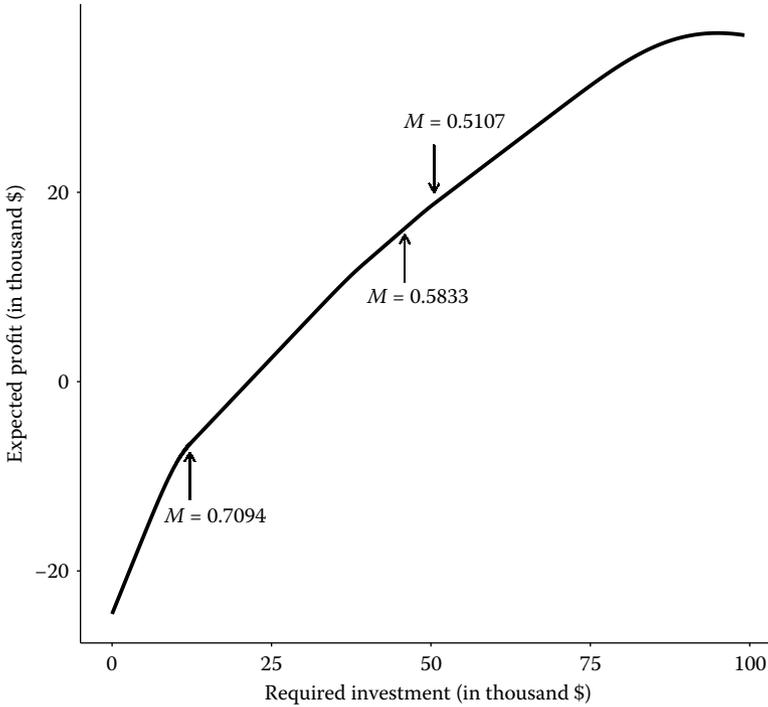


Figure 9.2 Exchange curve for the multi-item, single-period situation.

9.3.1 Numerical Illustration

Suppose The Ski Bum was faced with decisions on four items rather than just the one illustrated in Section 9.2.3. The manager is willing to accept that in each case total demand is normally distributed. For the case of normally distributed demand, Equation 9.12 becomes

$$p_{u \geq}(k_i) = \frac{(M + 1)v_i - g_i}{p_i - g_i + B_i} \tag{9.13}$$

where, as earlier, $p_{u \geq}(k_i)$ is the probability that a unit normal variable takes on a value of k_i or larger (tabulated in Appendix II) and

$$Q_i = \max(0, \hat{x}_i + k_i \sigma_i) \tag{9.14}$$

where

\hat{x}_i = forecast demand for item i

σ_i = standard deviation of the demand for item i (here we have simplified the notation somewhat to avoid a subscript)

The relevant parameter values are estimated to be as shown in the top six rows of Table 9.4. The manager has a budget of \$70,000 to allocate among these four items. She would like to know how to do this. She is also quite interested in the profit impact if the budget is changed somewhat.

Table 9.4 Numerical Illustration of Multi-Item, Single-Period Problem

	<i>Item 1</i>	<i>Item 2</i>	<i>Item 3</i>	<i>Item 4</i>	<i>Total</i>
p_i \$/unit	50.30	40.00	32.00	6.10	–
v_i \$/unit	35.10	15.00	28.00	4.80	–
g_i \$/unit	25.00	12.50	15.10	2.00	–
B_i \$/unit	9.70	0	10.30	1.50	–
\hat{x}_i units	900	800	1,200	2,300	–
σ_i units	122	200	170	200	–
Using $M = 0$ (i.e., no budget restriction):					
Q_i from Equations 9.12 and 9.13	968	1,067	1,211	2,300	–
$E[P(Q_i)]$ from Equation 9.8	12,222	19,100	2,959	2,543	\$36,824
$Q_i v_i$ \$	33,978	16,006	33,907	11,040	\$94,930
Using $M = 0.4$:					
Q_i	840	900	995	2,099	–
$E[P(Q_i)]$	11,325	18,660	1,613	2,334	\$33,932
$Q_i v_i$	29,470	13,495	27,861	10,074	\$80,900
Using $M = 0.6$:					
Q_i	750	841	0 ^a	0 ^a	–
$E[P(Q_i)]$	9,723	18,224	–12,360	–3,450	\$12,137
$Q_i v_i$	26,331	12,630	0	0	\$38,951
Using $M = 0.5093$ ^b :					
Q_i	798	867	694	1,995	–
$E[P(Q_i)]$	10,652	18,438	–2,439	2,104	\$28,755
$Q_i v_i$	27,996	13,006	19,430	9,575	\$70,007

^a The right-hand side of Equation 9.12 exceeds unity, implying that k_i (or equivalently Q_i) should be set at its lowest possible value. The lowest value for Q_i is zero from Equation 9.14.

^b By trial and error, it was found that this value of M was required to produce a total used budget of approximately 70,000.

The calculations for several values of the multiplier M are shown in the lower part of Table 9.4 and the results are shown graphically in Figure 9.2. If there is no budget constraint, all items would be stocked, a budget of \$94,930 would be needed, and the expected profit would be \$36,824. We can see that the expected total profit is \$28,755 when the proposed budget of approximately \$70,000 is used. However, increasing the budget could lead to a substantial improvement in the

total profit, a useful piece of information for the owner. The far right part of the curve shows that the total expected profit would start to decrease if the budget was too high and the manager was required to invest the whole amount in stocks of the four items (a negative value of M applies in this region). Three significant values of M are indicated in Figure 9.2. These values correspond to points at which the critical fractile (Equation 9.13) for one of the products equals zero, indicating it is unattractive to produce and stock that product. As shown in Table 9.4, a value of $M \approx 0.5093$ generated a set of order quantities that consumed the proposed budget of \$70,000. As stated in the previous section, this value of M is the value in terms of expected profit of adding one more dollar to the budget; another dollar of budget would generate approximately \$0.51 in expected profit. Another important set of managerial insights that can be gleaned from this problem formulation relates to comparisons across the items competing for the same budget. Notice that when the budget is quite constrained some products are not sufficiently attractive to order at all. For example, at $M = 0.6$ (leading to a budget of \$38,951), items 3 and 4 are not ordered at all. See Problem 9.26.

9.4 Postponed Product Differentiation

Because of the enormous variety and stiff delivery time requirements characterizing many style goods, such as fashion apparel, firms have been developing creative ways to reduce costs and improve delivery performance. One firm that has generated much interest is Benetton. By changing the order of production—their revised system switched the order from dye then knit, to knit then dye—Benetton was able to gain dramatic operational benefits. By switching the order of operations, Benetton was able to produce “plain” sweaters and then later dye them, only committing them to their final product form when more accurate demand information was available. This approach is known as delayed (or postponed) product differentiation or *postponement*.

Lee and Tang (1998) develop formal models to show when operations reversal is desirable. They show that when the major source of demand uncertainty is due to the option mix, but the total demand for all options is fairly stable, then the sequence of operations should have one or more of the following properties:

1. The first operation has the longer lead time if the value added of the two operations are roughly equal.
2. The value added of the first operation is less than the second operation, if the lead times are the same.
3. The first operation takes longer to perform, but the value added per unit of lead time for the first operation is smaller than or equal to the value added per unit of lead time for the second stage.

They provide several other properties, and they note that if total demand is highly variable, then the operation order suggested by these properties should be exactly reversed. See also Swaminathan and Tayur (1998), Fisher et al. (1999), and Randall and Ulrich (2001) for other examples of using the ideas of postponed product differentiation to manage multiple end items.

To adopt a postponement strategy, a firm typically manufactures a semifinished or base product that can later be finished into a final product. The Benetton setting describes such an example where the base product is the un-dyed sweater. This may require a redesign of the product and/or the process. Lee et al. (1993) discuss such a redesign effort. Ramdas (2003) provides

a broader discussion about managing product variety, including a discussion of manufacturing process capabilities that are essential in adopting postponement strategies.

There are two key factors that drive the potential operational improvement obtained by postponing product differentiation. First, by delaying some of the manufacturing value-add, the per-unit overstocking cost is reduced. Second, by delaying the commitment of the product into its final form, a firm can improve performance, either reducing safety stock or improving the level of service. We explore each of these two factors next, providing a numerical illustration in each case of the performance improvement obtained.

9.4.1 Value of Delayed Financial Commitment

We use the subscript 0 to represent the base product (e.g., the un-dyed sweater in the Benetton setting discussed above). Let t represent the additional transformation cost to convert a base product into a finished product. This means the cost to acquire (or produce) a base product and then transform that base product is $v_0 + t$. As before, let g denote the salvage value for the finished product. Let g_0 denote the salvage value for the base product. Here, we assume that the base product can be transformed into the finished item at the time demand is realized. This means that a firm will stock out when supply of the base product is insufficient and will incur the transformation cost only for products that are sold. In Section 9.5, we discuss the substantially more complicated situation when the transformation decision must be made before final demand occurs but after the firm has updated their demand forecast.

With these assumptions, the cost of understocking is $c_u = p - v_0 - t + B$ while the cost of overstocking is $v_0 - g_0$. As in Section 9.2.2, let EO, ES represent expected leftover units and expected units short, respectively. Letting Q_0 represent the quantity of base product acquired, the expected profit is

$$E[P(Q_0)] = (p - v_0 - t)\hat{x} - (v_0 - g_0)EO - (p - v_0 - t + B)ES \quad (9.15)$$

Following Equation 9.1, we get that the best quantity satisfies:

$$p_{u <} (Q_0^*) = \frac{c_u}{c_u + c_o} = \frac{p - v_0 - t + B}{p - g_0 - t + B} \quad (9.16)$$

Whether or not there is a benefit to the delayed financial commitment in this single-item setting depends on the relationship between the costs in the original system and this new system. To see why there is potential for savings, let us first simplify to the case where the total unit cost is the same in the two systems, $v = v_0 + t$, and the two salvage values are the same $g_0 = g$. Note that these relationships may not hold in general, but this simplification allows us to understand the source of the potential savings. Under these assumptions, the only difference between the two systems is that the cost of overstocking is lower in the postponed system (v_0 instead of $v_0 + t$) since the cost t is not incurred unless the demand is realized.

9.4.1.1 Numerical Illustration

Suppose The Ski Bum has historically purchased and sold a particular ski/binding package where the bindings are mounted (attached to the skis) by the distributor. The manager assumes that demand is approximately normally distributed with a mean of 100 and a standard deviation of 20.

This particular ski package sells for \$500, costs \$300, and can be sold at the end of the season for \$100. The manager estimates that failing to satisfy a demand will result in a future lost sale of a comparable ski package that would result in a similar margin contribution. So, the manager uses a penalty cost equal to the current product margin of \$200. In our notation:

$$\begin{aligned} p &= \$500/\text{pair} \\ v &= \$300/\text{pair} \\ g &= \$100/\text{pair} \\ B &= \$200/\text{pair} \end{aligned}$$

This is a standard newsvendor problem (no delayed product differentiation). Using the same approach as in Section 9.2.3, we obtain the results summarized in Table 9.5.

Now, suppose The Ski Bum management can purchase the skis and bindings separately for a total of \$275. It is estimated that it will cost approximately \$25 to mount the bindings in the store at the time of customer demand. Management also estimates that they can sell the skis and bindings at the end of the season and recoup the same \$100 salvage value. In our notation:

$$\begin{aligned} p &= \$500/\text{pair} \\ v_0 &= \$275/\text{pair} \\ t &= \$25/\text{pair} \\ g_0 &= \$100/\text{pair} \\ B &= \$200/\text{pair} \end{aligned}$$

Following Equation 9.16,

$$p_{u <}(Q^*) = \frac{p - v_0 - t + B}{p - g_0 - t + B} = \frac{500 - 275 - 25 + 200}{500 - 100 - 25 + 200} = \frac{400}{575} \quad (9.17)$$

Since demand is assumed to be normally distributed, we can use Appendix II or a spreadsheet formula to get $k = 0.512$. The results are summarized in Table 9.5 for easy comparison to the nonpostponed case.

From Table 9.5, we see that adopting this delayed differentiation results in a slightly higher stocking quantity and slightly higher profit. In this example, the cost of understocking is the same in both systems but the cost of overstocking is lower in the postponed system since the $t = \$25$ transformation cost is not incurred for items purchased but not sold. Since the cost consequences of overstocking are less severe, the firm is willing to purchase more. This also results in an increase in profit of a little more than 2%. Certainly no manager would turn down a 2% increase in profit, but as we will see in the next section, the potential benefits of postponement really kick in when multiple items are served from the same base product.

In some settings, the total unit and transformation cost is higher when postponement is adopted. It can be verified for this example that if the transformation cost is $t = 28.53$, the profit in the two systems is the same. That is, The Ski Bum could tolerate slightly higher total product costs under postponement and still obtain the same expected profit.

9.4.2 Value of Flexibility

We now consider the case where there are multiple end items created from the same base product. First, we address the special case where the cost of understocking $p_i - v_0 - t_i + B_i$ (where t_i is the

Table 9.5 Numerical Illustration of Single-Item, Single-Period Problem with and without Postponed Product Differentiation

	<i>Original System</i>	<i>Postponed System</i>
p \$/unit	500	500
v \$/unit	300	–
v_0 \$/unit	–	275
t \$/unit	–	25
g \$/unit	100	100
B \$/unit	200	200
\hat{x}_i units	100	100
σ_i units	20	20
k	0.431	0.512
Q or Q_0	109	110
$E[P(Q)]$ or $E[P(Q_0)]$	\$15,637	\$15,976

transformation cost for item i) is the same for all products. Making this simplifying assumption allows us to use the standard newsvendor formulation to calculate the optimal base product quantity. We subsequently relax this limiting assumption below and discuss how to address the more general multiple-item setting.

To see why the relative cost of understocking across items is critical to the analysis, consider the situation where a firm discovers they are in short supply after observing demand. They must make a choice of how to prioritize demand for different items. At this point in the decision timeline, the firm knows they are in short supply and thus will not incur overstocking costs. This means they will seek to minimize the understocking cost by first serving the demand for the item with the highest understocking cost, then the next one, and so on. If we assume that all understocking costs are the same, we can formulate the expected profit, and determine the ideal base product quantity by calculating only the *total* expected units short rather than the expected units short for each item.

Let \hat{x}_i, σ_i represent mean and standard deviation of demand for item $i = 1, \dots, n$, and let \hat{x}_0, σ_0 represent the mean and standard deviation of total demand across all items. We obtain the mean for total demand by adding the individual means, $\hat{x}_0 = \sum_{i=1}^n \hat{x}_i$. Here we will assume that demands for these items are mutually independent, meaning we obtain the variance for total demand by adding the variance of the individual demands, $\sigma_0^2 = \sum_{i=1}^n \sigma_i^2$. It may not be the case in practice that the demands are independent.

The cost of overstocking is the same as in the previous section $c_o = v_0 - g_0$. Under our assumption that the cost of understocking is the same for any stockout, $c_u = p_i - v_0 - t_i + B_i$, we can formulate the expected profit as

$$E[P(Q_0)] = \sum_{i=1}^n (p_i - v_0 - t_i) \hat{x}_i - c_o EO - c_u ES \quad (9.18)$$

where EO and ES are expected units left over and expected units short based on Q_0 and the total demand described by \hat{x}_0, σ_0 . Following Equation 9.1, we get that the best quantity satisfies

$$p_{u<}(Q_0^*) = \frac{c_u}{c_u + c_o} = \frac{p - v_0 - t_i + B_i}{p - g_0 - t_i + B_i} \tag{9.19}$$

We now consider the case where costs of understocking are different across items. For this setting, we need to know the expected units short *for each item* rather than just the total expected units short. Letting ES_i represent expected units short for item i , the expected profit function is

$$E[P(Q_0)] = \sum_{i=1}^n (p_i - v_0 - t_i)\hat{x}_i - (v_0 - g_0)EO - \sum_{i=1}^n (p_i - v_0 - t_i + B_i)ES_i \tag{9.20}$$

Our experience is that for most settings where this newsvendor model is appropriate, demand is occurring over a moderately long selling season and retailers tend to serve demand first-come first-serve. We first address this setting and discuss below the case where the retailer may be able to prioritize the use of the base product for more critical end-item demand.

If the retailer follows a policy of meeting end-item demand as it occurs, we can simplify the expected profit given in Equation 9.20 by using an *average* cost of understocking. We recommend using the demand-weighted average cost of understocking. That is, if item 1 has twice the average demand as item 2, we would give twice the weight to the cost of understocking for item 1. Formally, let \bar{c}_u be the weighted-average cost of understocking, obtained as follows:

$$\bar{c}_u = \frac{\sum_{i=1}^n \hat{x}_i (p_i - v_0 - t_i + B_i)}{\sum_{i=1}^n \hat{x}_i} \tag{9.21}$$

Now we can use the same, single-item newsvendor approach as before, with \bar{c}_u as the cost of understocking and $(v_0 - g_0)$ as the cost of overstocking.

If a firm has the ability to prioritize use of the base product, it should choose to meet more demand for items with higher understocking costs at the expense of items with lower-understocking costs. To illustrate this effect, we consider the setting where the retailer sees all demands at once. The exact approach requires some effort in determining the expected units short for each item, ES_i . We comment on this approach further below, but first we suggest an approximate approach based on the observation that the item with the lowest understocking cost that will have the most missed sales since demand for that item will always be prioritized last. Let c_u^{min} represent the lowest understocking cost among the items. If we assume that *all* missed sales incurred are for this item, we can modify Equation 9.18 to be

$$E[P(Q_0)] = \sum_{i=1}^n (p_i - v_0 - t_i)\hat{x}_i - c_o EO - c_u^{min} ES \tag{9.22}$$

and the optimal solution can then be found using the same newsvendor approach we have already seen several times with overstocking and understocking costs c_o, c_u^{min} . This simplified approach risks overstating the expected profit slightly as we assumed that all missed sales incur the lowest-possible understocking cost.

It is possible to find values for expected units short *for each item*. To illustrate how, suppose we have only two items and that item 1 has the higher understocking cost (meaning demand for

item 1 will be met first). As noted above, we can calculate the total ES based on Q_0 and \hat{x}_0, σ_0 . Since demand for item 1 will always be met first if possible, we can calculate ES_1 by using the same Q_0 and \hat{x}_1, σ_1 . If we know the total expected units short and the expected units short for item 1, expected units short for item 2 must be the difference between the total and item 1 missed sales, $ES_2 = ES - ES_1$.

9.4.2.1 Numerical Illustration

One of the popular snowboards sold by The Ski Bum comes with either a glossy or matte finish. The type of finish is determined at the time the topsheet is applied to the board. Historically, the retailer has purchased and sold matte and glossy snowboards separately but is now considering switching to a postponement system where the desired finish can be applied in the shop at the time of the customer order.

The snowboard with matte finish sells for \$400 while the glossy-finish board sells for \$450. The unit costs for either board are \$200 if purchased from the manufacturer with the finish already applied. It costs \$190 to buy an unfinished board and \$20 to complete the finishing in the shop. Any type of board can be sold at the end of season for \$50. The manager estimates that failing to satisfy a demand will result in future lost business of \$300. In our notation (let subscripts g, m represent glossy and matte):

$$\begin{aligned} p_g, p_m &= \$450, \$400 \\ v_g, v_m &= \$190 \\ t_g, t_m &= \$20 \\ g_g &= \$50 \\ B_g, B_m &= \$300 \end{aligned}$$

The manager estimates that demand for glossy-finish boards is normally distributed with mean 90 and standard deviation 25 while the demand for matte boards is normally distributed with mean 80 and standard deviation 20. Demand for the boards occurs throughout the winter selling season, so The Ski Bum meets demands as they come and cannot prioritize one over the other. This means the approach of using the weighted average understocking cost is appropriate. With postponement, the cost of understocking for each of the two boards is obtained using Equation 9.21

$$\begin{aligned} \bar{c}_u &= \frac{\hat{x}_g(p_g - v_g + B_g)}{(\hat{x}_g + \hat{x}_m)} + \frac{\hat{x}_m(p_m - v_m + B_m)}{(\hat{x}_g + \hat{x}_m)} \\ &= \left(\frac{90}{90 + 80}\right) (\$560) + \left(\frac{80}{90 + 80}\right) (\$510) = \$536.47 \end{aligned}$$

The cost of overstocking an unfinished board is the cost of the board minus the salvage value, $v_0 - g_0 = \$190 - \$50 = \$140$. To obtain the ideal stocking quantity of unfinished boards, we need the mean and standard deviation of *total* demand. The average demand is the sum of the individual-item averages, $\hat{x}_0 = \hat{x}_g + \hat{x}_m = 170$. As noted above, we assume that item demands are independent, so we get the standard deviation by taking the square root of the sum of the variances, $\sigma_0 = \sqrt{\sigma_g^2 + \sigma_m^2} = 32$. The ideal quantity is then obtained with this mean and standard deviation and the critical fractile obtained using the costs of overstocking and understocking.

Table 9.6 Numerical Illustration of Single-Item, Single-Period Problem with and without Postponed Product Differentiation

	<i>Original System—Glossy</i>	<i>Original System—Matte</i>	<i>Postponed System</i>
p \$/unit	450	400	450/400
v \$/unit	200	200	–
v_0 \$/unit	–	–	190
t \$/unit	–	–	20
g \$/unit	50	50	50
B \$/unit	300	300	300
\hat{x}_i units	90	80	170
σ_i units	25	20	32
k	0.792	0.736	0.817
Q or Q_0	110	95	196
$E[P(Q)]$ or $E[P(Q_0)]$	\$17,397	\$12,045	\$32,312
ES	3.05	2.69	3.73
EO	22.84	17.41	29.89

$$p_{u^<}(Q^*) = \frac{\bar{c}_u}{\bar{c}_u + c_o} = \frac{536.47}{536.47 + 140} = 0.79 \quad (9.23)$$

Since demand is assumed to be normally distributed, we can use Appendix II or a spreadsheet formula to get $k = 0.817$ and an ideal order quantity for the base product of 196. It is straightforward to follow the approaches outlined above to determine quantities and expected profits for the case where the two snowboards are managed separately, without postponement. Results are summarized in Table 9.6 for easy comparison to the nonpostponed case. From Table 9.6, we see the flexibility benefit made possible by delaying product differentiation. Even though the total product and transformation cost is higher, the postponed system generates higher expected profit, \$32,312 instead of \$29,442; a gain of almost 10%. Examining the expected units short and expected units left over (ES , EO), it is clear that the increase in profit comes from reducing the mismatch between supply and demand. Total expected missed sales drops from 5.74 to 3.73, and total units left over drops from 40.25 to 29.89.

The preceding analysis assumed that The Ski Bum served each demand as it arrived and could not prioritize demand for one item over the other. Suppose instead that The Ski Bum had to order snowboards well in advance of the ski season, in the summer perhaps, but all customers would pre-order in the fall. In such a case, The Ski Bum could then determine which demands to meet. It should be clear that they would prefer to sell a glossy board since it has a higher cost of understocking. Using the approach outlined in the previous section, the only change that needs to be made is that the lower cost of understocking of \$510 replaces the weighted-average cost of

understocking of \$536.47. This results in a value of k of 0.776, a slightly lower-order quantity of 195 and a slightly higher profit of \$32,453. Of course, we would expect the profit to increase given that missed sales are now only for the matte boards and thus cost only \$510 per missed sale.

9.5 More than One Period in Which to Prepare for the Selling Season

Particularly in a production context (e.g., style goods in the garment industry) there may be an extended length of time (divided, for convenience, into several time periods) in which replenishment commitments are made before the actual selling season begins. There are likely to be production constraints on the total amounts that can be acquired in each time period. Furthermore, forecasts of total demand are almost certain to change during these preseason periods, perhaps as a consequence of firm customer orders being received. We would like to be able to modify the decision rule of the previous section to cope with these new complexities effectively.

Hausman and Peterson (1972) have shown that an “optimal” solution to the problem is out of the question. However, they have proposed two heuristic decision rules that appear to perform quite well, at least on the numerical examples tested. The heuristics are closely related to the single-period procedure of Section 9.3, but allowance is made for production capability in all remaining periods. Related references include Crowston et al. (1973), Bitran et al. (1986), Matsuo (1990), Bradford and Sugrue (1990), and Fisher and Raman (1996). The latter paper allows for updating of forecasts based on sales during the early part of the season and develops a heuristic solution. Experience with a fashion apparel manufacturer is reported. See also Abillama and Singh (1996) who model a situation in which the capacity to produce within a season is limited. Therefore, production is started prior to the season, and any shortages of realized demand are produced if capacity is available in season. They find optimal order quantities for single and multiple products, and describe a number of properties of the optimal solution.

9.6 Multiperiod Newsvendor Problem

In many cases, obsolescence occurs after several time periods, implying that more than one order may be placed as the demand declines or simply evaporates. Section 8.5 of Chapter 8 discussed the case of gradually declining demands and the sizing of the final replenishment. We now introduce literature that addresses the problem of a product that faces sudden obsolescence at some unknown future time.

For multiperiod production/inventory problems, Ignall and Veinott (1969) have developed conditions under which a relatively simple policy, known as a myopic policy, is optimal for the sequence of periods. A myopic policy is one that selects the production quantities for the current period to minimize expected costs in the current period alone. Obviously, a myopic policy is much simpler to compute than is a more general multiperiod policy. However, the multiperiod style goods problem discussed above does not satisfy the conditions for which a myopic policy is optimal.

Joglekar and Lee (1993) provide an exact formulation of the inventory costs for items with deterministic demand that may face sudden obsolescence at an unknown future time. Jain and Silver (1994) generalize the problem somewhat by incorporating the potential for a longer lifetime. At the end of each period, the remaining inventory either becomes worthless or is usable for at least one more period. They use variations of the lot sizing heuristics introduced in Chapter 5 to solve

the problem of how much to order each period. See also David and Mehrez (1995) who introduce a problem in which all items become useless at once, but at a random time in the future. See also Rosenfield (1989).

Song and Zipkin (1990) develop a flexible model that can analyze sudden obsolescence as well as gradual declining demand when demand is probabilistic and is distributed as a Poisson random variable. They show that in many cases, an inventory hump occurs. When goods are ordered prior to the decline in demand, inventory reaches a level that is too high for the lower demand rate. It takes some time to work off the excess inventory. Therefore, simple myopic policies can produce significant errors, as can simple adjustments to the inventory holding cost rate. (The latter approach is commonly used. The carrying charge, r , is inflated to account for the possibility of obsolescence.) Song and Zipkin show that inventory policies that explicitly account for obsolescence perform much better than these heuristic methods. Unfortunately, these policies are not easy to compute for most demand distributions. If disposal of stock is possible, the problem becomes somewhat more manageable. Masters (1991), on the other hand, shows that the common approach of inflating r is indeed optimal when demand is deterministic and the time to obsolescence is exponentially distributed. See also Cobbaert and Van Oudheusden (1996).

Related work on multiperiod problems includes Liao and Yang (1994) and Kurawarwala and Matsuo (1996).

9.7 Other Issues Relevant to the Control of Style Goods

In this section, we provide the highlights of other research that has been done related to the control of style goods. For more detailed analyses, see the referenced literature at the end of this chapter.

9.7.1 Updating of Forecasts

This topic has been treated in Chapter 3, but we now reexamine those aspects of particular relevance to the style goods problem. In a multiperiod problem, opportunities exist to update forecasts. Several alternative methods exist, including

1. *Exploitation of the properties of the forecasts made by decision makers.* Hausman (1969) found that under certain circumstances, ratios of successive forecasts turn out to be independent variables with a specific form of probability distribution. Fisher et al. (1994b) discovered that members of the buying committee of a fashion skiwear company tended to listen to a few dominant members in developing estimates of the next season's demand. When the committee members independently forecasted demand, the average of the forecasts tended to be much more accurate. See also Chambers and Eglese (1986), Fisher et al. (1994a), and Fisher and Raman (1996).
2. *Taking advantage of the observation that sales at the retail level tend to be proportional to inventory displayed.* Wolfe (1968) used this approach to revise forecasts based on demands observed early in the selling period. See also Baker and Urban (1988), Bar-Lev et al. (1994), Kurawarwala and Matsuo (1996), Gerchak and Wang (1994), Urban (1995), Padmanabhan and Vrat (1995), and Giri et al. (1996).
3. *Simple extrapolation methods using a particular mathematical form for the cumulative sales as a function of time.* Hertz and Schaffir (1960) assumed a normal-shaped cumulative curve. Gilding and Lock (1980), in the context of a furniture manufacturer, dealt with the case

of an artificial seasonality induced by sales promotions. They found that a Gompertz curve provides a reasonable fit to cumulative sales:

$$Y_{ult} = \ln Y_t + ae^{-bt} \quad (9.24)$$

where

Y_{ult} = represents the ultimate (total) sales of the item

Y_t = cumulative sales as of time t

and a and b are constants. See also Franses (1994).

4. *Bayesian procedures.* In the Bayesian approach, one or more parameters of the probability distribution of demand are assumed unknown. Prior knowledge is encoded in the form of probability distributions over the possible values of these parameters. As demands are observed in the early part of the season, these probability distributions are appropriately modified to take account of the additional information. References include Chang and Fyffe (1971), Murray and Silver (1966), Riter (1967), Harpaz et al. (1982), Azoury (1985), Lovejoy (1990), Lariviere and Porteus (1999), and Eppen and Iyer (1997b). Also, Fisher et al. (1994b) found that in the firm they studied, using information from only the first 20% of the season's sales data significantly improved forecast accuracy.
5. *Use of information about patterns of past product demand in conjunction with estimates of the total life cycle sales of the current product.* Managers can estimate the duration of the product life cycle for a short cycle product based on the experiences with previous versions of the product. Then estimates about total sales, the timing of peak sales (such as in December), and other seasonal variations, can be added to develop specific forecasts of demand. Multiperiod variations of the newsvendor inventory model are then used to control inventory. See Kurawarwala and Matsuo (1998, 1996), and Lau and Lau (1996a). See also Fisher et al. (1994b), who advocated dividing items into those which are more predictable and those which are less predictable. Thus, productive capacity can be devoted to predictable items first. The smaller amounts of capacity available just before, or during, the season can be reserved for less predictable items.

9.7.2 *Reorders and Markdowns*

Wolfe (1968), using the assumption about sales being proportional to the inventory level, developed methods of determining each of the following:

1. The expected time T_F to sell a fraction F of the initial inventory if a fraction f has been sold by time $t(f < F)$. A style can be tentatively identified as a fast mover (with associated possible reorder) or slow mover (with associated possible markdown) depending on the value of T_F for a particular selected F value.
2. The associated order-up-to-level if a reorder is to be placed at a specific time.
3. The timing of a markdown as a function of the fraction of initial inventory sold to date, the current price, the markdown price, the salvage value of leftover material, and the assumed known ratio of the sales rate after the markdown to that before it.

Feng and Gallego (1995) studied the problem assuming that there is a fixed number of units available for sale in a season. They determined when to lower or raise the price as a function of the time until the end of the season, if there is but one chance to change the price. Based on the number of items left, they provide a threshold time when the price should be changed. Khouja (1995) shows that multiple markdowns provide higher expected profit than a single markdown. See also Carlson (1982a,b), Lau and Lau (1988); Arcelus and Rowcroft (1992), and Khouja (1996).

9.7.3 Reserving Capacity Ahead of Time

It is common in the apparel industry for a firm to reserve some portion of the capacity of one or several suppliers. One fashion-forward apparel firm we have worked with limits its activities to designing garments and arranging to have them manufactured. All manufacturing and logistics is performed by outside contractors. To ensure that enough production capacity is available to meet peak-season demand, this firm agrees to buy a certain number of, say, sport coats from a supplier over the year. Then, specific styles and colors are determined as the season approaches. Jain and Silver (1995) examine this problem for an item with uncertainty in its demand and uncertainty in the capacity of the supplier. Dedicated capacity can be ensured by paying a premium to the supplier. The two decision variables are the replenishment quantity to request and the amount of capacity to reserve. It turns out to be very easy to select the best value of the replenishment quantity. On the other hand, they are only able to characterize the general behavior of the expected profit as a function of the level of dedicated capacity. See also Ciarallo et al. (1994), So and Song (1998), and Costa and Silver (1996). In particular, the latter authors examine a multiperiod version of the Jain and Silver (1995) paper. Bassok and Anupindi (1997) examined the case in which a commitment is made to the supplier to buy a certain minimum quantity over a multiple period horizon. Anupindi and Bassok (1995) studied the case in which the buyer provides a forecast but the supplier may allow some flexibility to deviate from the forecast. Eppen and Iyer (1997a) studied a similar system in which the buyer orders a given number of items, but some are held back by the vendor. If the buyer does not purchase these, a penalty is paid. If they are ordered, they can be purchased at the original price. See also Tsay (1999) and Eppen and Iyer (1997b). Liao and Yang (1994) analyzed a system in which more than one replenishment is triggered at once. Cachon (1999) studied a case in which multiple buyers are forced to place their orders according to their supplier's production schedule so that the variability of demand experienced by the supplier is reduced. Nevertheless, total supply chain costs increase, and the buyers end up incurring most of the higher cost.

9.7.4 Inventory Policies for Common Components

In a manufacturing firm that can assemble finished products to customer order, it is often desirable to hold safety stock of components, rather than of end items. If the assembly time is short relative to the lead time for purchasing or making components, and the value added at the assembly stage is high, holding component inventory makes sense. In addition, if there is significant commonality among components, total safety stocks may be reduced by holding common components and then assembling them to order. Component commonality refers to the situation in which certain components are used to make more than one finished product.

A number of researchers have investigated the question of how much stock to hold when there is component commonality. This literature has usually viewed the problem in a newsvendor context because the long component lead time yields a single decision of how much component inventory

to order. A second-stage decision in the problem involves which finished products to assemble once demand has been realized.

Dogramaci (1979) shows that component commonality can lower costs due to a reduction in the forecast error. Baker (1985) shows how the typical relationship between the safety factor and finished product service level is altered when the safety factor is chosen for common components. Choosing a safety factor of components to satisfy, say, 90% of demand will not give a 90% service level of finished products because the common components can be allocated to several products. See also Bagchi and Gutierrez (1992).

Gerchak and Henig (1986) show that (1) the inventory of product-specific components always increases when other components are combined into common parts and the objective is to maximize profit; and (2) the inventory of the common components could increase, decrease, or stay the same. They also show that the multiperiod problem satisfies the conditions for a myopic optimal solution, so that qualitative insights from the newsvendor-type solutions hold in the more general case. Gerchak et al. (1988) extend the analysis to the case in which there is a service-level requirement, and show that property (1) does not necessarily hold. See also Baker et al. (1986). When there is a capacity or storage constraint, Gerchak and Henig (1989) show that property (1) again may not hold. See also Jönsson and Silver (1989a,b), Jönsson et al. (1993).

Eynan and Rosenblatt (1996) relax a prevalent assumption that the common component has the same cost as the product-specific components it replaces. They show that commonality may not be desirable. See also Eynan and Rosenblatt (1997). Vakharia et al. (1996) illustrate how commonality may increase the disruption on the shop floor by increasing the variability of the load at certain work centers.

Eynan and Rosenblatt (1995), in an interesting paper, compare an assemble-in-advance strategy with an assemble-to-order strategy. If products are made in advance, additional holding costs may be incurred, whereas if they are made to order, processing costs are higher. The authors give simple newsvendor-type solutions for the optimal inventory levels. They also provide conditions under which it is optimal to choose each strategy, or a combination of the two. See also Lee and Tang (1997) and Moon and Choi (1997).

Srinivasan et al. (1992) develop approximations and an algorithm for a large-scale commonality problem at IBM. Other references on the commonality problem include Collier (1982), McClain et al. (1984), Guerrero (1985), Baker (1988), Aneja and Singh (1990), Tsubone et al. (1994), Agrawal and Cohen (2001), Balakrishnan et al. (1996), Eynan (1996), and Zhang (1997).

9.7.5 *Other Research*

We now comment briefly on other research pertaining to the general problem area of this chapter. Gerchak and Mossman (1992) show that the optimal order quantity in the newsvendor problem may actually increase when demand variability decreases. The order quantity increases as demand variability increases if and only if

$$p_{x \leq}(\hat{x}) < \frac{c_u}{c_u + c_o}$$

which is often true. They show that, under quite general conditions, as uncertainty increases, the order quantity always moves away from mean demand, and the optimal cost increases.

Additional research in this area includes Eeckhoudt et al. (1995) and Kim et al. (1995).

9.8 Inventory Control of Perishable Items

In the earlier parts of the book, we have not explicitly dealt with perishability, other than possibly through a rather crude adjustment of the inventory carrying charge. Perishability refers to the physical deterioration of units of a product. Demand continues for further units of the item. In contrast, when obsolescence occurs, there is negligible further demand for the SKU involved.

Perishable items can be divided into two categories, fixed or random, depending on the lifetime of a unit of the item. For items with a fixed lifetime, the utility of each unit stays essentially constant for a fixed period of time, then drops appreciably (perhaps decreed by law). An example is blood stored in blood banks. For items where the lifetime is a random variable, the utility can decrease throughout the lifetime in a fashion that may or may not depend on the age of the unit involved (examples include fresh produce, certain types of volatile chemicals, drugs, etc.). See Abdel-Malek and Ziegler (1988).

For the fixed lifetime case, Nahmias (1982) provides an extensive and excellent review of the research on this topic. Here, we briefly comment on the research since that time. Schmidt and Nahmias (1985) consider $(S - 1, S)$ inventory policies for fixed lifetime items with Poisson demands, and positive replenishment lead times. Ravichandran (1995) analyzes the same system using (s, S) policies, and shows that it is extremely difficult to analyze. Chiu (1995) allows for a positive replenishment lead time and develops approximations for (s, Q) policies. See also Cohen et al. (1983) and Kalpakam and Sapna (1996).

Most research that analyzes the random lifetime problem assumes that inventory on hand deteriorates at a constant rate. In other words, a constant fraction of inventory on hand becomes useless each unit of time. Raafat (1991) reviews some of the literature in this area. Chung and Ting (1994) investigate a system in which demand decreases linearly with time, and where a constant fraction of items deteriorates each period. They develop a good heuristic for the amount of inventory to order. (See also Kim 1995.) Hariga and Benkherouf (1994) expand the problem to include time-varying demand. They employ heuristics like those presented in Chapter 5. Benkherouf (1995) presents an optimal procedure for the case in which demand rates decrease over a known time horizon and items deteriorate at a constant rate. See also Benkherouf and Mahmoud (1996) and Balkhi and Benkherouf (1996). Aggarwal and Jaggi (1995) extend the analysis to allow for delayed payment to the supplier. Kalpakam and Sapna (1994) analyze an (s, S) system with Poisson demands and lifetimes that follow an exponential distribution. Lead times are positive and probabilistic as well. They describe properties of the long-run expected cost. When demand is probabilistic, as in this latter case, the analysis becomes extremely complicated. Rosenfield (1989) examines the probabilistic case when each item can perish at a random time. He determines the optimal amount of inventory to dispose of when demands arrive according to a Poisson process and batch sizes are one unit. See also Padmanabhan and Vrat (1995), Abad (1996), and Chapter 8 of this book.

In an interesting paper, Vaughn (1994) blends the fixed lifetime and random lifetime literatures. In his model, neither the vendor nor the customer can observe whether the item has deteriorated. Therefore, the vendor assigns a “sell by” or “use by” date to each item. The useful life of the item is a random variable, but the shelf life is deterministic and is given by the chosen “sell by” date. This paper examines the interaction between the optimal order up to level S and the optimal “sell by” date.

For most of this research, inventory is depleted according to a first-in-first-out (FIFO) policy; that is, the oldest inventory is used first. FIFO is an optimal issuing policy over a wide range of assumptions, particularly where the issuing organization has complete control over the issuing actions. However, where the customer makes the selection (e.g., in retail food distribution where

the expiration date is shown on each unit), a last-in-first-out (LIFO) policy is likely to be observed. See Pierskalla and Roach (1972) and Albright (1976).

9.9 Summary

In this chapter, we have presented some decision logic for coping with items subject to obsolescence or perishability. In the style goods version of the problem, the logic exploited the property of not being able (for economic reasons) to carry stock over from one selling season to the next. With this chapter, we have completed our coverage of individual-item inventory management.

Problems

- 9.1 Alexander Norman owns several retail fur stores in a large North American city. In the spring of each year, he must decide on the number of each type of fur coat to order from his manufacturing supplier for the upcoming winter season. For a particular muskrat line, his cost per coat is \$150 and the retail selling price is \$210. He estimates an average sales of 100 coats but with considerable uncertainty, which he is willing to express as a uniform distribution between 75 and 125. Any coat not sold at the end of the winter can be disposed of at cost (\$150) to a discount house. However, Norman feels that on any such coat he has lost money because of the capital tied up in the inventory for the whole season. He estimates a loss of \$0.15 for every dollar tied up in a coat that must be sold off at the end of the season.
 - a. How many coats should he order?
 - b. One of the factors contributing to the uncertainty in sales is the unknown level of retail luxury tax on coats that will be established by the government early in the fall. Norman has connections in the government and manages to learn the tax level prior to his buying decision. This changes his probability distribution to a normal form with a mean of 110 and a standard deviation of 15. Now what is the best order quantity? How much was the inside information concerning the tax worth to him?
- 9.2 Consider the situation where tooling is set up to make a major piece of equipment and spare parts can be made relatively inexpensively. After this one production run of the product and its spares, the dyes, etc. will be discarded. Discuss the relationship between the problem of how many spares to run and the single-period problem.
- 9.3 Neighborhood Hardware Ltd. acts as a central buying agent and distributor for a large number of retail hardware outlets in Canada. The product line is divided into six major categories, with a different buyer being responsible for each single category. One category is miscellaneous equipment for outdoor work around the home. The buyer for this group, Mr. Harry Lock, seeks assistance from a recently hired analyst, J. D. Smith, in the computer division of the company. In particular, he is concerned with the acquisition of a particular type of small snowblower that must be ordered several months before the winter. Smith, after considerable discussions with Lock, has the latter's agreement with the following data:

Unit acquisition cost is \$60.00/unit

Selling price is \$100.00/unit

Any units unsold at the end of the winter will be marked down to \$51/unit, ensuring a complete clearance and thus avoiding the prohibitive expense of storage until the next season. The probability distribution of regular demand is estimated to be

<i>Hundreds of Units, j_o</i>	3	4	5	6	7	8
$P_j(j_o)$	0.1	0.1	0.4	0.2	0.1	0.1

- a. What is the expected demand?
 - b. What is the standard deviation of demand?
 - c. To maximize expected profit, how many units should Smith (using a discrete demand model) tell Lock to acquire?
 - d. What is the expected profit under the strategy of (c).
 - e. Suppose Smith instead decides to fit a normal distribution, having the same mean and standard deviation, to the above discrete distribution. With this normal model, what is the recommended order quantity, rounded to the nearest hundred units?
 - f. If the discrete distribution is the true one, what cost penalty is incurred by the use of the somewhat simpler normal model?
- 9.4 Suppose the supplier in Problem 9.3 offers a discount such that the unit cost is \$55 if an order of at least 750 units is placed? Should Neighborhood take advantage of this offer?
- 9.5 In reality, Mr. Lock of Neighborhood Hardware has to order three different snowblowers from the same supplier. He tells the analyst, Smith, that he will go along with his normal model (whatever that is!) but, in no event, will he allocate more than \$70,000 for the acquisition of snowblowers. In addition to item SB-1, described in Problem 9.3, characteristics of the other two items are:

<i>Item</i>	<i>Unit Acquisition Cost (\$/Unit)</i>	<i>Selling Price (\$/Unit)</i>	<i>Clearance Price (\$/Unit)</i>	<i>Demand</i>	
				<i>Mean (Units)</i>	<i>Standard Deviation (Units)</i>
SB-2	80	110	70	300	50
SB-3	130	200	120	200	40

- a. What should Smith suggest as the order quantities of the three items?
 - b. What would be the approximate change in expected profit if Lock agreed to a \$5,000 increase in the budget allocated for snowblowers?
- Note:* Assume in this problem that there are no quantity discounts available.
- 9.6 The *Calgary Herald* has a policy of not permitting returns from hotels and stores on unsold newspapers. To compensate for this fact, they have lowered the value of v somewhat.

- a. From the standpoint of a particular store, how would you evaluate the effect of a change from v_1, g_1 to $v_2, 0$ where $g_1 > 0$ and $v_2 < v_1$?
 - b. Discuss the rationale for the *Herald's* policy.
- 9.7 Professor Leo Libin, an expert in inventory theory, feels that his wife is considering the preparation of far too much food for an upcoming family get-together. He feels that the newsvendor formulation is relevant to this situation.
- a. What kinds of data (objective or subjective) should he collect?
 - b. His wife says that certain types of leftovers can be put in the freezer. Briefly indicate what impact this has on the analysis.
- 9.8 Consider a company with N warehouses and a certain amount (W units) of stock of a particular SKU to allocate among the warehouses. The next allocation will not be made until one period from now. Let

$$f_x(x_0) = \text{probability density of demand in a period for warehouse } i (i = 1, 2, \dots, N)$$

Develop a rule for allocating the W units of stock under each of the following costing assumptions (treated separately):

- a. A cost H_i incurred at warehouse i if any shortage occurs.
 - b. A cost B_i is charged per unit short at warehouse i .
- 9.9 Consider a perishable item with a unit acquisition cost v , a fixed ordering cost A , a known demand rate D , and a carrying charge r . The perishability is reflected in the following behavior of the selling price p as a function of the age t of a unit:

$$p(t) = p - bt \quad 0 < t < SL$$

where

b = positive constant

SL = shelf life

Any units of age SL are disposed of at a unit value of g , which is less than both v and $p - b(SL)$. Set up a decision rule for finding the profit maximizing value of the order quantity Q .

- 9.10 Consider a sporting goods company that carries ski equipment in stores across Canada. Because of unusual weather conditions 1 year, near the end of the ski season, the company finds itself with a surplus of ski equipment in the West and with not enough on hand in the East to meet the demand. At a meeting of marketing managers a person from the East requests that the surplus goods in the West be rush-shipped to the East. On the other hand, a Western manager suggests a discount sale there to get rid of the stock. In an effort to resolve the conflict, the President asks the Management Sciences Group to analyze the problem and report back with a recommended decision within 1 week. Briefly outline an analysis you would undertake, including the types of data that would be required. Also, if possible, suggest another alternative (i.e., differing from the two suggested by the marketing managers).
- 9.11 Although its derivation in the appendix is for the case of continuous units, Equation 9.3 also holds for the case of discrete demand and units where $ES = \sum_{x_0 > Q} (x_0 - Q)p_x(x_0)$.

- a. For an item with the following data, compute the expected profit from Equation 9.3 for Q equal to each possible (nonzero probability) value of x and thus find the best Q value. Verify that the use of Equation 9.10 gives the same result.

<i>Number of Units, x_0</i>	1	2	3	<i>Total</i>
$p_x(x_0) = \Pr(x = x_0)$	0.25	0.4	0.35	1
$p_x < (x_0) = \Pr(x < x_0)$	0	0.25	0.65	1

$p = \$10$
 $v = \$5$
 $g = \$1$
 $B = \$1$

- b. Prove analytically that Equation 9.3 holds, in general, for discrete demand units.
- 9.12 Mr. Jones, a recent MBA graduate, the only entrepreneur in his graduating class, will establish a distinctive newspaper concession on the corner of two busy streets near his alma mater. After conducting in-depth interviews, he has decided to stock three daily newspapers: the *Hong Kong Chronicle*, the *New York Times*, and the *Barrie Bugle*. Having successfully avoided all courses requiring arithmetic during his academic career, he has contacted you to advise him on the kind of ordering policy that he should follow.
- a. You first make the assumption that the daily demand for each of the three papers is normally distributed. The mean number of papers sold during the first 4 days of the week appears to be 50% greater than on Fridays and Saturdays. Mr. Jones has a small kiosk (acquired through a student loan from the bank) that can stock only 500 papers each day. In keeping with analytic modeling tradition, you assume, for mathematical convenience, that the cost of being short of any of the three papers on any particular day is \$0.06 and the cost of not selling a paper on any particular day is \$0.02. Recommend inventory stocking rules for Mr. Jones.
- b. Illustrate with the following numerical example for a typical Tuesday:

<i>Paper</i>	<i>Mean Daily Demand</i>	<i>Variance Daily Demand</i>
Chronicle	400	10,000
Times	200	8,100
Bugle	300	10,000

- c. After collecting your exorbitant consulting fee, you confess to your bartender, explaining the shortcomings of the numerical example and ordering policy above. Criticize your answers to (a) and (b) in light of reality.
- d. How would you ascertain whether it was worthwhile for Jones to go deeper into debt to buy a larger kiosk (capable of holding 1,000 copies)?

- 9.13 A local vendor of newspapers feels that dissatisfaction of customers leads to future lost sales. In fact, she feels that the average demand for a particular newspaper is related to the service level as follows:

$$\hat{x} = 1,000 + 500P$$

where \hat{x} is the average demand in copies/period and P is the fraction of demand that is satisfied. Demands are normally distributed and the standard deviation (per period) is equal to 200, independent of the service level. Other (possibly) relevant factors are:

Fixed cost per day to get any papers in = \$2

Cost per paper (for vendor) = \$0.07

Selling price per paper = \$0.15

Salvage value per paper = \$0.02

- a. Ignoring the effect of disservice on demand, what is the best number of papers to acquire per period?
 - b. Taking account of the effect, now what is the best number to acquire?
 - c. What average profit is the vendor losing if she proceeds as in (a) instead of as in (b)?
- 9.14 The Food Services Department of a large university is concerned with the decision making related to the purchase of bread for use in the cafeterias of the university. The daily demand can be assumed to be uniformly distributed between 20 and 60 loaves. There are 20 slices in a loaf of bread and the selling price is \$0.04 per slice. Any demand for bread when there is none on hand is lost but there is no associated loss of goodwill, etc. Any bread left over at the end of a day is used as ingredients for cooking (e.g., in bread pudding, stuffing, etc.). The value of such use is estimated to be \$0.05 per loaf. The supplier, Easton Bakeries, in an effort to induce high consumption offers a discount structure as follows:

Cost per loaf for first 50 loaves = \$0.30

Cost per any loaf over 50 loaves = \$0.20

The fixed cost of a procurement from the bakery = \$5.00

- a. Show the expected daily profit as a function of the number of loaves of bread ordered per day.
 - b. What quantity of bread should be purchased per day?
 - c. Outline how you would determine the daily loss in profit (compared with the optimal strategy) that would result if the university simply ordered the expected demand?
 - d. What are some complexities that would have to be included to make the problem formulation more realistic?
- 9.15 The hardware store in Hanover, New Hampshire orders lawn mowers each January. The cost per mower is \$300, and they sell for \$425. Any mowers left over at the end of the

selling season are sold during a Labor Day Sale for \$250 each. Assume that the following data applies:

<i>Demand</i>	0	1	2	3	4	5
Probability	0.10	0.15	0.30	0.20	0.15	0.10

How many mowers should the store order?

- 9.16 A retail store sells a seasonal product for \$10 per unit. The cost to the store is \$6 per unit. An end-of-season sale is required for all excess units that are sold at half price. Assume that demand for the units is normal with mean of 500 and standard deviation of 100.
- What is the recommended order quantity?
 - What is the probability that at least some customers will ask to purchase the product after the outlet is sold out, assuming you use the order quantity recommended in (a)?
 - What is the effect of reducing the standard deviation of demand to 75 (through improved forecasting or through the use of early demand information from some customers)?
- 9.17 An exclusive local men's fashion store is considering the number and styles of ties to purchase for the upcoming spring and summer season. He has informed you that ties are generally coordinated with other items (suits and shirts). He has also told you that a tie can often make the difference between a sale and a no-sale on a suit. Finally, he has informed you that ties he generally carries range from classic ties to stylish ties to seasonal colored ties. He is unsure of the factors that he should be considering in deciding what stock of ties to carry. Make some suggestions about these factors.
- 9.18 A small discount retailer of snorkeling equipment must decide how many snorkeling masks to order for this summer season. The selling price for these masks is \$20, while the cost to the retailer is only \$12. The retailer can usually buy additional masks from competitors at their retail price of \$25. Masks left over at the end of the short selling season are sold at a special discount price of \$10 per pair. The manager of the store assumes that demand is approximately normally distributed with a mean of 600 masks and a standard deviation of 200 masks.
- How many should she order?
 - What is the effect on the order quantity if she could reduce the standard deviation to 100?
- 9.19 In the text, we discussed the basic newsvendor problem where there is but a single ordering opportunity per selling period (day). In fact, we assumed that the order had to be placed prior to the actual selling period (day). Let us now consider the following modified version of the problem. After all the demand is known for a particular day, the newsvendor has an additional opportunity to order any copies that he still needs above and beyond what he ordered prior to the day. These new copies arrive quickly enough to satisfy customers, thus avoiding lost sales and any loss of good will. Let
- v_1 = unit acquisition cost for copies ordered prior to the day in question (i.e., by the usual assumed method), \$/copy
 - v_2 = unit acquisition cost for the emergency additional order, \$/copy
- Assume $g < v_1 < v_2 < p$ where
- g is the unit salvage value, \$/copy and
 - p is the selling price, \$/copy

Let Q_1 and Q_2 be the order quantities before and after the demand is known. Let x be the demand in a day.

Note: Part (b) is a marginal approach; parts (c) through (e) are total cost approaches toward obtaining the same decision rule.

- a. How is Q_2 determined given Q_1 and x ?
 - b. Use a marginal cost approach to develop a decision rule that the best Q_1 must satisfy. Assume that x and Q_1 are continuous variables and that the probability density function of x is denoted by $f_x(x_0)$.
 - c. Write an expression or expressions for the profit as a function of Q_1 and the specific value of x observed.
 - d. Develop an expression for the expected profit.
 - e. Using the result of (d), develop a decision rule that the best value of Q_1 must satisfy.
 - f. Indicate how the following complexity would alter the development in (a) and (b) only: now there is a probability π ($0 < \pi < 1$) that the emergency order will be filled (before we assumed that $\pi = 1$).
- 9.20 One of the assumptions in the basic newsvendor problem is that there is a constant underage cost c_u per unit short, independent of the size of the shortage. Suppose instead that there was a cost c_1 if any shortage occurred and an additional cost c_2 if the shortage exceeded some size Y (and no cost per unit short).
- a. Why would a marginal analysis now be extremely difficult, if not impossible?
 - b. Indicate how you would develop a decision rule for determining the best order quantity Q . (Assume that there is an overage cost of c_o per unit leftover). For simplicity, treat Q and demand as continuous variables.
- 9.21 A manufacturer of summer style goods has two opportunities to make replenishments. The first opportunity, at a unit cost v_1 , is during the previous winter when very little in the way of firm orders have yet been received. The second opportunity, at unit cost v_2 (where $v_2 > v_1$), is in the early spring after a (random) number x of firm orders have been received. The ultimate total number of units demanded is denoted by y , and x and y have a joint probability density function $f_{x,y}(x_0, y_0)$. The selling price per unit is p and any units left can be scrapped at a unit value of g . Let Q_1 be the quantity produced in the winter and Q_2 be the quantity produced in the spring.
- a. Why would x and y have a joint distribution as opposed to being independent?
 - b. Develop an expression for the best value of Q_2 given values of Q_1 and x .
 - c. Outline how you would go about obtaining the best Q_1 value.
- 9.22 Jack G. Gold, the manager of the bookstore at Lawson College, must decide on how many copies of a particular text to order in connection with an elective course to be offered by Professor Patrick Dunphy of the college's Department of Philosophy. Gold decides to seek the assistance of Professor William F. Jennings of the Department of Industrial Engineering who has some familiarity with inventory models. Jennings ascertains from Gold that historically there has been considerable variance between the number of books that Dunphy says he needs for a course and the actual number purchased by the students. The particular text under consideration is supplied by a publisher that allows up to 20% returns from the bookstore, with the latter only incurring the shipping costs (i.e., the wholesale purchase price is completely refunded). To illustrate, up to 20 copies could be handled in this fashion if the original order was for 100 copies. No refund is possible on any amount beyond the 20%. Gold indicates that any copies that could not be returned would be marked down to

50% of his cost (the wholesale price), which would clear them out. Gold points out that his permitted markup on all books is 15% of his wholesale price. Suppose that, through analysis of historical records and discussions with Gold and Dunphy, Jennings is able to develop a reasonable probability distribution of total demand for the text.

- a. Outline how Jennings might develop a decision rule for Gold to use in deciding on how many copies of Dunphy's text to order.
 - b. Mention some complexities ignored in your analysis.
- 9.23 Consider a perishable commodity that is outdated in exactly three basic periods. Suppose that an ordering decision is made at the beginning of each period and the replenishment arrives instantaneously. An order-up-to-level (R, S) system is used. The perishability of the item means that any units procured at the start of period t and still on-hand at the end of period $t + 2$ must be discarded. Any demand not satisfied is lost (i.e., there is no backordering). Simulate the behavior of such a system for 20 periods assuming $S = 30$ units and the demand in the 20 consecutive periods are 11, 9, 3, 4, 14, 11, 7, 9, 11, 10, 12, 15, 8, 11, 8, 8, 14, 15, 5, 11. In particular, what is the average of each of the following?
- a. Inventory carried forward to the next period
 - b. Number of units discarded per period
 - c. Lost sales per period

Assume that at the start of period 1 (before an order is placed) there is no older stock, that is, we place an order for 30 units.

- 9.24 The demand for a particular spare part appears to be decreasing rapidly. Any demand occurs only in the month of December each year. Because of warranty considerations, the company has guaranteed the availability of spare parts through to (and including) the 2017 year. The inventory of the item in mid-November 2016 is 10 units. The demands in December 2016 and December 2017 are assumed to be independent random variables x and y with probability mass functions as follows:

x_0	0	10	20	30
$p_x(x_0)$	0.2	0.4	0.3	0.1

y_0	0	10	20
$p_y(y_0)$	0.4	0.5	0.1

There is a setup cost of \$50 associated with producing any units of the item. The unit variable cost of a produced piece is \$20/unit and it costs \$4.50 to carry each unit in inventory from December 2016 to December 2017. In any month of December when the total demand exceeds the stock available, additional spare parts are purchased from a competitor at a cost in dollars of $60 + 35z$, where z is the number of units purchased (i.e., the size of the shortage). Any unit left over at the end of December 2017 will have a salvage value of \$5/unit. Production each year must take place before the December demand is known.

- a. Suppose at the beginning of December 2017 there were no units in stock. Determine the quantity that should be produced at that time and the associated expected costs.

- b. What are two complicating factors that make a standard newsvendor analysis inappropriate for determining the best quantity to produce prior to the December 2016 season (i.e., in late November, 2016).
- c. Briefly outline (but do not attempt to carry out any of the numerical details of) a method for determining the best quantity to produce prior to the December 2016 season.
- 9.25 a. Build a spreadsheet to determine the order quantity for the basic newsvendor problem for a given discrete demand distribution.
- b. Repeat for the normal distribution.
- 9.26 Consider the multi-item newsvendor setting described in Section 9.3. Equation 9.12 defines the optimal quantity for a given item i and a particular budget M .
- a. In order for an item i to be stocked, the right-hand side of Equation 9.12 must be positive. Write an inequality with the right-hand side of Equation 9.12 greater than or equal to zero and rearrange this inequality to get a condition where the budget M is isolated on one side of the inequality.
- b. How can you use this inequality to determine which items will be stocked for a given budget level M ?
- c. What is the economic interpretation of this inequality?

Appendix 9A: Derivations

9A.1 Basic Newsvendor Results

9A.1.1 Profit Maximization

Each unit purchased costs v , each unit sold produces a revenue of p , each unit disposed as salvage gives a revenue of g , and there is an additional cost B associated with each unit of demand not satisfied. If a quantity Q is stocked and a demand x_0 occurs, the profit is

$$P(Q, x_0) = \begin{cases} -Qv + px_0 + g(Q - x_0) & \text{if } x_0 \leq Q \\ -Qv + pQ - B(x_0 - Q) & \text{if } x_0 \geq Q \end{cases} \quad (9A.1)$$

The expected value of the profit, as a function of Q , is given by

$$E[P(Q)] = \int_0^{\infty} P(Q, x_0) f_x(x_0) dx_0$$

By substituting from Equation 9A.1, we obtain

$$\begin{aligned} E[P(Q)] &= \int_0^Q [-Qv + px_0 + g(Q - x_0)] f_x(x_0) dx_0 + \int_Q^{\infty} [-Qv + pQ - B(x_0 - Q)] f_x(x_0) dx_0 \\ &= -Qv + gQ \int_0^Q f_x(x_0) dx_0 + (p - g) \int_0^Q x_0 f_x(x_0) dx_0 + pQ \int_Q^{\infty} f_x(x_0) dx_0 \end{aligned}$$

$$\begin{aligned}
 & - B \int_Q^\infty (x_0 - Q) f_x(x_0) dx_0 \\
 = & -Qv + gQ \int_0^Q f_x(x_0) dx_0 + (p - g) \left[\int_0^\infty x_0 f_x(x_0) dx_0 - \int_Q^\infty (x_0 - Q) f_x(x_0) dx_0 \right. \\
 & \left. - Q \int_Q^\infty f_x(x_0) dx_0 \right] + pQ \int_Q^\infty f_x(x_0) dx_0 - B \int_Q^\infty (x_0 - Q) f_x(x_0) dx_0 \\
 = & -Qv + gQ \int_0^Q f_x(x_0) dx_0 + (p - g) \left[\hat{x} - ES - Q \int_Q^\infty f_x(x_0) dx_0 \right] \\
 & + pQ \int_Q^\infty f_x(x_0) dx_0 - B(ES) \\
 = & (p - g)\hat{x} - (v - g)Q - (p - g + B)ES \tag{9A.2}
 \end{aligned}$$

where ES is the expected number of units short. This is Equation 9.3.*

We can also represent Equation 9A.2 as perfect information profit minus demand–supply mismatch costs. To do this, we use the following relationship between expected units short ES and expected units left over which we denote EO .

$$\begin{aligned}
 ES - EO &= \int_Q^\infty (x_0 - Q) f_x(x_0) dx_0 - \int_0^Q (Q - x_0) f_x(x_0) dx_0 \\
 &= \int_Q^\infty (x_0 - Q) f_x(x_0) dx_0 + \int_0^Q (x_0 - Q) f_x(x_0) dx_0 = \hat{x} - Q
 \end{aligned}$$

By this relationship, $Q = \hat{x} - ES + EO$, so Equation 9A.2 becomes

$$\begin{aligned}
 E[P(Q)] &= (p - g)\hat{x} - (v - g)(\hat{x} - ES + EO) - (p - g + B)ES \\
 &= (p - v)\hat{x} - (v - g)EO - (p - v + B)ES \tag{9A.3}
 \end{aligned}$$

Equation 9A.3 is then the profit under perfect information (the per unit margin, $p - v$, times the expected demand) minus the per unit overage cost (unit cost minus salvage value) times expected units left over, and underage cost (per unit margin $p - v$ plus the penalty cost B) times expected units short.

* Lau (1997) provides a similar development for expected costs and gives formulas for uniform or exponential demands. As shown in Appendix II for the normal distribution, $ES = \sigma_x G_u(Q - \hat{x}/\sigma_x)$, which gives Equation 9.8.

To find the maximizing value of Q , we set

$$\frac{dE[P(Q)]}{dQ} = 0 \tag{9A.4}$$

By applying Leibnitz's rule,* we get $\frac{dES}{dQ} = p_{x<}(Q) - 1$, so that

$$\frac{dE[P(Q)]}{dQ} = 0 = -(v - g) - (p - g + B)[p_{x<}(Q) - 1]$$

or

$$p_{x<}(Q) = \frac{p - v + B}{p - g + B} \tag{9A.5}$$

which is Equation 9.2.

The expected profit at the optimal value of Q is found by substituting the right-hand side of Equation 9A.4 into the expected profit function:

$$\begin{aligned} E[P(Q^*)] &= (p - g)\hat{x} - (v - g)Q^* - (p - g + B)ES \\ &= (p - g)\hat{x} - (v - g)Q^* - (p - g + B) \int_{Q^*}^{\infty} (x_0 - Q^*)f'_x(x_0)dx_0 \\ &= (p - g)\hat{x} - (v - g)Q^* - (p - g + B) \int_{Q^*}^{\infty} x_0f'_x(x_0)dx_0 + (p - g + B)Q^* \int_{Q^*}^{\infty} f'_x(x_0)dx_0 \\ &= (p - g)\hat{x} - (v - g)Q^* - (p - g + B) \int_{Q^*}^{\infty} x_0f'_x(x_0)dx_0 \\ &\quad + (p - g + B)Q^*[1 - p_{x<}(Q^*)] \end{aligned}$$

By using Equation 9A.4, we get

$$\begin{aligned} E[P(Q^*)] &= (p - g)\hat{x} - (v - g)Q^* - (p - g + B) \int_{Q^*}^{\infty} x_0f'_x(x_0)dx_0 + (v - g)Q^* \\ &= (p - g)\hat{x} - (p - g + B) \int_{Q^*}^{\infty} x_0f'_x(x_0)dx_0 \end{aligned}$$

* Leibnitz's rule states that $\frac{d}{dy} \int_{a_1(y)}^{a_2(y)} b(x, y)dx = \int_{a_1(y)}^{a_2(y)} [\partial b(x, y)/\partial y]dx + b(a_2(y), y) \frac{da_2(y)}{dy} - b(a_1(y), y) \frac{da_1(y)}{dy}$.

By a result in Appendix II, we obtain

$$E[P(Q^*)] = -B\hat{x} + (p - g + B) \int_0^{Q^*} x_0 f_x(x_0) dx_0$$

It is easily verified that the second derivative is negative, thus ensuring that we have found a profit maximizing value of Q .

9A.1.2 Normally Distributed Demand

If we set

$$k = \frac{Q - x}{\sigma_x} \tag{9A.6}$$

then, as shown in Appendix II, Equation 9A.5 simplifies to

$$p_{x <}(k) = \frac{p - v + B}{p - g + B} \tag{9A.7}$$

where

$$p_{u <}(k) = \Pr\{\text{a unit normal variable takes on a value less than } k\}$$

Using the fact that

$$p_{u \geq}(k) = 1 - p_{u <}(k) \tag{9A.8}$$

Equation 9A.7 can be rewritten as

$$p_{u \geq}(k) = \frac{v - g}{p - g + B}$$

Now we show the derivation of Equation 9.9. Equation 9.8 is

$$E[P(Q)] = (p - g)\hat{x} - (v - g)Q - (p - g + B)\sigma_x G_u\left(\frac{Q - \hat{x}}{\sigma_x}\right)$$

Using the expression for $G_u(k) = [f_u(k) - kp_{u \geq}(k)]$ from Appendix II and the expression from Equation 9A.8, we obtain

$$\begin{aligned} E[P(Q)] &= (p - g)\hat{x} - (v - g)(\hat{x} + k\sigma_x) - (p - g + B)\sigma_x[f_u(k) - kp_{u \geq}(k)] \\ &= (p - g)\hat{x} - (v - g)\hat{x} - (v - g)k\sigma_x - (p - g + B)\sigma_x f_u(k) + (p - g + B)k\sigma_x p_{u \geq}(k) \\ &= (p - v)\hat{x} - (v - g)k\sigma_x - (p - g + B)\sigma_x f_u(k) + (p - g + B)k\sigma_x \left(\frac{v - g}{p - g + B}\right) \\ &= (p - v)\hat{x} - (v - g)k\sigma_x - (p - g + B)\sigma_x f_u(k) + (v - g)k\sigma_x \\ &= (p - v)\hat{x} - (p - g + B)\sigma_x f_u(k) \end{aligned}$$

9A.1.3 Discrete Case

The expected total cost in the discrete case is

$$ETC(Q) = \sum_{x_0=0}^{Q-1} c_0(Q - x_0)p_x(x_0) + \sum_{x_0=Q}^{\infty} c_u(x_0 - Q)p_x(x_0) \quad (9A.9)$$

where

$$p_x(x_0) = \Pr\{x = x_0\}$$

Let

$$\Delta ETC(Q) = ETC(Q + 1) - ETC(Q) \quad (9A.10)$$

Then, $\Delta ETC(Q)$ is the change in expected total cost when we switch from Q to $Q + 1$. For a convex cost function (the case here), the best Q (or one of the best Q values) will be the lowest Q where $\Delta ETC(Q)$ is greater than zero. Therefore, we select the smallest Q for which

$$\Delta ETC(Q) > 0 \quad (9A.11)$$

Substituting from Equations 9A.9 and 9A.10 leads, after considerable simplification, to Equation 9.10.

9A.2 Constrained Multi-Item Situation

The problem is to select the Q_i 's ($i = 1, 2, \dots, n$) to maximize

$$\sum_{i=1}^n E[P(Q_i)]$$

subject to

$$\sum_{i=1}^n Q_i v_i = W \quad (9A.12)$$

where, from Equation 9A.2

$$E[P(Q_i)] = -Q_i v_i + \int_0^{Q_i} [p_i x_0 + g_i(Q_i - x_0)] f_x(x_0) dx_0 + \int_{Q_i}^{\infty} [p_i Q_i - B_i(x_0 - Q_i)] f_x(x_0) dx_0 \quad (9A.13)$$

Equation 9A.13 and

$$f_x(x_0) dx_0 = \Pr\{\text{Total demand for product } i \text{ lies between } x_0 \text{ and } x_0 + dx_0\}$$

The Lagrangian approach (see Appendix I) is to select the multiplier M and Q_i 's to maximize

$$L(Q_i\text{'s}, M) = \sum_{i=1}^n E[P(Q_i)] - M \left[\sum_{i=1}^n Q_i v_i - W \right] \tag{9A.14}$$

This is accomplished by setting partial derivatives to zero. Setting the partial derivative with respect to M equal to zero simply produces Equation 9A.12, implying that an optimal solution must exactly consume the budget W .

Substituting Equation 9A.13 into Equation 9A.14, we have

$$L(Q_i\text{'s}, M) = \sum_{i=1}^n (-Q_i v_i + I_1 + I_2) - M \left[\sum_{i=1}^n Q_i v_i - W \right] \tag{9A.15}$$

where I_1 and I_2 are the two integrals in Equation 9A.13. The key point is that I_1 and I_2 do not involve the v_i 's.

Now Equation 9A.15 can be rewritten as

$$L(Q_i\text{'s}, M) = \sum_{i=1}^n [-Q_i(M + 1)v_i + I_1 + I_2] + MW$$

The expression in square brackets is the same as $E[P(Q_i)]$, except that v_i is replaced by $(M + 1)v_i$. Therefore, when we set

$$\frac{\partial L}{\partial Q_i} = 0$$

we obtain the same result as the independent Q^* for item i except that v_i is replaced by $(M + 1)v_i$; that is, select Q_i to satisfy

$$ip_{x<}(Q_i) = \frac{p_i - (M + 1)v_i + B_i}{p_i - g_i + B_i} \tag{9A.16}$$

A given value of M implies a value for each Q_i which in turn implies a value of

$$\sum_{i=1}^n Q_i v_i$$

Equation 9A.16 shows that the higher the value of M , the lower must be $ip_{x<}(Q_i)$, that is, the lower must be Q_i .

References

Abad, P. L. 1996. Optimal pricing and lot-sizing under conditions of perishability and partial backordering. *Management Science* 42(8), 1093–1104.

Abdel-Malek, L. L. and H. Ziegler. 1988. Age dependent perishability in two-echelon serial inventory systems. *Computers and Operations Research* 15(3), 227–238.

Abillama, W. R. and M. R. Singh. 1996. *Optimal Order Quantities of Style Products*. PhD thesis, The Tuck School of Business Administration, Dartmouth College.

- Aggarwal, S. P. and C. K. Jaggi. 1995. Ordering policies of deteriorating items under permissible delay in payments. *Journal of the Operational Research Society* 46, 658–662.
- Agrawal, N. and M. Cohen. 2001. Optimal material control in an assembly system with component commonality. *Naval Research Logistics* 48(5), 409–429.
- Albright, S. C. 1976. A Bayesian Approach to a Generalized Secretary Problem. Working Paper. Graduate School of Business, Indiana University.
- Aneja, Y. P. and N. Singh. 1990. Scheduling production of common components at a single facility. *IIE Transactions* 22(3), 234–237.
- Anupindi, R. and Y. Bassok. 1995. Analysis of supply contracts with total minimum commitment and flexibility. In *2nd International Symposium in Logistics*, University of Nottingham, UK.
- Arcelus, F. J. and J. E. Rowcroft. 1992. All-units quantity-freight discounts with disposals. *European Journal of Operational Research* 57, 77–88.
- Azoury, K. S. 1985. Bayes solution to dynamic inventory models under unknown demand distribution. *Management Science* 31(9), 1150–1160.
- Bagchi, U. and G. Gutierrez. 1992. Effect of increasing component commonality on service level and holding cost. *Naval Research Logistics* 39, 815–832.
- Baker, K. R. 1985. Safety stocks and component commonality. *Journal of Operations Management* 6(1), 13–22.
- Baker, K. R. 1988. Scheduling the production of components at a common facility. *IIE Transactions* 20(1), 32–35.
- Baker, K. R., M. J. Magazine, and H. L. W. Nuttle. 1986. The effect of commonality on safety stock in a simple inventory model. *Management Science* 32(8), 982–988.
- Baker, R. C. and T. L. Urban. 1988. Single-period inventory dependent demand models. *OMEGA* 16(6), 605–607.
- Balakrishnan, A., R. L. Francis, and S. J. Grotzinger. 1996. Bottleneck resource allocation in manufacturing. *Management Science* 42(11), 1611–1625.
- Balkhi, Z. T. and L. Benkherouf. 1996. A production lot size inventory model for deteriorating items and arbitrary production and demand rates. *European Journal of Operational Research* 92, 302–309.
- Bar-Lev, S., M. Parlar, and D. Perry. 1994. On the EOQ model with inventory-level-dependent demand rate and random yield. *Operations Research Letters* 16, 167–176.
- Bassok, Y. and R. Anupindi. 1997. Analysis of supply contracts with total minimum commitment. *IIE Transactions* 29(5), 373–381.
- Benkherouf, L. 1995. On an inventory model with deteriorating items and decreasing time-varying demand and shortages. *European Journal of Operational Research* 86, 293–299.
- Benkherouf, L. and M. G. Mahmoud. 1996. On an inventory model with deteriorating items and increasing time-varying demand and shortages. *Journal of the Operational Research Society* 47, 188–200.
- Bitran, G., E. Haas, and H. Matsuo. 1986. Production planning of style goods with high setup costs and forecast revisions. *Operations Research* 34(2), 226–236.
- Bradford, J. and P. Sugrue. 1990. A Bayesian approach to the two-period style goods inventory problem with single replenishment and heterogeneous Poisson demands. *Journal of the Operational Research Society* 41(3), 211–218.
- Cachon, G. P. 1999. Managing supply chain demand variability with scheduled ordering policies. *Management Science* 45(6), 843–856.
- Carlson, P. C. 1982a. A system for the control of retail fashion inventories. In *Spring National Conference, Institute of Industrial Engineers*, New Orleans, pp. 505–510.
- Carlson, P. C. 1982b. Fashion retailing: Orders, reorders and markdowns. In *Fall National Conference, Institute of Industrial Engineers*, Cincinnati, pp. 315–321.
- Chambers, M. and R. Eglese. 1986. Use of preview exercises to forecast demand for new lines in mail order. *Journal of the Operational Research Society* 37(3), 267–273.
- Chang, S. H. and D. E. Fyffe. 1971. Estimation of forecast errors for seasonal style-goods sales. *Management Science* 18(2), B89–B96.

- Chiu, H. N. 1995. An approximation to the continuous review inventory model with perishable items and lead times. *European Journal of Operational Research* 87, 93–108.
- Chung, K. and P. Ting. 1994. On replenishment schedule for deteriorating items with time-proportional demand. *Production Planning & Control* 5(4), 392–396.
- Ciarallo, F., R. Akella, and T. Morton. 1994. A periodic review, production planning model with uncertain capacity and uncertain demand—Optimality of extended myopic policies. *Management Science* 40(3), 320–332.
- Cobbaert, K. and D. Van Oudheusden. 1996. Inventory models for fast moving spare parts subject to “sudden death” obsolescence. *International Journal of Production Economics* 44, 239–248.
- Cohen, M. A., W. P. Pierskalla, and R. J. Sasseti. 1983. The impact of adenine and inventory utilization decisions on blood inventory management. *Transfusion* 23(1), 54–58.
- Collier, D. 1982. Aggregate safety stock levels and component part commonality. *Management Science* 28(11), 1296–1303.
- Costa, D. and E. A. Silver. 1996. Exact and approximate algorithms for the multi-period procurement problem where dedicated supplier capacity can be reserved. *OR Spektrum* 18(4), 197–207.
- Crowston, W. B., M. Wagner, and J. F. Williams. 1973. Economic lot size determination in multi-stage assembly systems. *Management Science* 19(5), 517–527.
- David, I. and A. Mehrez. 1995. An inventory model with exogenous failures. *Operations Research* 43(3), 902–903.
- Dogramaci, A. 1979. Design of common components considering implications of inventory costs and forecasting. *AIIE Transactions* 11(2), 129–135.
- Eeckhoudt, L., C. Gollier, and H. Schlesinger. 1995. The risk-averse (and prudent) newsboy. *Management Science* 41(5), 786–794.
- Eppen, G. D. and A. V. Iyer. 1997a. Backup agreements in fashion buying—The value of upstream flexibility. *Management Science* 43(11), 1469–1484.
- Eppen, G. D. and A. V. Iyer. 1997b. Improved fashion buying using Bayesian updates. *Operations Research* 45(6), 805–819.
- Erlebacher, S. J. 2000. Optimal and heuristic solutions for the multi-item newsvendor problem with a single capacity constraint. *Production and Operations Management* 9(3), 303–318.
- Eynan, A. 1996. The impact of demands’ correlation on the effectiveness of component commonality. *International Journal of Production Research* 34(6), 1581–1602.
- Eynan, A. and M. J. Rosenblatt. 1995. Assemble to order and assemble in advance in a single-period stochastic environment. *Naval Research Logistics* 42, 861–872.
- Eynan, A. and M. J. Rosenblatt. 1996. Component commonality effects on inventory costs. *IIE Transactions* 28, 93–104.
- Eynan, A. and M. J. Rosenblatt. 1997. An analysis of purchasing costs as the number of products components is reduced. *Production and Operations Management* 6(4), 388–397.
- Feng, Y. and G. Gallego. 1995. Optimal starting times for end-of season sales and optimal stopping times for promotional fares. *Management Science* 41(8), 1371–1391.
- Fisher, M. and A. Raman. 1996. Reducing the cost of demand uncertainty through accurate response to early sales. *Operations Research* 44(1), 87–99.
- Fisher, M., K. Ramdas, and K. Ulrich. 1999. Component sharing in the management of product variety: A study of automotive braking systems. *Management Science* 45(3), 297–315.
- Fisher, M. L., J. H. Hammond, W. R. Obermeyer, and A. Raman. 1994a. Accurate response: The key to profiting from QR. *Bobbin February* 35, 48–62.
- Fisher, M. L., J. H. Hammond, W. R. Obermeyer, and A. Raman. 1994b. Making supply meet demand in an uncertain world. *Harvard Business Review May–June* 72(3), 83–93.
- Franses, P. H. 1994. Fitting a Gompertz curve. *Journal of the Operational Research Society* 45(1), 109–113.
- Gallego, G. and I. Moon. 1993. The distribution free newsboy problem: Review and extensions. *Journal of the Operational Research Society* 44(8), 825–834.
- Gerchak, Y. and M. Henig. 1986. An inventory model with component commonality. *Operations Research Letters* 5(3), 157–160.

- Gerchak, Y. and M. Henig. 1989. Component commonality in assemble-to-order systems: Models and properties. *Naval Research Logistics* 36, 61–68.
- Gerchak, Y., M. J. Magazine, and B. Gamble. 1988. Component commonality with service level requirements. *Management Science* 34(6), 753–760.
- Gerchak, Y. and D. Mossman. 1992. On the effect of demand randomness on inventories and costs. *Operations Research* 40(4), 804–807.
- Gerchak, Y. and Y. Wang. 1994. Periodic review inventory models with inventory-level-dependent demand. *Naval Research Logistics* 41(1), 99–116.
- Gilding, D. B. and C. Lock. 1980. Determination of stock for sale promotions. *Journal of the Operational Research Society* 31(4), 311–318.
- Giri, B. C., S. Pal, A. Goswami, and K. S. Chaudhuri. 1996. An inventory model for deteriorating items with stock-dependent demand rate. *European Journal of Operational Research* 95, 604–610.
- Graves, S. 1982. Using Lagrangian techniques to solve hierarchical production planning problems. *Management Science* 28(3), 260–275.
- Guerrero, H. H. 1985. The effect of various production strategies on product structures with commonality. *Journal of Operations Management* 5(4), 395–410.
- Hariga, M. and L. Benkherouf. 1994. Optimal and heuristic inventory replenishment models for deteriorating items with exponential time-varying demand. *European Journal of Operational Research* 79, 123–137.
- Harpaz, G., W. Y. Lee, and R. L. Winkler. 1982. Optimal output decisions of a competitive firm. *Management Science* 28, 589–602.
- Hausman, W. 1969. Minimizing customer line items backordered in inventory control. *Management Science* 15(12), 628–634.
- Hausman, W. H. 1982. On optimal repair kits under a job completion criterion. *Management Science* 28(11), 1350–1351.
- Hausman, W. H. and R. Peterson. 1972. Multiproduct production scheduling for style goods with limited capacity, forecast revisions and terminal delivery. *Management Science* 18(7), 370–383.
- Hertz, D. and K. Schaffir. 1960. A forecasting method for management of seasonal style-goods inventories. *Operations Research* 8(1), 45–52.
- Ignall, E. and A. Veinott. 1969. Optimality of myopic inventory policies for several substitute products. *Management Science* 15(5), 284–304.
- Jain, K. and E. A. Silver. 1994. Lot sizing for a product subject to obsolescence or perishability. *European Journal of Operational Research* 75, 287–295.
- Jain, K. and E. A. Silver. 1995. The single-period procurement problem where dedicated supplier capacity can be reserved. *Naval Research Logistics* 42(6), 915–934.
- Joglekar, P. and P. Lee. 1993. An exact formulation of inventory costs and optimal lot size in face of sudden obsolescence. *Operations Research Letters* 14, 283–290.
- Jönsson, H., K. Jörnsten, and E. A. Silver. 1993. Application of the scenario aggregation approach to a two-stage, stochastic, common component, inventory problem with a budget constraint. *European Journal of Operational Research* 68(2), 196–211.
- Jönsson, H. and E. A. Silver. 1989a. Common component inventory problems with a budget constraint: Heuristics and upper bounds. *Engineering Costs and Production Economics* 18, 71–81.
- Jönsson, H. and E. A. Silver. 1989b. Optimal and heuristic solutions for a simple common component inventory problem. *Engineering Costs and Production Economics* 16, 257–267.
- Jönsson, H. and E. A. Silver. 1996. Some insights regarding selecting sets of scenarios in combinatorial stochastic problems. *International Journal of Production Economics* 45(1), 463–472.
- Kalpakam, S. and K. P. Sapna. 1994. Continuous review (s, S) inventory system with random lifetimes and positive leadtimes. *Operations Research Letters* 16, 115–119.
- Kalpakam, S. and K. P. Sapna. 1996. A lost sales ($S - 1, S$) perishable inventory system with renewal demand. *Naval Research Logistics* 43, 129–142.
- Khouja, M. 1995. The newsboy problem under progressive multiple discounts. *European Journal of Operational Research* 84, 458–466.

- Khouja, M. 1996. The newsboy problem with multiple discounts offered by suppliers and retailers. *Decision Sciences* 27(3), 589–599.
- Kim, D., D. Morrison, and C. Tang. 1995. A tactical model for airing new seasonal products. *European Journal of Operational Research* 84, 240–264.
- Kim, D. H. 1995. A heuristic for replenishment of deteriorating items with a linear trend in demand. *International Journal of Production Economics* 39, 265–270.
- Kurawarwala, A. A. and H. Matsuo. 1996. Forecasting and inventory management of short life cycle products. *Operations Research* 44(1), 131–150.
- Kurawarwala, A. A. and H. Matsuo. 1998. Product growth models for medium-term forecasting of short life cycle products. *Technological Forecasting and Social Change* 57(3), 169–196.
- Lariviere, M. A. and E. L. Porteus. 1999. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science* 45(3), 346–363.
- Lau, A. H.-L. and H.-S. Lau. 1988. The newsboy problem with price-dependent demand distribution. *IIE Transactions* 20(2), 168–175.
- Lau, H. 1980. The newsboy problem under alternative optimization objectives. *Journal of the Operational Research Society* 31, 525–535.
- Lau, H. 1997. Simple formulas for the expected costs in the newsboy problem: An educational note. *European Journal of Operational Research* 100, 557–561.
- Lau, H. and A. Lau. 1994. The multi-product multi-constraint newsboy problem: Applications, formulation and solution. *Journal of Operations Management* 13(2), 153–162.
- Lau, H. and A. Lau. 1996a. Estimating the demand distributions of single-period items having frequent stockouts. *European Journal of Operational Research* 92, 254–265.
- Lau, H.-S. and A. H.-L. Lau. 1996b. The newsstand problem: A capacitated multiple-product single-period inventory problem. *European Journal of Operational Research* 94(1), 29–42.
- Lee, H., C. Billington, and B. Carter. 1993. Hewlett-Packard gains control of inventory and service through design for localization. *Interfaces* 23(4), 1–11.
- Lee, H. L. and C. S. Tang. 1997. Modelling the costs and benefits of delayed product differentiation. *Management Science* 43(1), 40–53.
- Lee, H. L. and C. S. Tang. 1998. Variability reduction through operations reversal. *Management Science* 44(2), 162–172.
- Li, J., H. Lau, and A. H. Lau. 1990. Some analytical results for a two-product newsboy problem. *Decision Sciences* 21, 710–726.
- Li, J., H. S. Lau, and A. H. L. Lau. 1991. A two-product newsboy problem with satisficing objective and independent exponential demands. *IIE Transactions* 23(1), 29–39.
- Liao, C. and W. Yang. 1994. An inventory system with multiple replenishment scheduling. *Operations Research Letters* 15, 213–222.
- Lovejoy, W. S. 1990. Myopic policies for some inventory models with uncertain demand distributions. *Management Science* 36(6), 724–738.
- Mamer, J. and A. Shogan. 1987. A constrained capital budgeting problem with application to repair kit selection. *Management Science* 33, 800–806.
- Mamer, J. and S. A. Smith. 1982. Optimizing field repair kits based on job completion rate. *Management Science* 26(11), 1328–1333.
- Mamer, J. W. and S. A. Smith. 1985. Job completion based inventory systems: Optimal policies for repair kits and spare machines. *Management Science* 31(6), 703–718.
- Masters, J. M. 1991. A note on the effect of sudden obsolescence on the optimal lot size. *Decision Sciences* 22(5), 1180–1186.
- Matsuo, H. 1990. A stochastic sequencing problem for style goods with forecast revisions and hierarchical structure. *Management Science* 36(3), 332–347.
- McClain, J., W. Maxwell, J. Muckstadt, L. J. Thomas, and E. Weiss. 1984. Comment on “aggregate safety stock levels and component part commonality”. *Management Science* 30(6), 772–773.
- Moon, I. and S. Choi. 1995. The distribution free newsboy problem with balking. *Journal of the Operational Research Society* 46, 537–542.

- Moon, I. and S. Choi. 1997. The distribution free procedures for make-to-order, make-to-stock, and composite policies. *International Journal of Production Economics* 47(1), 21–28.
- Moon, I., E. A. Silver et al. 2000. The multi-item newsvendor problem with a budget constraint and fixed ordering costs. *Journal of the Operational Research Society* 51(5), 602–608.
- Murray, G. R. and E. A. Silver. 1966. A Bayesian analysis of the style goods inventory problem. *Management Science* 12(11), 785–797.
- Nahmias, S. 1982. Perishable inventory theory: A review. *Operations Research* 30(4), 680–708.
- Nahmias, S. and C. P. Schmidt. 1984. An efficient heuristic for the multi-item newsboy problem with a single constraint. *Naval Research Logistics* 31, 463–474.
- Padmanabhan, G. and P. Vrat. 1995. EOQ models for perishable items under stock dependent selling rate. *European Journal of Operational Research* 86, 281–292.
- Pierskalla, W. P. and C. Roach. 1972. Optimal issuing policies for perishable inventory. *Management Science* 18(11), 603–614.
- Porteus, E. L. 1990. The impact of inspection delay on process and inspection lot sizing. *Management Science* 36(8), 999–1007.
- Raafat, F. 1991. Survey of literature on continuously deteriorating inventory models. *Journal of the Operational Research Society* 42(1), 27–37.
- Ramdass, K. 2003. Managing product variety; an integrative review and research directions. *Production and Operations Management* 12(1), 79–101.
- Randall, T. and K. Ulrich. 2001. Product variety, supply chain structure, and firm performance: Analysis of the U.S. bicycle industry. *Management Science* 47(12), 1588–1604.
- Ravichandran, N. 1995. Stochastic analysis of a continuous review perishable inventory system with positive lead time and Poisson demand. *European Journal of Operational Research* 84, 444–457.
- Riter, C. 1967. The merchandising decision under uncertainty. *Journal of Marketing* 31, 44–47.
- Rosenfield, D. 1989. Disposal of excess inventory. *Operations Research* 37(3), 404–409.
- Sankarasubramanian, E. and S. Kumaraswamy. 1983. Optimal ordering quantity for pre-determined level of profit. *Management Science* 29, 512–514.
- Scarf, H. 1958. A min-max solution of an inventory problem. In K. Arrow, S. Karlin, and H. Scarf (Eds.), *Studies in the Mathematical Theory of Inventory and Production*, pp. 201–209. California: Stanford University Press.
- Schmidt, C. P. and S. Nahmias. 1985. Optimal policy for a two-stage assembly system under random demand. *Operations Research* 33(5), 1130–1145.
- Scudder, G. and S. T. March. 1984. On optimizing field repair kits based on job completion rate. *Management Science* 30(8), 1025–1028.
- Smith, S., J. Chambers, and E. Shlifer. 1980. Optimal inventories based on job completion rate for repairs requiring multiple items. *Management Science* 26(8), 849–852.
- So, K. C. and J.-S. Song. 1998. Price, delivery time guarantees and capacity selection. *European Journal of Operational Research* 111(1), 28–49.
- Song, J.-S. and P. H. Zipkin. 1990. Managing inventory with the prospect of obsolescence. *Operations Research* 44(1), 215–222.
- Srinivasan, R., R. Jayaraman, R. Roundy, and S. Tayur. 1992. *Procurement of common components in a stochastic environment*. IBM Thomas J. Watson Research Division: RC-18580,12.
- Swaminathan, J. and S. R. Tayur. 1998. Managing broader product lines through delayed differentiation using vanilla boxes. *Management Science* 44(12), S161–S172.
- Tsay, A. A. 1999. The quantity flexibility contract and supplier-customer incentives. *Management Science* 45(10), 1339–1358.
- Tsubone, H., H. Matsuura, and S. Satoh. 1994. Component part commonality and process flexibility effects on manufacturing performance. *International Journal of Production Research* 32(10), 2479–2493.
- Urban, T. L. 1995. Inventory models with the demand rate dependent on stock and shortage levels. *International Journal of Production Economics* 40, 21–28.
- Vakharia, A. J., D. A. Parmenter, and S. M. Sanchez. 1996. The operating impact of parts commonality. *Journal of Operations Management* 14(1), 3–18.

- Vaughn, T. S. 1994. A model of the perishable inventory system with reference to consumer-realized product expiration. *Journal of the Operational Research Society* 45(5), 519–528.
- Wadsworth, G. P. 1959. Probability. In *Notes on Operations Research*, pp. 26–29. Cambridge, MA: Technology Press.
- Wolfe, H. B. 1968. A model for control of style merchandise. *Industrial Management Review* 9(2), 69–82.
- Zhang, A. X. 1997. Demand fulfillment rates in an assemble-to-order system with multiple products and dependent demands. *Production and Operations Management* 6(3), 309–324.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

MANAGING INVENTORY ACROSS MULTIPLE LOCATIONS AND MULTIPLE FIRMS

IV

In Sections II and III, we have been primarily concerned with the control of individual items at a single location. With just a few exceptions, which addressed the aggregate consequences of a particular decision rule, the rules and systems developed have treated the individual items and locations in isolation. In Section IV, we now explicitly consider coordinated control of both items and locations, generally in nonproduction situations. There are three important contexts where such coordination often leads to substantial cost savings. First, in Chapter 10, we deal with a single stocking point (as in Chapters 4 through 9) but now consider coordination of replenishments of a group of items to reduce costs. Such coordination makes sense if one of the following conditions occurs:

1. Several items are purchased from the same supplier.
2. Several items share the same mode of transportation.
3. Several items are produced on the same piece of equipment.

The bulk of this chapter focuses on nonproduction situations, but we do provide a substantial treatment of one particular production situation.

Chapter 11 deals with the second context where coordination can lead to appreciable cost savings—multiechelon inventory within the same firm. Material is stored at more than one geographic location, and one must account for the possibility that the supplying location is short of stock when an order is placed. It is clear that decisions concerning the same item at different locations should not be made independently.

Chapter 12 addresses the third context—managing inventories across the supply chain. This topic shares many of the complexities of the multiechelon situation, except that we are now also dealing with the coordination of multiple firms across the supply chain.

Because of these coordination issues, exact analyses are an order of magnitude more complex than those of earlier chapters. Therefore, we are frequently forced to resort to heuristic decision rules. In Chapters 13 through 16, we address coordination issues within production environments.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Chapter 10

Coordinated Replenishments at a Single Stocking Point

In this chapter, we address an important situation in which it makes sense to coordinate the control of different stock-keeping units. Specifically, the items are all stocked at the same location and they share a common supplier or mode of transportation. Alternatively, they share a common production facility. In the latter case, there is the added complexity of a finite productive capacity to be shared each period. Conceptually, the items to be coordinated could be the same item at different, parallel locations. We emphasize the word *parallel* because coordination in a multiechelon (serial) situation (one location feeding stock into another) is quite different. The latter situation will be addressed in Chapter 11.

As we will see in Section 10.1, there are several possible advantages in coordinating the replenishments or production of a group of items. The potential disadvantages will also be discussed. In Section 10.2, for the case of deterministic demand, the economic order quantity analysis of Chapter 4 (which there assumed independent control of items) is extended to the situation where there is a major fixed (setup) cost associated with a replenishment of a family of coordinated items and a minor fixed (setup) cost for each item involved in the particular replenishment. Section 10.3 extends the quantity discount arguments of Chapter 4 to the case where the discount is based on the magnitude of the total replenishment of a group of items—for example, a freight rate reduction if a carload size replenishment is achieved. Then, in Section 10.4, we discuss the complexity of the situation in which items are to be coordinated but demand is probabilistic and there are no quantity discounts. In Section 10.5, we deal with the situation where group discounts are of paramount importance so that, when a group order is placed, we normally strive to achieve a certain overall size (or dollar value) of order. Section 10.6 moves on to the production environment in which there may be a capacity constraint. In this section, we discuss both deterministic and probabilistic demand, as well as constant and time-varying demand. It should be noted that in the literature the terminology *joint replenishment* is sometimes used in lieu of *coordinated replenishment*. Finally, in Section 10.7, we briefly discuss the related problem of shipment consolidation.

10.1 Advantages and Disadvantages of Coordination

There are a number of reasons for coordinating items when making replenishment decisions. These include the following:

1. *Savings on unit purchase costs:* When a group of items is ordered from the same vendor, a quantity discount in purchase cost may be realized if the total order is greater than some breakpoint quantity. It may be uneconomical to order this much of a single item, but it could certainly make sense to coordinate several items to achieve a total order size as large as the breakpoint. For example, a distributor incurs a fixed cost for negotiating a replenishment order from her supplier. Adding a line item to the order, however, incurs a smaller fixed cost. In some cases, a vendor-imposed minimum order quantity may dictate the same sort of joint consideration. One book publisher is required by its printer to order at least four products together. And the printer produces in only three quantities: 5,000, 10,000, or 15,000.
2. *Savings on unit transportation costs:* The discussion is basically the same as above. Now, grouping individual item orders may be advisable to achieve a quantity such as a carload or a container load on a ship. Malden Mills was a textile firm that invented Polar Fleece, a unique fabric used in outdoor jackets and other items. One of our students examined shipping policies and costs as a summer project. It turns out that, because of significant discounts, shipments from their Massachusetts factory to Europe were always created to completely fill a shipping container for the ocean voyage.
3. *Savings on ordering costs:* In some cases, where the fixed (setup) cost of placing a replenishment order is high, it might make sense to put several items on a single order to reduce the annual total of these fixed costs. This is particularly relevant where the replenishment is by in-house production. In such a case, the major component of the fixed ordering cost is the manufacturing setup cost. An illustration would be bottling beer products. There are major changeover costs in converting the production line from one quality of beer to another. In contrast, the costs of changing from one container type to another are rather minor.
4. *Ease of scheduling:* Coordinated handling of a vendor group can facilitate scheduling of buyer time, receiving and inspection workload, and so forth. In fact, we have found that, by and large, managers and purchasing agents alike tend to think and deal in terms of vendors or suppliers rather than individual stock-keeping units.

On the other hand, there are possible disadvantages of using coordinated replenishment procedures. These include the following:

1. *A possible increase in the average inventory level:* When items are coordinated, some will be reordered earlier than if they were treated independently.
2. *An increase in system control costs:* By the very nature of the problem, coordinated control is more complex than independent control of individual items. Therefore, under coordinated control, review costs, computation costs, and so on are likely to be higher.
3. *Reduced flexibility:* Not being able to work with items independently reduces flexibility in dealing with unusual situations. One possible result is reduced stability of customer service on an individual-item basis.

10.2 Deterministic Case: Selection of Replenishment Quantities in a Family of Items

In this section, we consider the case where there is a family of coordinated items defined in the following manner. There is a major setup cost (A) associated with a replenishment of the family. In the procurement context, this is the fixed (or header) cost of placing an order, independent of the number of distinct items involved in the order. In the production environment, this is the changeover cost associated with converting the facility from the production of some other family to production within the family of interest. Then there is a minor setup cost (a_i) associated with including item i in a replenishment of the family. In the procurement context, a_i is often called the line cost, the cost of adding one more item or line to the requisition. From a production standpoint, a_i represents the relatively minor cost of switching to production of item i from production of some other item within the same family. In general, we are concerned with items sharing capacity, such as a container on a ship or a piece of production equipment, or with items sharing costs, such as a major fixed ordering cost. For a review of the literature on deterministic and probabilistic models, see Goyal and Satir (1989); Van Eijs (1993), and Khouja and Goyal (2008).*

10.2.1 Assumptions

All of the assumptions behind the derivation of the economic order quantity (Section 4.1 of Chapter 4) are retained except that now coordination of items is allowed in an effort to reduce setup costs. We review the assumptions:

1. The demand rate of each item is constant and deterministic (this deterministic assumption will be relaxed in Sections 10.4 through 10.6).[†]
2. The replenishment quantity of an item need not be an integral number of units (the extension to integral units is as discussed for a single item in Chapter 4).
3. The unit variable cost of any of the items does not depend on the quantity; in particular, there are no discounts in either the unit purchase cost or the unit transportation cost (we will relax this assumption in Section 10.3).
4. The replenishment lead time is of zero duration, but the extension to a fixed, known, nonzero duration, independent of the magnitude of the replenishment, is straightforward.
5. No shortages are allowed (Sections 10.4 through 10.6 will deal with the situation where shortages can occur).
6. The entire order quantity is delivered at the same time.

* For a related problem of finding groups of raw materials that should be ordered together, see Rosenblatt and Finger (1983); Chakravarty (1984); Rosenblatt and Kaspi (1985); Chakravarty and Martin (1988); Van Eijs et al. (1992); Hong and Hayya (1992); and Dekker et al. (1997).

[†] The situation of time-varying but deterministic demand patterns with no capacity constraint (as was the case in Chapter 5) has been analyzed by Kao (1979) See Section 10.6.2 for the case with a capacity constraint.

10.2.2 Decision Rule

In the independent EOQ analysis of Chapter 4, we showed that the EOQ expressed as a time supply was given by

$$T_{\text{EOQ}} = \sqrt{\frac{2A}{Dvr}}$$

Note that T_{EOQ} increases as the ratio A/Dv increases. Within our present context, this indicates that an item i with a high setup cost a_i and a low dollar usage rate $D_i v_i$ should probably be replenished less frequently (higher time supply) than an item j having a low a_j and a high $D_j v_j$. Because of the assumptions of deterministic demand, no shortages permitted, and instantaneous delivery, it makes sense to include an item in a replenishment only when its inventory drops to the zero level. Therefore, a reasonable type of policy to consider is the use of a time interval (T) between replenishments of the family and a set of m_i 's, where m_i , an integer, is the number of T intervals that the replenishment quantity of item i will last. That is, item i will be included in every m_i th replenishment of the family.* For example, if $m_{17} = 3$, item 17 will only be included in every third replenishment of the family, with a replenishment quantity sufficient to last a time interval of duration $3T$. Each time it will be replenished just as its stock hits the zero level. We wish to select the values of T and the m_i 's to keep the total relevant costs as low as possible.

Figure 10.1 contains a more detailed example. Items 1 and 3 are ordered every time an order is placed (i.e., m_1 and $m_3 = 1$). Item 2 is ordered every other time, so that $m_2 = 2$.

As shown in the Appendix of this chapter, the integer m_i 's must be selected to minimize

$$\left(A + \sum_{i=1}^n \frac{a_i}{m_i} \right) \sum_{i=1}^n m_i D_i v_i \quad (10.1)$$

where

A = major setup cost for the family, in dollars

a_i = minor setup cost for item i , in dollars

D_i = demand rate of item i , in units/unit time

v_i = unit variable cost of item i , in \$/unit

n = number of items in the family (the items are numbered 1, 2, 3, ..., $n - 1$, n)

m_i = the integer number of T intervals that the replenishment quantity of item i will last

Also, once the best m_i 's are known, the corresponding value of T is given by

$$T^*(m_1, \dots, m_n) = \sqrt{\frac{2(A + \sum a_i/m_i)}{r \sum m_i D_i v_i}} \quad (10.2)$$

* Jackson et al. (1985) have developed an efficient procedure for the more restrictive situation where $m_i = 1, 2, 4, 8$, etc. (i.e., a multiple of 2) and T itself is also such a multiple of some basic time period (such as a day or a week). This type of restriction is commonly found in practice. Moreover, Jackson et al. (1985) have shown that their solution is, at most, 6% costlier than the "optimal" solution to the problem. We do not present the details here since their solution is more difficult conceptually than is ours. See also Roundy (1985, 1986); Atkins and Iyogun (1987); Jackson (1988); Zoller (1990); Atkins (1991); Atkins and Sun (1995), and Section 10.4 where we return to the "powers-of-two" solution. Klein and Ventura (1995) present a simple procedure for finding the optimal cycle times if T is restricted to take on one of a set of possible discrete values.

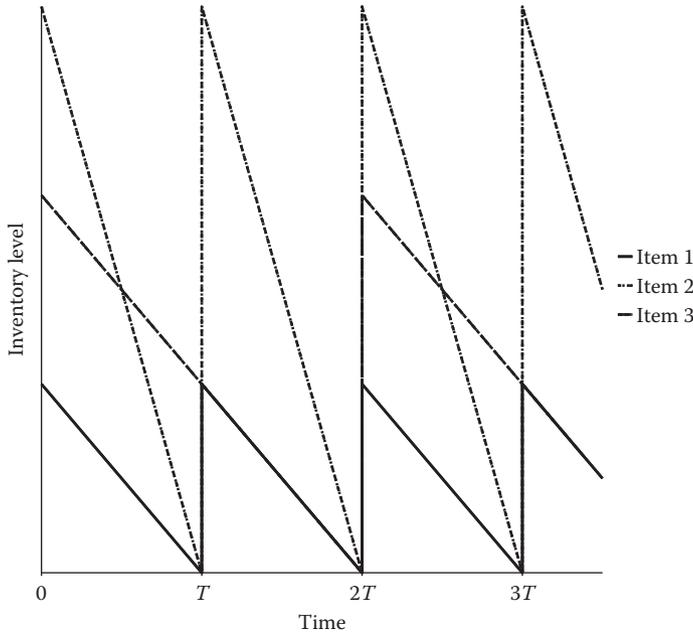


Figure 10.1 Behavior of inventory level with time.

It is worth noting that r does not appear in Equation 10.1; that is, the best values of the m_i 's do not depend on the carrying charge.

Goyal (1974) has proposed a search procedure for finding the best set of m_i 's. (See related research by Naddor (1975); Chakravarty (1985); Bastian (1986); Goyal (1988); Kaspi and Rosenblatt (1991); Goyal and Deshmukh (1993); Van Eijs (1993); Hariga (1994); and Viswanathan and Goyal (1997). Khouja and Goyal (2008) provides an excellent overview of this early as well as discussing more recent work.) We suggest, instead, the use of the following much simpler (noniterative) procedure which is derived in the Appendix of this chapter:

Step 1 Number the items such that

$$\frac{a_i}{D_i v_i}$$

is smallest for item 1. Set $m_1 = 1$.

Step 2 Evaluate

$$m_i = \sqrt{\frac{a_i}{D_i v_i} \frac{D_1 v_1}{A + a_1}} \tag{10.3}$$

rounded to the nearest integer greater than zero.

Step 3 Evaluate T^* using the m_i 's of Step 2 in Equation 10.2.

Step 4 Determine

$$Q_i v_i = m_i D_i v_i T^* \quad i = 1, 2, \dots, n \tag{10.4}$$

In numerous tests (see Silver 1975), this procedure has produced results at or near the best possible solution. Note, however, that the results are not always optimal because the m_i values are rounded.

10.2.2.1 Numerical Illustration

A plant uses the same machinery to place three different liquids into assorted container sizes. One of these liquids (known as product XSC) is packaged in four different sized containers: 1, 5, 10, and 50 liters. The major setup cost to convert to this type of liquid is estimated to be \$40. It costs \$15 to switch from one container size to another. The demand rates of this family of items are shown in Table 10.1. The carrying charge has been established as 0.24 \$/\$/year. Moreover, there is substantial excess capacity in the packaging operation, so it is reasonable to ignore capacity considerations.

Step 2 of the decision rule gives

$$m_2 = \sqrt{\frac{15}{12,500} \frac{86,000}{55}} = 1.37 \rightarrow 1$$

$$m_3 = \sqrt{\frac{15}{1,400} \frac{86,000}{55}} = 4.09 \rightarrow 4$$

$$m_4 = \sqrt{\frac{15}{3,000} \frac{86,000}{55}} = 2.80 \rightarrow 3$$

These entries are shown in Table 10.1.

Step 3 gives

$$T^* = \sqrt{\frac{2(40 + 15/1 + 15/1 + 15/4 + 15/3)}{0.24[1(86,000) + 1(12,500) + 4(1,400) + 3(3,000]}}$$

$$= 0.0762 \text{ year} \approx 4 \text{ weeks}$$

With Step 4, we obtain the run quantities

$$Q_1 v_1 = (1)(86,000)(0.0762) = \$6,550 \text{ (run quantity in dollars of 10-liter container)}$$

$$Q_2 v_2 = (1)(12,500)(0.0762) \approx \$950$$

Table 10.1 The XSC Family of Items Problem

Item <i>i</i> Description	1 ^a 10 Liters	2 1 Liter	3 5 Liters	4 50 Liters
$D_i v_i$ (\$/year)	86,000	12,500	1,400	3,000
m_i	1	1	4	3
$T^* = 0.0762 \text{ year}$				
$Q_i v_i$ (\$)	6,550	950	430	690

Note: $A = \$40$, $a_i = \$15$ (independent of i), $r = 0.24$ \$/\$/year.

^a As discussed in the decision logic, the items have been numbered such that item 1 has the smallest value of $a_i/D_i v_i$. (Here, the largest value of $D_i v_i$ because all the a_i 's are the same.)

$$Q_3v_3 = (4)(1,400)(0.0762) \approx \$430$$

$$Q_4v_4 = (3)(3,000)(0.0762) \approx \$690$$

To convert these run quantities to units, we would have to divide by the v_i values.

10.2.3 A Bound on the Cost Penalty of the Heuristic Solution

By definition, a heuristic solution is not guaranteed to give a solution with a low cost penalty. Fortunately for this problem, it is possible, as shown in the Appendix of this chapter, to find a simple lower bound for the cost of the best (T, m_i) 's policy, namely,

$$\text{TRC}_{\text{bound}} = \sqrt{2(A + a_1)D_1v_1r} + \sum_{j=2}^n \sqrt{2a_jD_jv_jr} \quad (10.5)$$

There is an interesting interpretation of the bound when we remember that an optimal policy must have item 1 included in every replenishment of the family. The first term is the total relevant cost per unit time of an EOQ strategy (see Equation 4.5 of Chapter 4) for item 1 considered alone, if we associate the full cost $A + a_1$ with each replenishment of item 1. The second term represents a summation of the total relevant costs per unit time of an EOQ strategy of each of the other items, where a cost of only a_j is associated with a replenishment of item j ($j \neq 1$).

For our numerical illustration, the value of this bound works out to be \$2,054.15/year. As shown in the Appendix of this chapter, the cost of any (T, m_i) 's solution is given by

$$\text{TRC}(T, m_i \text{'s}) = \frac{A + \sum_{i=1}^n a_i/m_i}{T} + \sum_{i=1}^n \frac{D_i m_i T v_i r}{2} \quad (10.6)$$

The solution developed by the heuristic has a cost of \$2,067.65/year (evaluated by substituting the *integer* m_i values, found by the heuristic, into Equation 10.6). The bound clearly indicates that the heuristic solution is very close to (if not right at) the optimum for the particular example.

10.3 Deterministic Case with Group Discounts

As mentioned earlier, unit price or freight rate discounts may be offered on the total dollar value or the total volume of a replenishment made up of several items (e.g., the use of containers for shipments from Massachusetts to Europe). The inventory control manager would like to ascertain when to take advantage of such a discount. Therefore, knowing the discount structure, the major setup cost (A) of a group of items, the item characteristics (a_i 's, D_i 's, and v_i 's), and the carrying charge (r), we want to develop decision logic for selecting the appropriate individual order quantities. The sum of the individual quantities totals a group quantity that determines whether or not a particular discount is achieved. We mention that sometimes discount structures are designed for much smaller customers than the organization under consideration. In such cases, discounts can be achieved by individual-item orders so that the more complex coordinated control is not necessarily needed.

As in the case of a single item, treated in Chapter 4, taking advantage of an offered discount reduces the replenishment costs (both the fixed and unit costs) but increases the inventory carrying costs. At first glance, it would appear that the analysis for multiple items should not be much more difficult than that for a single item. This would be the case if every item was included in every replenishment. However, we know from the preceding section that an $m_i > 1$ is likely for an item having a high value of $a_i/D_i v_i$; that is, such items should not necessarily be included in every replenishment, even to help achieve a quantity discount level. Therefore, it is conceivable that the best strategy might be one where on certain replenishments a discount was achieved, while on others it was not (even though all demand rates are assumed known and at constant levels). The analysis of such a strategy would be quite complex because the replenishment cycles would no longer all be of the same duration, T (the ones where quantity discounts were achieved would be longer than the others). Rather than attempting to explicitly model such complex possibilities, we suggest the following reasonable compromise solution (illustrated for the case of a single possible discount based on the total replenishment size, in units).

Our approach parallels that used in Section 4.5 of Chapter 4. We consider three possible solutions. The first (and if it is feasible, it is the best to use) is where a coordinated analysis, assuming a quantity discount, leads to total replenishment quantities that are always sufficient to achieve the discount. The second case is where the best result is achieved right at the breakpoint. Finally, the third possibility is the coordinated solution without a quantity discount. Therefore, first, the method of Section 10.2 is used to ascertain the m_i 's and T , assuming that the discount is achieved. The *smallest** replenishment quantity is computed and compared with the breakpoint quantity (Q_b) required. If it exceeds Q_b , we use the m_i 's and T developed. If it is less than Q_b , then we must compare the cost of the best solution without a discount to the cost of the solution where the smallest replenishment quantity is right at the breakpoint. Whichever of these has the lower cost is then the solution to use. The steps are presented using a discount based on the total number of units ordered, while the numerical illustration that follows uses a dollar value discount. The details are as follows:

Step 1 Compute the m_i 's and T as in Section 10.2 but assuming that each

$$v_i = v_{0i}(1 - d)$$

where

v_{0i} = basic unit cost of item i without a discount

d = fractional discount when the total replenishment equals or exceeds the breakpoint quantity Q_b

It should be noted that the set of m_i 's does not depend on the size of the discount as long as the unit cost of each item is reduced by the same percentage discount (which is typically the

* What we are assuming here is that if a discount is to be achieved, it must be achieved by every replenishment; that is, we are ignoring the more complex possibility of achieving a discount on only some of the replenishments. The smallest replenishments are those where only the items with $m_i = 1$ are included. Lu (1995) develops a heuristic method for finding the order quantities when it is possible to take the discount on some replenishments and not on others. They find that this additional flexibility can yield significant savings. See also Russell and Krajewski (1992); Sadriani and Yoon (1992, 1994); Katz et al. (1994); and Xu et al. (2000).

case). This can be seen from Equation 10.3, where, if we use $v_i = v_{0i}(1 - d)$ for all i , the $(1 - d)$ terms in the numerator and denominator cancel.

Compute the size of the smallest family replenishment

$$Q_{sm} = (\text{Summation of order quantities of all items having } m_i = 1) \tag{10.7}$$

If $Q_{sm} \geq Q_b$, use the m_i 's, T , and Q_i 's found above. If not, proceed to Step 2.

Step 2 Scale up the family cycle time T (found in Step 1) until the smallest replenishment size equals the quantity breakpoint. This is achieved at

$$T_b = \frac{Q_b}{\text{Summation of } D_i \text{ of all items having } m_i = 1} \tag{10.8}$$

The m_i 's found in Step 1 are maintained. The cost of this breakpoint solution is evaluated using the following total relevant cost expression:

$$\text{TRC}(T_b, m_i's) = (1 - d) \sum_{i=1}^n D_i v_{0i} + \frac{A + \sum_{i=1}^n \frac{a_i}{m_i}}{T_b} + \frac{r(1 - d)T_b}{2} \sum_{i=1}^n m_i D_i v_{0i} \tag{10.9}$$

Step 3 Use the procedure of Section 10.2 to find the m_i 's, T , and Q_i 's without a discount (as mentioned earlier, the best m_i 's here must be the same as those found in Step 1). As shown in the Appendix of this chapter, the total relevant costs of this solution are given by a somewhat simpler expression than Equation 10.9, namely,

$$\text{TRC}(\text{best } T \text{ and } m_i's) = \sum_{i=1}^n D_i v_{0i} + \sqrt{2 \left(A + \sum_{i=1}^n \frac{a_i}{m_i} \right) r \sum_{i=1}^n m_i D_i v_{0i}} \tag{10.10}$$

Step 4 Compare the TRC values found in Steps 2 and 3 and use the m_i 's, T , and Q_i 's associated with the lower of these. If the solution right at the breakpoint is used, care must be taken (because of the likely required integer nature of the Q 's) to ensure that the breakpoint is actually achieved. Adjustment of one or more Q values to the next higher integers may be necessary.

Prior to showing a numerical illustration of the above procedure, we make a few remarks related to the work of Schneider (1982), which was developed in the context of the food distribution industry. Specifically, he incorporates the idea of a modular order quantity (MOQ) for each item. Then, anytime the item is ordered, the size of the order must be an integer multiple of MOQ. Furthermore, recognizing that demand is probabilistic, he presents an approximation for the average inventory level of each item as a function of MOQ, the item's demand rate, and the vendor order quantity (the breakpoint). This permits the selection of the best MOQ for each possible breakpoint so that the various breakpoint solutions can be costed out. Other research in this area includes Chakravarty (1985).

10.3.1 Numerical Illustration

Consider three parts, EDS-031, EDS-032, and EDS-033, used in the assembly of a stabilization processor. These three products are purchased from the same domestic supplier who offers a discount of 5% if the value of the total replenishment quantity is at least \$600. We use \$1.50 as the basic cost of placing an order with one item involved. The inclusion of each additional item costs \$0.50. It therefore follows that

$$A = \$1.50 - \$0.50 = \$1.00$$

and

$$a_i = \$0.50 \quad \text{for all } i$$

An r value of 0.24 \$/\$/year is to be used and the D_i 's and v_{0i} 's are shown in columns 1 and 2 of Table 10.2.

Following the above procedure, Step 1 produces the results shown in columns 4 through 6 of Table 10.2. We can see from Equation 10.7 that

$$Q_{sm} \text{ in dollars} = 275 + 110 = \$385$$

which is below the quantity breakpoint of \$600. Therefore, we proceed to Step 2. The use of Equation 10.8, suitably modified because the breakpoint is expressed in dollars, gives

$$T_b = \frac{\$600}{\$6,000/\text{year} + \$2,400/\text{year}} = 0.07143 \text{ year}$$

The cost of the breakpoint solution is from Equation 10.9:

$$\begin{aligned} \text{TRC}(T_b, m_i' \text{'s}) &= 0.95(8,470) + \frac{1.00 + 0.50 + 0.50 + 0.10}{0.07143} \\ &\quad + \frac{0.24(0.95)}{2}(0.07143)(6,000 + 2,400 + 350) \\ &= 8,047 + 29 + 71 \\ &= \$8,147/\text{year} \end{aligned}$$

In Step 3, we know that the set of best m_i 's is $m_1 = m_2 = 1$ and $m_3 = 5$ (from Step 1). Using Equation 10.10, we have

$$\begin{aligned} \text{TRC}(\text{best } T \text{ and } m_i' \text{'s}) &= 8,470 + \sqrt{2(2.10)0.24(8,750)} \\ &= 8,470 + 94 \\ &= \$8,564/\text{year} > \$8,147/\text{year} \end{aligned}$$

Therefore, we should use the solution of Step 2—which assures that the smallest replenishment quantity is at the breakpoint. This is achieved by using*

* The T and Q values have been adjusted slightly upward to ensure that $Q_1 v_{01} + Q_2 v_{02}$ is at least as large as the breakpoint value of \$600.

Table 10.2 Numerical Example with Group Quantity Discount

Results of Step 1								
<i>Item i</i>	<i>Code</i>	D_i (Units/Year) (1)	v_{0i} (\$/Unit) (2)	$D_i v_{0i}$ (\$/Year) (3)	m_i (4)	T (Year) (5)	$Q_i v_{0i}$ (Δ) (6)	
1	EDS-001	12,000	0.5	6,000	1		275	}385
2	EDS-002	8,000	0.3	2,400	1	0.0459	110	
3	EDS-003	700	0.1	70	5		16	
				8,470				
385 < 600 Therefore go to Step 2 Final solution $T = 0.0715$ year								
<i>Item i</i>	m_i	Q_i (Units)	$Q_i v_i$ (Δ)					
1	1	858	429.00	} \$600.60				
2	1	572	171.60					
3	5	250	25.00					

Note: $A = \$1$, $a_i = a = \$0.50$, $r = 0.24$ \$/\$/year.

- $T = 0.0715$ year
- $Q_1 = 858$ units ($D_1 T$) every group replenishment
- $Q_2 = 572$ units ($D_2 T$) every group replenishment
- $Q_3 = 250$ units ($5D_3 T$) every fifth group replenishment

We can see that a sizable saving of (8,564–8,147) or \$417/year (around 4.9%) is achieved by taking the discount.

10.4 Case of Probabilistic Demand and No Quantity Discounts

Probabilistic demand greatly complicates the decision problem in a coordinated control context. Several questions must be answered:

1. How often do we review the status of items? (That is, the choice of the review interval, R .)
2. When do we reorder the group?
3. How much do we order? (In particular, whether or not to achieve a discount quantity on a particular order.)
4. How do we allocate the total order among the individual items?

Because of the coordination, it no longer follows that an item will always be at its reorder point when it is included in a replenishment. Usually, it will be above in that some other item in the family will have triggered the order. This complicates matters in two ways. First, it is now more difficult to ascertain the average inventory level of an item. Second, and more serious, the service implications of any particular reorder point are much more difficult to evaluate than in the case of individual-item control.

In the next two sections, we present an overview of two important types of coordinated control systems under probabilistic demand, particularly emphasizing the second approach.

10.4.1 (S, c, s) , or Can-Order, Systems

Can-order systems are specifically geared to the situation where savings in setup costs are of primary concern (e.g., where several products are run on the same piece of equipment) as opposed to achieving a specified total replenishment size (for quantity discount purposes). Balintfy (1964) was the first to propose the use of an (S, c, s) system, a special type of continuous review system for controlling coordinated items. In such a system, whenever item i 's inventory position drops to or below s_i (called its must-order point), it triggers a replenishment action that raises item i 's level to its order-up-to-level S_i . At the same time, any other item j (within the associated family) with its inventory position at or below its can-order point c_j is included in the replenishment. If item j is included, a quantity is ordered sufficient to raise its level to S_j . The idea of having a can-order point is to allow an item j , whose inventory position is low enough (at c_j or lower), to be included in the order triggered by item i , thus eliminating an extra major setup cost that would likely occur in the near future when item j reaches its must-order point. On the other hand, inclusion of item j in the order is not worthwhile if its inventory position is high enough—above its can-order point (i.e., above c_j). The behavior of a typical item under such a system of control is shown in Figure 10.2.

For additional information on can-order systems, including how to establish appropriate values of the s , c , and S , see Ignall (1969); Silver (1974); Thompstone and Silver (1975); Silver (1981); Federgruen et al. (1984); Goyal and Satir (1989); Zheng (1994); Schultz and Johansen (1999); and Nielsen and Larsen (2005). Related work includes Gross et al. (1972).

10.4.2 Periodic Review System

Atkins and Iyogun (1988) suggest a procedure for this problem that outperforms can-order policies, and is easier to compute. McGee and Pyke (1996) apply this procedure at a fastener manufacturer using a spreadsheet to find the production quantities and safety stock. They modified the procedure to account for sequence-dependent setup times and a production capacity constraint. The Atkins and Iyogun (1988) procedure essentially allocates the major replenishment cost in small amounts to the products that are produced or bought most frequently, keeping the expected time to the next replenishment for these products in balance.

We now describe this procedure in some detail. First compute the EOQ and equivalent time supply for each product i using only the minor setup cost a_i :

$$\text{EOQ}_i = \sqrt{\frac{2a_i D_i}{rv_i}}$$

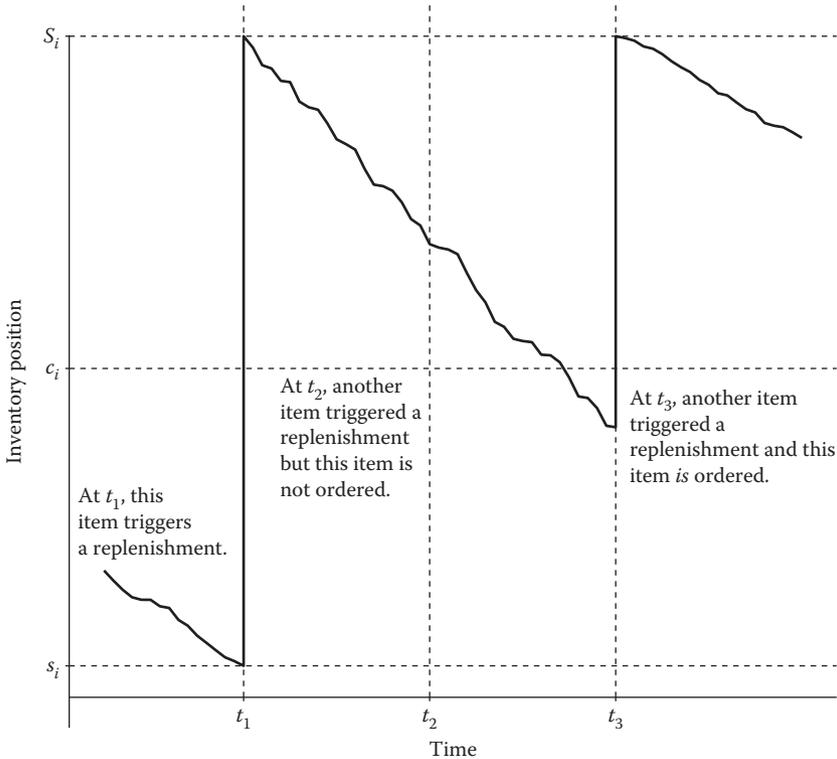


Figure 10.2 Behavior of an item under (S, c, s) control.

So the time supply, or expected *runout time*, in weeks is $EOQ_i(50)/D_i$, for product i assuming a 50-week year.* The result is a set of time supplies for each product in the family. Then choose the product with the smallest time supply, and denote it product 1 for now. A small portion, α_1 , of the major setup cost is allocated to this first product. The idea is to eventually allocate the entire major setup cost to several products in the family. If all the costs were allocated to one product, its EOQ would increase significantly, thereby increasing its time supply. It would then be likely that this product would be purchased or produced less frequently than every cycle, and the setup cost would be incorrect. Hence, the algorithm calls for allocating only a small portion of the major setup cost to the most frequently purchased product.

Allocating more and more α_1 to the first product increases its time supply, eventually to the point that it matches the time supply of the second most frequently purchased product (based only on its a_i), denoted product 2 for now. Now allocate more of the major setup cost to the first product, and some to product 2, keeping the time supplies in balance. When these two equal time supplies increase to the time supply of the third most frequently purchased product, begin allocating α_3 to this third product. This process continues until the entire major setup is allocated—that is, $\sum_i \alpha_i = 1$. Therefore, each time a family is ordered, the entire setup cost is accounted for.

* This is essentially the same as Step 1 in Section 10.2.2.

If we express the total setup cost for an item as $(\alpha_i A + a_i)$, it is easy to solve for α_i for a given time supply:

$$\alpha_i = \frac{T_i^2 r_i v_i D_i}{2A} - \frac{a_i}{A} \quad (10.11)$$

where T_i is the time supply, EOQ_i/D_i . This expression can be calculated in a spreadsheet to quickly find all values of α_i such that the sum is 1 and the T_i values are equal for all products that have $\alpha_i > 0$. The time supply of the set of products for which $\alpha_i > 0$ is called the *base period*, or *base cycle*.

Atkins and Iyogun (1988) advocate rounding all products that are not purchased every cycle to a multiple of the base period. McGee and Pyke (1996) use powers-of-two multiples of the base period, so that each product's cycle is either the base period or $2^1, 2^2, 2^3$, etc. times the base period. Powers-of-two multiples create a greater sense of order, allowing regular preventive maintenance, communication with customers about production schedules, and easier scheduling of operators. See also the references in footnote^(*).

Finally, safety stock is determined by using the periodic review (R, S) policies of Chapter 6, using the cycle time for each product as the R value.

10.4.2.1 Numerical Illustration

Four products are ordered by a distributor from a single vendor. The product data are given below, and the following information has been gathered. The fixed cost to place an order is \$23, but once an order is placed, an item can be added to the order for \$3. Thus, we set $A = \$20$, and $a = \$3$. The carrying charge is 24% and the delivery lead time is 1 week. Assume a 50-week year.

Product	Annual Demand	Unit Cost	Standard Deviation of Monthly Demand	EOQ	Time Supply in Weeks
1	450	\$8.00	19	38	4.17
2	2,000	\$12.50	50	63	1.58
3	200	\$3.52	4	38	9.42
4	3,000	\$33.30	88	47	0.79

The first step is to compute the EOQ and time supply based on the \$3 minor ordering cost only. This calculation is shown in the table, where, for example, the EOQ for product 1 is

$$EOQ_1 = \sqrt{\frac{2a_1 D_1}{rv_1}} = \sqrt{\frac{2(3)(450)}{0.24(8)}} = 37.5 \approx 38$$

* What we are assuming here is that if a discount is to be achieved, it must be achieved by every replenishment; that is, we are ignoring the more complex possibility of achieving a discount on only some of the replenishments. The smallest replenishments are those where only the items with $m_i = 1$ are included. Lu (1995) develops a heuristic method for finding the order quantities when it is possible to take the discount on some replenishments and not on others. They find that this additional flexibility can yield significant savings. See also Russell and Krajewski (1992); Sadriani and Yoon (1992, 1994); Katz et al. (1994); and Xu et al. (2000).

Table 10.3 Solution to the Periodic Review Example

(1) Product	(2) α	(3) Time Supply in Weeks	(4) Rounded to Powers-of-Two	(5) Average Order Size
1	0.0000	4.17	4	36
2	0.1102	2.08	2	80
3	0.0000	9.42	8	32
4	0.8898	2.08	2	120

and the time supply is $37.5/450 = 0.0833$ years, or 4.17 weeks.

At this point, the major ordering cost has not been counted. So, we assign a small portion of the \$20 cost to product 4, which has the smallest time supply. For example, if $\alpha_4 = 0.1$, the total setup cost is $\$20(0.1) + 3 = \5 , and the new EOQ is 61, implying a time supply of 1.02 weeks. The time supply is longer, as we would expect, but it is still less than that of product 2. As we gradually increase α_4 the time supply of product 4 increases until it reaches 1.58, the time supply of product 2. This occurs at $\alpha_4 = 0.45$, which can be found from Equation 10.11, using 1.58 for T .

Now, increase α_4 slightly while increasing α_2 as well, keeping the time supplies of the two in balance. For example, if $\alpha_4 = 0.70$, and $\alpha_2 = 0.063$, the time supplies of each become 1.88 weeks. If the time supplies were to increase to 4.17, the time supply of product 1, we would increase α_1 also. This process would continue until $\sum_i \alpha_i = 1$. In this example, however, the sum of the α values reaches 1.0 before their time supplies increase to 4.17. The values are $\alpha_4 = 0.8898$, and $\alpha_2 = 0.1102$, which total to 1.0, and the time supplies of products 2 and 4 become 2.08 weeks.

Now, we have time supplies given in column 3 of Table 10.3.

The next step is to round the base period, 2.08, to an integer, or 2.0, and then round all other time supplies to powers-of-two multiples of the base period. Thus, 4.17 is rounded to $2 \times 2^1 = 4$, and 9.42 is rounded to $2 \times 2^2 = 8$.

Therefore, products 2 and 4 will be ordered every 2 weeks, while product 1 will be ordered every 4 weeks, or on every other replenishment cycle. Product 3 will be ordered every fourth cycle. On average, the amount ordered will be as given in column 5 of Table 10.3.

The final step is to use methods of Chapter 6 to find order-up-to-levels for each product, using R as given in column 4 and $L = 1$ week. This entire procedure can be built on a spreadsheet and searching for the α values. See Problem 10.9.

Tests of this procedure against the can-order policy as well as a number of other policies show that it performs very well in terms of total cost. If the major setup cost is close to zero, can-order policies perform better. However, finding optimal values of s , c , and S is not easy. See Iyogun (1991); Golany and Lev-er (1992), and Pantumsinchai (1992). For other research on this problem, see Sivazlian and Wei (1990); and Eynan and Kropp (1998). See also Song (1998) for finding the fill rate for a multi-item order.

10.5 Probabilistic Demand and Quantity Discounts

In contrast with the situation in the previous section, now it is attractive to achieve a specified group order size. One obvious example is where a discount is offered if the group order size exceeds

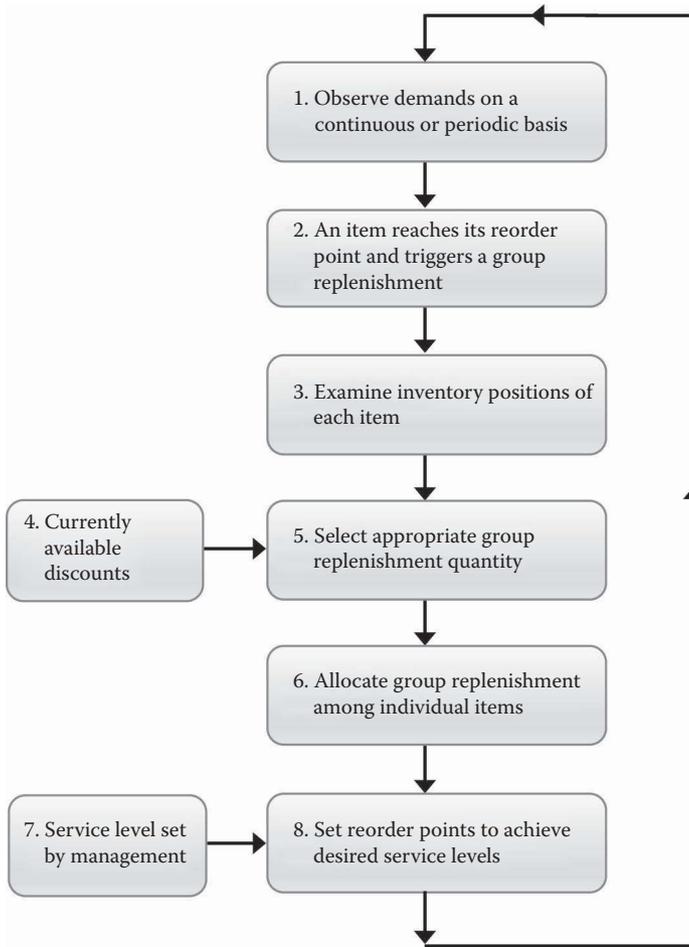


Figure 10.3 Components of a coordinated control system with quantity discounts and probabilistic demand.

some level. Another is where several products are shipped on a periodic basis in a single truck, rail car, or boat. In either case, it is important to answer the four questions posed near the start of Section 10.4.

We first provide an overview of a system developed by Miltenburg (1982) that is adapted for use on a computer. Figure 10.3 should be helpful in understanding the following discussion of Miltenburg's system.

Block 1 in Figure 10.3 indicates that either continuous or periodic review can be used. In a coordinated situation (particularly the transportation setting described earlier), a convenient periodic review interval such as a day or a week would typically be used.

As indicated in Block 2, a reorder point system is used. At a review instant, if the inventory position of any item in the family is at or below its reorder point, then a group replenishment is triggered. The setting of the reorder points will be discussed later.

The choice of an appropriate group replenishment (Block 5), of course, depends on the current inventory status of each item (Block 3) and what quantity discounts are currently available

(Block 4). In this decision, the methodology of Section 10.2 could be used; however, it is based on a steady-state model and does not take into account the current inventory situation. Thus, instead, the more appropriate decision logic of the “special opportunity to buy” situation (Section 4.9.5 of Chapter 4) is used to evaluate the various possible group replenishment sizes. We comment below on a case in which a full truckload defines the group replenishment size.

Once an overall replenishment size is selected, it must be allocated (Block 6) among the individual items. This is done to maximize the time until the next replenishment. After any given allocation, the individual stock levels are depleted by demands. The first to hit its reorder point triggers a new group order at the next review instant. Because of the probabilistic nature of the demand, the time until the next reorder is a random variable. Miltenburg models the demands for separate items as independent, diffusion processes (a diffusion process, besides being mathematically convenient, has the attractive property of implying that the total demand over any particular interval of time has a normal distribution).

Just as in Section 10.4.1, often an item will be reordered when it is above its reorder point. This is because some other item in the family will trigger the replenishment. The excess stock above the reorder point is known as residual stock. Account must be taken of this residual stock because it provides safety stock above and beyond the usual safety stock built into the reorder point. Unfortunately, the probability distribution of the residual stock of an item depends on the inventory positions of all of the items after the allocation is made. Thus, after every allocation, these distributions must be evaluated and used to establish the reorder points (Block 8) in order to provide a desired service level (Block 7).

For the case of periodic review, Miltenburg has found that a normal distribution provides a reasonable fit to the distribution of residual stock. In contrast, for the case of continuous review, a more appropriate fit is achieved by a spike at 0 and a truncated normal distribution above it (see Problem 10.8).

For other research in this area, see Low and Waddington (1967); Johnston (1980); Agnihotri et al. (1982); Miltenburg and Silver (1984a,b,c, 1988); Miltenburg (1985, 1987); Ernst and Pyke (1992); Van der Duyn Schouten et al. (1994); and Ertl et al. (2000).

10.5.1 A Full Truckload Application

In this section, we present an example of the Miltenburg system. One distributor of laminated wood products ordered extremely long beams from its Oregon supplier. The supplier required that the beams be shipped in full-truckload quantities because of the high fixed cost of sending a truck. This cost was due in part to police escorts that were required in many of the small towns the trucks had to travel through. How should the distributor determine when to order the next truckload, and how many of each beam should be ordered?

To solve this problem, we could apply the can-order system described in Section 10.4.1, but there is no guarantee that the order quantities will generate a full truckload. The periodic system of Atkins and Iyogun illustrated in Section 10.4.2 is another possibility, but again, we are faced with the problem of filling the truck exactly. The solution we describe is an application of the Miltenburg approach and is discussed by Carlson and Miltenburg (1988). It is called the service point method, and it has been used successfully in an industrial application. We shall assume that a periodic review system is employed (Block 1 of Figure 10.3). The steps are as follows:

1. Compute the acceptable shortages per replenishment cycle (ASPRC). This is a management decision, and can be computed as follows: $ASPRC = \sum_{i=1}^n Y_i(1 - P_2)$, where Y_i is the

expected usage of item i in the current replenishment cycle. (Y_i is simply the expected order size, Q_i , because Q_i is the average number of units used in each cycle), P_2 is the fill rate discussed in Chapter 6, and n is the number of items in the family. (This is Block 7.)

2. At every review instant, compute the expected shortages per replenishment cycle (ESPRC) if an order is not placed. Recall that ESPRC was computed as $\sigma_L G_u(k)$ in Chapter 6. In this case, however, k is different: $k = (IP - \hat{x}_{R+L})/\sigma_{R+L}$, where IP is the inventory position at the time of the review. (Now we are not concerned with s , but with the inventory position at the time of the review.) Also, because the next order, if we do not order now, will arrive $R + L$ periods later (R until the next review time, and L for delivery), we use $R + L$. (This is an application of Block 3.)
3. If $ESPRC > ASPRC$, place an order. If not, wait until the next review period, and check again. (This is a variation on Block 2.)
4. If the decision is to order, allocate the truck capacity to all products so as to maximize the time to the next replenishment. This last step is a bit more complicated, but can be approximated by assigning the truck capacity in proportion to the average demand of each product, adjusted for the current inventory positions, IP_i . That is, allocate (or order) an amount, Q_i , of product i , so to equate $(IP_i + Q_i)/D_i$ for all i , such that $\sum Q_i = TC$, where TC is the truck capacity.* (The truck capacity is the gross replenishment size, or Block 5, and the allocation is Block 6.)

10.5.2 Numerical Illustration

As an illustration, assume that the distributor of laminated beams stocks just three products that are supplied by the Oregon manufacturer. The data for each product are given below. In addition, we shall simplify the problem slightly by assuming that the truck holds 850 beams of any size. In reality, we would need to insure that the truck capacity is scaled by beam sizes. Management has set a 98% fill rate target for each product.

	Product			Total
	1	2	3	
Annual demand, D_i	1,200	3,000	6,000	10,200
Y_i	100	250	500	850
\hat{x}_{L+R}	20	50	100	
σ_{L+R}	5	15	40	

We can see that, on average, a truckload composed of 100, 250, and 500 beams of products 1, 2, and 3, respectively, is ordered each month. Of course, any particular order will likely be composed of different amounts than these.

At a given review instant, assume that the inventory positions of the three products are 15, 45, and 125, respectively. We must decide whether to place an order. The ASPRC is calculated based

* See Carlson and Miltenburg (1988) for more detail.

on the product data and fill rate target:

$$ASPRC = \sum_{i=1}^n Y_i(1 - P_2) = (100 + 250 + 500)(1 - 0.98) = 17$$

So the k value for product 1 is computed as follows:

$$k_1 = (IP_1 - \hat{x}_{L+R})/\sigma_{L+R} = (15 - 20)/5 = -1$$

where the subscript denotes the product number. $G_u(k)$ is therefore 1.0833 (Table II.1 of Appendix II, or a spreadsheet as described in Appendix III), and

$$ESPRC = \sigma_{L+R}G_u(k) = 5(1.0322) = 5.42$$

Repeating the calculations for products 2 and 3 gives the below results.

	Product		
	1	2	3
IP	15	45	125
\hat{x}_{L+R}	20	50	100
σ_{L+R}	5	15	40
k	-1.000	-0.333	0.625
$G_u(k)$	1.0833	0.5876	0.1619
ESPRC	5.42	8.81	6.48

The total ESPRC is 20.71, which is higher than the acceptable number of 17. Therefore, an order should be placed.

As mentioned above, the allocation of truck capacity to the three products is somewhat complicated. For this example, we illustrate a simple approach that provides a reasonable approximation. In reality, one might want to consult the research papers mentioned above and apply an approach that could provide lower-cost solutions. For our illustration, we begin the allocation by ordering enough of each product to raise its inventory position to 80% of Y_i . (The 80% value was chosen based on work by Miltenburg (1985).) Because the inventory position of product 1 is 15, $100(0.80) - 15 = 65$ beams are initially allocated to this product. Likewise, products 2 and 3 are allocated $250(0.80) - 45 = 155$ and $500(0.80) - 125 = 275$ beams, respectively. This totals 495 beams, leaving $750 - 495 = 355$ beams to allocate. If we allocate based on average demand rates, product 1 should receive $1,200/(10,200) = 0.118$ of the total. This translates to $0.118(355) = 42$ beams. Results for all products are shown below. The truck is filled by 107 beams of product 1, 259 beams of product 2, and 484 beams of product 3. Notice that $(IP_i + Q_i)/D_i$ is the same (1.22) for each product.

	<i>Product</i>			<i>Totals</i>
	<i>1</i>	<i>2</i>	<i>3</i>	
Proportion to allocate	0.118	0.294	0.588	1
Initial allocation	65	155	275	495
Remaining allocation	42	104	209	355
Total allocation, Q_j	107	259	484	850

In other research, Pantumsinchai (1992) tests a rule similar to the service point method against can-order policies and the periodic method of Atkins and Iyogun, and finds that the service point and periodic methods are quite good. For other research on the problem of finding order quantities with limits on storage space or budgets, see Roundy (1985); Hall (1988); Ventura and Klein (1988); Dagpunar (1988); Schneider and Rinks (1989); Sivazlian and Wei (1990); Mehrez and Benn-Arieh (1991); Lee (1994); Chen and Min (1994); Atkins (1991); Hariri et al. (1995); Martel et al. (1995); Güder et al. (1995); Güder and Zydiak (1999); and Haksever and Moussourakis (2005).

10.6 Production Environment

Throughout this chapter, we have noted that these coordination issues also apply in the production setting, where major and minor setups are determined by the time and cost of changing over production equipment among families and within families. Unfortunately, unless the factory is not capacity-constrained, the methods developed above cannot be applied unless adjustments are made. In this section, we discuss several dimensions of the coordination problem in a production setting. The methods we describe apply most commonly in the lower right corner of the product process matrix introduced in Chapter 2. In other words, they are applied in continuous flow processes, although they may also be applicable to batch flow processes.

10.6.1 Case of Constant Demand and Capacity: Economic Lot Scheduling Problem

A pharmaceutical manufacturer with which one of the authors consulted was faced with a bottleneck in the packaging lines. Changeover from one product to another entailed a major cleaning that could take up to 8 hours, while changes of label took only 1/2 hour. How should this firm schedule production at the packaging lines to replenish finished goods inventory in time to meet their service goals? One possibility is a can-order system. However, any system that reacts to demand occurrences is in danger of creating an idle facility for some time, followed by an overload due to multiple families reaching their reorder points at the same time. The factory simply may not have enough capacity to meet the demand. A better choice is to be proactive by creating a sequence of production that will be followed periodically, similar to the procedure described in Section 10.4.2. The solution to the economic lot scheduling problem (ELSP) is designed to do precisely that.

In the discussion that follows, we shall assume that if setup times are dependent on the sequence of production, a particular sequence of products can be developed and will be followed.* Thus, we will know the correct setup cost and time to use for each family and each item within the family. Membership in a family is defined by relatively low setup time or cost between members. Also, in this section we shall assume that demand is constant over time. In other words, we assume that the pharmaceutical manufacturer faces very little demand fluctuation and that there is no seasonality or trend in the demand over time. (In Sections 10.6.2 and 10.6.3, we will relax this restrictive assumption.) This constant demand case is essentially the same as the economic order quantity context discussed in Section 4.1 of Chapter 4, except there are several items with differing demand rates and finite production rates, and there is, of course, a production capacity restriction. A simple selection of independent finite-rate economic run quantities of the separate items (see Equation 4.19 of Chapter 4) can easily lead to an infeasible solution because two or more items are likely to require production at the same time.

The ELSP is to find a cycle length, a production sequence, production times, and idle times, so that the production sequence can be completed in the chosen cycle, the cycle can be repeated over time, demand can be fully met, and annual inventory and setup costs can be minimized. As it happens, this problem is notoriously difficult to solve.† Two factors contribute to this difficulty: the need to satisfy a production capacity constraint, and the need to have only one product in production at a time (a so-called “synchronization” constraint). A multitude of papers have been written on the ELSP. Here we provide a procedure for a simple case, and then briefly review some of the literature that extends the research into more complex situations.

Let us begin by introducing several terms used frequently in the literature. A cyclic schedule is one in which the entire system is periodic, which means that, regardless of the sequence of production, the schedule repeats itself over time. One variant of a cyclic schedule is to choose a base period and to restrict the order interval of each item to be an integer multiple of that base period. This variant includes the powers-of-two multiples mentioned above.‡ Whatever the multiple, the batch sizes of a given item will be the same over time.§ A special case of the base period approach is a pure rotation schedule, in which each product is produced once in each cycle.¶ In other words, each product shares a common order interval, T . Of course, there is no guarantee that this schedule will be optimal. Therefore, when we restrict the choice of schedules to this type, we are resorting to the use of a heuristic. A second variant of a cyclic schedule is to allow items to be produced more than once in the overall period of the system, and to allow the batch sizes for each item to differ over that period.** We will focus here on pure rotation schedules.

* An example of sequence-dependent setups is a fastener manufacturer that has long setup times between fastener sizes, but short setup times within size. (Products that share the same size may differ in the cut of the thread, the finish, or other factors.) Another is beer production that sequences products from high quality to low quality. Within a family (or a given quality level), there are different container sizes. See Dobson (1992) and McGee and Pyke (1996).

† See, for example, Hsu (1983) and Gallego and Shaw (1997).

‡ See Maxwell and Singh (1983); Jackson et al. (1985); and Roundy (1985, 1989) who have suggested the use of powers-of-two multiples of the base period for all products in the family. They show that the cost of this restriction is very low—within 6% of the optimal if a fixed base cycle is used, and within 2% if the base cycle is a variable.

§ See Elmaghraby (1978); Hsu (1983); Schweitzer and Silver (1983); and Axsäter (1987).

¶ See, for example, Gallego (1990).

** See Dobson (1987); Roundy (1989); Gallego (1992); and Zipkin (1991b).

In a pure rotation schedule, there is no chance of two products requiring production equipment at the same time because each product is scheduled only once during each cycle. Therefore, if a cycle length can be found that satisfies the capacity constraint, we will have a feasible solution to the pure rotation heuristic. In developing a pure rotation schedule, we shall focus the discussion on finding the reorder interval, rather than the batch size, partly because it is often easier to think in terms of frequency of production than in terms of number of units. The mathematical representation of the problem also becomes easier. Implementation is much easier as well because the planner can adjust batch sizes a little if demand varies, while keeping to the same reorder interval. This regularity facilitates preventive maintenance, operator scheduling, and coordination with other machines. (See Muckstadt and Roundy 1993.)

Some notation is required:

T = the common order interval, or time supply, for each product, in units of time

p_i = the production rate for product i , in units per unit time

A_i = setup cost for item i , in dollars

K_i = setup time for item i , in units of time

D_i = demand rate of item i , in units/unit time

v_i = unit variable cost of item i , in \$/unit

n = number of items in the family (the items are numbered $1, 2, 3, \dots, n - 1, n$)

The total relevant costs for item i are (following the finite production rate treatment in Section 4.8)

$$\text{TRC}_i(T) = \frac{A_i}{T} + \frac{rv_i D_i (p_i - D_i) T}{2p_i}$$

Notice that $D_i T$ = the batch size for item i . Now let $H_i = (rv_i D_i (p_i - D_i))/2p_i$ to simplify the notation. The total amount of time to produce a batch of product i is $K_i + (D_i T/p_i)$, or the setup plus processing times. Therefore, we need a constraint on the total time in a cycle:

$$\sum_{i=1}^n \left(K_i + \frac{D_i T}{p_i} \right) \leq T \quad (10.12)$$

The problem is to minimize

$$\sum_{i=1}^n \text{TRC}_i(T) \quad (10.13)$$

subject to

$$\sum_{i=1}^n \left(K_i + \frac{D_i T}{p_i} \right) \leq T \quad (10.14)$$

Equation 10.14 can be written as

$$\frac{\sum_{i=1}^n K_i}{1 - \sum_{i=1}^n \frac{D_i}{p_i}} \leq T \quad (10.15)$$

Taking the derivative of Equation 10.13 with respect to T and setting it equal to zero gives

$$T' = \sqrt{\frac{\sum_{i=1}^n A_i}{\sum_{i=1}^n H_i}} \tag{10.16}$$

If the capacity constraint Equation 10.12 is satisfied, the optimal T value is given by T' . Otherwise, T needs to be increased to insure that there is enough capacity. Then the optimal value is exactly the constraint (because $\sum_{i=1}^n TRC_i(T)$ is a convex function of T):

$$T'' = \frac{\sum_{i=1}^n K_i}{1 - \sum_{i=1}^n (D_i/p_i)} \tag{10.17}$$

Therefore, the optimal order interval is $\max(T', T'')$.

10.6.1.1 Numerical Illustration

A pharmaceutical manufacturer produces three bottle sizes of one product on one packaging machine. No other products are assigned to this equipment, and although other equipment is used for this product, this machine is the clear bottleneck. The data for the item are given in the following table:

Item	1	2	3
D_i (units/year)	1,850	1,150	800
v_i (\$/unit)	50	350	85
p_i (units/year)	5,000	3,500	3,000
A_i (\$)	125	100	110
K_i (years)	0.00068	0.00171	0.00091
H_i (\$)	7,284	33,781	6,233

The carrying charge, r , is 25%. The setup times for the three items are 6, 15, and 8 hours, respectively. So, $0.00068 = 6/[(365 \text{ days})(24 \text{ hours per day})]$, because the equipment is run 24 hours per day. Using Equation 10.16, we find

$$T' = \sqrt{\frac{125 + 100 + 110}{7,284 + 33,781 + 6,233}} = 0.0842$$

From Equation 10.15, we find that

$$\frac{0.00068 + 0.00171 + 0.00091}{1 - (1,850/5,000 + 1,150/3,500 + 800/3,000)} = 0.0952 > 0.0842$$

Therefore, the order interval is too low and the capacity constraint is violated. So we set $T'' = 0.0952$ using Equation 10.17. Converting this value to weeks yields a production cycle

of $0.0952(52) = 4.95$, or 5 weeks. Every 5 weeks, set up and produce $(5/52)(1,850) = 178$ units of product 1, followed by 111 units of product 2 and 77 units of product 3. (How much time is spent for each batch?)

Gallego and Queyranne (1995) show that there is an easy way to determine how close the pure rotation schedule may be to the optimal. First compute the optimal time supply for each item taken independently using Equation 4.7 of Chapter 4 or $T_i^* = \sqrt{A_i/H_i}$. In our illustration, these values are 0.1310, 0.0544, and 0.1328, respectively. Then take the ratio of the largest to the smallest of these values, and call it R . So $R = 0.1328/0.0544 = 2.44$. The ratio of the cost of the pure rotation schedule (TRC_{PRS}) to the optimal (TRC_{OPT}) is then

$$\frac{\text{TRC}_{\text{PRS}}}{\text{TRC}_{\text{OPT}}} \leq \frac{1}{2} \left(\sqrt{R} + 1/\sqrt{R} \right)$$

In our illustration, this bound is 1.10. Thus, the cost of the pure rotation schedule is no more than 10% more than the cost of the unknown optimal solution.

We now mention a small portion of the literature on the ELSP. The earliest work in this field has been reviewed by Elmaghraby (1978). More recently, Dobson (1987) developed a formulation of the problem that allows lot sizes and therefore cycle times to vary over time. (This is the second variant of cyclic schedules mentioned above.) He was also one of the first to explicitly consider setup time, rather than using setup cost as a surrogate of setup time. See also Maxwell (1964) and Delporte and Thomas (1977). Zipkin (1991a) built on Dobson's work by taking the sequences from Dobson as givens, and then using a heuristic to find the production run times and machine idle times for each product. Taken together, Zipkin and Dobson provided a heuristic that computes feasible schedules with reasonable effort. Bourland and Yano (1997) test several heuristics and show their performance.

Additional research has addressed the ELSP with the added complication of delivery schedules when a supplier would rather run larger lot sizes than the customer desires. See Hahm and Yano (1992, 1995a,b). Silver (1995) and Viswanathan and Goyal (1997) consider the situation in which a family of items follows a cyclic schedule, but there is a limit on shelf life. The cycle length and production rate are adjusted to insure a feasible schedule. Taylor and Bolander (1996) ask a series of questions designed to help the reader understand whether MRP (Chapter 15) is the correct tool for scheduling the factory. The focus of the article is to determine whether process flow scheduling, like ELSP, is appropriate. See also Taylor and Bolander (1990, 1991, 1994) and Bolander and Taylor (1990).

In the realm of changing the givens, Gallego and Moon (1992) examine a multiple product factory that employs a cyclic schedule to minimize holding and setup costs. When setup times can be reduced, at the expense of setup costs, by externalizing setup operations, they show that dramatic savings are possible for highly utilized facilities. Allen (1990) modifies the ELSP to allow production rates to be decision variables. He then develops a graphical method for finding the production rates and cycle times for a two-product problem. The results were successfully applied in a chemical firm. See also Moon et al. (1991); Hwang et al. (1993); Gallego (1993); Gallego and Moon (1995); and Moon (1994).

Other references include Salomon (1991); Bretthauer et al. (1994); Dobson and Yano (1994); Ramudhin and Ratliff (1995); Davis (1995); Bramel et al. (2000); Gallego et al. (1996); Taylor et al. (1997); and Brahimi et al. (2006)

10.6.2 Case of Time-Varying Demand and Capacity: Capacitated Lot Sizing

The methods of Chapter 5 apply to the case of time-varying deterministic demand when there is no capacity constraint. A more difficult problem arises if it may not be possible to produce or purchase all of the desired parts in a given period. In addition, we allow production capacities to vary with time (but they are still assumed to be known with certainty). Under such circumstances, as in Chapter 5, it is most common to consider discrete time intervals such as days or weeks, each with a known capacity and known individual-item demand rates. The objective is to keep the total (across all items and all periods out to some horizon) of setup and carrying costs as low as possible, subject to (1) no backlogging of demand or loss of sales, and (2) no violations of production capacity constraints.

In Chapter 5, for the single-item, uncapacitated situation, we were able to argue that a replenishment quantity need only be placed when the inventory level is zero,* and also that we can restrict our attention to replenishment quantities that would cover exactly an integer number of periods of requirements. This is certainly not the case with capacitated production as evidenced by the simple, single-item example shown in Table 10.4. The only feasible schedule is to produce at capacity (i.e., 10 units) each period. Clearly, the quantity produced in period 1 lasts for 1.3 periods and, moreover, production must take place in period 2 even though the starting inventory in that period is nonzero (3 units). Thus, the nature of the solution, even for a single-item problem, is considerably more complicated when production capacities are taken into account. Not surprisingly, a mathematically optimal solution is out of the question when we are dealing with the multi-item, time-varying, capacitated case. We must again turn to the use of a heuristic solution procedure.

We outline the nature of a heuristic procedure first developed by Dixon (1979), and also presented in Dixon and Silver (1981). Subsequent industrial applications are described by Van Wassenhove and De Bodt (1983) and Van Wassenhove and Vanderhenst (1983). The latter, in particular, use the heuristic as part of an hierarchical planning framework. The procedure is forward-looking in nature; as with the basic Silver–Meal heuristic of Chapter 5, it develops the current period's production quantities using demand information for as few periods into the future as possible (a desirable characteristic when we recognize the increasing uncertainty in forecasts as we project further into the future). The heuristic makes use of the Silver–Meal heuristic's basic criterion—namely, that, if item 1 is to be produced in period 1, then we wish to select the production time supply (T_i) as the integer number of periods that produces the (first local) minimum of

Table 10.4 Single-Item, Capacitated Example

Period t	1	2	3	4	Total
Demand D_t	7	10	9	14	40
Capacity C_t	10	10	10	10	40

* The discussion here assumes a negligible lead time. If there is a known lead time L , then the order should be placed L time units before the inventory level hits 0.

the total relevant costs per unit time:

$$\text{TRCUT}(T_i) = \frac{A_i + b_i \sum_{j=1}^{T_i} (j-1)D_{ij}}{T_i}$$

where

$$\begin{aligned} A_i &= \text{setup cost for item } i, \text{ in } \$ \\ b_i \text{ (or } v_i r) &= \text{holding cost for item } i, \text{ in } \$/\text{unit/period} \\ D_{ij} &= \text{demand for item } i \text{ in period } j, \text{ in units} \end{aligned}$$

However, where capacity is limited, there is competition for a scarce resource and it is quite likely that not all lot sizes to be produced in a particular period can be such that the associated *TRCUT* is minimized. In order to describe how the heuristic deals with this difficulty, we need to introduce a new variable U_i , defined as follows:

$$U_i = \frac{\text{TRCUT}(T_i) - \text{TRCUT}(T_i + 1)}{k_i D_{i, T_i + 1}}$$

where k_i is the amount of production resource required per unit produced of item i .

We can see that U_i represents the marginal decrease in costs per unit time of item i per unit of capacity absorbed if the time supply produced of item i is increased from T_i to $T_i + 1$. Each item requiring production in the period under consideration has its T_i first set to unity (the minimum feasible amount that must be produced). Then, the item with the largest positive U_i has its T_i increased to $T_i + 1$. This is continued until there is insufficient capacity to increase any of the T_i 's by unity or until all U_i 's are negative (the latter situation would give the unconstrained Silver–Meal lot sizes for each of the items).

The above procedure clearly does not violate the capacity restriction in the particular period being scheduled. However, it does not guard against another type of infeasibility. Consider the situation where in some future period the total demand exceeds the production capacity of that period. Then, some of the requirements of that period must be satisfied by production in preceding periods. This is illustrated by our example of Table 10.4 where four of the units required in period 4 must be produced prior to that period. Thus, the heuristic also has a look-ahead feature built into it to ensure that, as the T_i 's are increased (as described in the previous paragraph), enough total production is accomplished to prevent future infeasibilities.

If the heuristic is applied to a problem having a horizon with well-defined ending conditions, then the above procedure is used to establish an initial solution of all production lots out to the horizon. In such a case, the authors discuss several possible adjustments for improving on the initial solution. However, extensive testing has revealed that only one of these types of adjustments tends to be attractive in terms of potential savings without too much extra computational effort. Specifically, an attempt should be made to eliminate each scheduled lot by, if possible, allocating the production to periods where there is capacity available and the item under consideration is already being produced. The lot should be eliminated only if the increased carrying costs do not exceed the setup cost saved. In the more common case of a rolling-horizon implementation with an indefinite ending point, we do not recommend any such adjustment procedure.

There have been a number of other heuristic procedures developed for solving the capacitated, multi-item, time-varying demand problem. See reviews by De Bodt et al. (1984); Thizy and Van Wassenhove (1985); Bitran and Matsuo (1986); Bahl and Zionts (1987); and Maes and

Van Wassenhove (1988). Also see, to name just a few of the many references, Florian and Klein (1971); Lambrecht and Vanderveken (1979); Baker et al. (1978); ter Haseborg (1982); Lambert and Luss (1982); Bolander and Taylor (1983); Bahl and Ritzman (1984); Blackburn and Millen (1984); Billington (1986); Maes and Van Wassenhove (1986); Taylor and Bolander (1986); Eppen and Martin (1987); Fogarty and Barringer (1987); Gunther (1987); Erenguc (1988); Prentis and Khumawala (1989); Trigeiro et al. (1989); Joneja (1990); Dixon and Poh (1990); Gilbert and Madan (1991); Maes et al. (1991); Pochet and Wolsey (1991); Salomon (1991); Thizy (1991); Campbell (1992); Diaby et al. (1992); Hill and Raturi (1992); Rajagopalan (1992); Toklu and Wilson (1992); Cattrysse et al. (1993); Mercan and Erenguc (1993); Tempelmeier and Helber (1994); Billington et al. (1994); Lotfi and Yoon (1994); Kirca (1995); Gardiner and Blackstone (1995); DeMatta and Guignard (1995); Potamianos et al. (1996); and Tempelmeier and Derstroff (1996). We make special mention of a procedure presented by Van Nunen and Wessels (1978) because it involves a conceptual approach very different from that of the Dixon and Silver heuristic. The latter builds up an initial feasible schedule with possible subsequent attempts at improvement. The van Nunen and Wessels approach is to first ignore the capacity restrictions and solve each individual item problem independently. The resulting overall solution is usually infeasible, violating one or more of the production capacities. The next step is to adjust the solution until it is feasible with as small as possible an increase in the costs. (See also Karni and Roll 1982.)

Variations on this problem are many. Eppen and Martin (1987); Pochet and Wolsey (1991); and Millar and Yang (1994) consider a related problem, but add the possibility of backordering. A number of authors have examined the case in which there is a cost to start up a facility for production, and a reservation cost that represents the cost of having the facility available regardless of whether it is producing or not. This reservation cost may represent the opportunity cost of tying up the facility or machine for the given product, and thus making it unavailable for other products. Therefore, a single item problem can be viewed as a subproblem of the more realistic multi-item case. This research applies when the facility produces a relatively small number of products, and a product may tie up the facility for more than one time period. See Karmarkar and Schrage (1985); Karmarkar et al. (1987); Sandbothe (1991); Hindi (1995); and Coleman and McKnew (1995).

10.6.3 Probabilistic Demand: The Stochastic Economic Lot Scheduling Problem

When demand, setup times or processing rates are not deterministic, production may not follow the production plan developed by deterministic approaches considered in the ELSP described in Section 10.6.1. If variability is high, there may be significant disruptions from using the deterministic solution in a probabilistic environment. Thus, adjustments must be made. The stochastic economic lot scheduling problem (SELSP) exactly parallels the ELSP with the added complexity of probabilistic (i.e., stochastic) demand, setup times, or processing rates.

There are two basic approaches to dealing with this problem. One is to develop a regular cyclic schedule using a solution to the deterministic problem, and then to develop a control rule that attempts to track or follow this schedule. The other is to develop a heuristic that directly decides which product to produce next and its production quantity. Therefore, the production sequence can vary over time.

Examples of the first approach include Bourland and Yano (1994) who consider the use of safety stock and idle time to account for random demand. They find that except when setup costs are

very high, it is better to use safety stock rather than built-in idle time. See also Gallego (1990) who develops a real-time scheduling tool in three steps. First, replace probabilistic demands by their average values, and compute an optimal or near-optimal target cyclic schedule. Next, formulate a control policy to recover the target cyclic schedule after a disruption perturbs the inventories. Finally, find safety stocks to minimize the long run average cost of following the target schedule with the recovery policy. See also Loerch and Muckstadt (1994); Gallego (1994); and Gallego and Moon (1996).

A particularly compelling solution is by Federgruen and Katalan (1996c) who propose a simple strategy: when the facility is assigned to a given product, continue production until either a target inventory level is reached or a specific production batch is completed. The different products are produced in a given sequence, possibly with idle times inserted between the completion of an item's production batch and the setup for the next item. Federgruen and Katalan (1996c) provide a rather complex but very fast and accurate algorithm for computing the approximately optimal policy. Markowitz et al. (2000) show that setup costs should not be used as a surrogate for setup times, and that the policy of Federgruen and Katalan (1996c) should not be used when setup times are negligible but setup costs are high. They allow for switching production from one product to another based on the inventories of all products. Contrast this with the base stock policy, which relies only on the status of the product currently being produced. See also Bowman and Muckstadt (1993); Federgruen and Katalan (1994, 1996a,b, 1998); Anupindi and Tayur (1998); and Bowman and Muckstadt (1995).

The second approach, which dynamically decides which product to produce next, will generally have lower total costs, but may not be better in a real application because of disruptions to schedules of other production stages. Leachman and Gascon (1988) develop a methodology based on target cyclic schedules that are updated dynamically based on estimates of the time when inventories of various products will run out. They assume periodic review and rely on the ELSP solution for deterministic nonstationary demand to find the next item to produce and the production quantity. The schedule is modified if an item is backordered or close to a stockout. See also Leachman et al. (1991); Wein (1992); Veatch and Wein (1996); Duenyas and Van Oyen (1995, 1996); Perez and Zipkin (1997); Gascon et al. (1994); Goncalves et al. (1994); Duenyas and Van Oyen (1996); Qui and Loulou (1995); Sox and Muckstadt (1995, 1996); and Shaoxiang and Lambrecht (1996). Carr et al. (1993) consider a strategy that makes B and C items to order, while making A items to stock.

For the probabilistic time-varying demand case, see Karmarkar and Yoo (1994) and Winands et al. (2011).

10.7 Shipping Consolidation

A problem related to the coordination of multiple items is that of shipment consolidation. When a customer order arrives, the vendor must decide whether to ship the order immediately, or to wait for more orders to arrive so that the truck will be more fully loaded. Clearly, the truck entails a fixed cost regardless of the number of units shipped. In addition, costs may increase for each additional unit included. The focus here is on unit weight or volume, rather than on line items in a customer order, as above. Higginson and Bookbinder (1994) pointed out that shipment consolidation decisions include:

- Which orders will be consolidated and which will be shipped individually?

- When will orders be released for possible shipping? Immediately, or after some time trigger or quantity trigger?
- Where will the consolidation take place? At the factory or at an off-site warehouse or terminal?
- Who will consolidate? The manufacturer, customer, or a third party?

Higginson and Bookbinder (1994) also tested three policies: a time policy that ships at a prespecified time, a quantity policy that ships when a given quantity is achieved, and a time/quantity policy that ships at the earliest of the time and quantity values. The shipper must trade off cost per unit with customer service in deciding on which policy to use. Some literature that may be helpful includes Tyworth et al. (1991); Van Eijs (1994); Bausch et al. (1995); Higginson and Bookbinder (1995); and Hall and Racer (1995).

10.8 Summary

In this chapter, we first showed the details of coordinated control under deterministic demand, both with and without quantity discounts. Then, we developed an appreciation of the complexity of the probabilistic demand situation. For the latter, we illustrated a useful model for coordinated control. We then extended the discussion to the production context and examined constant and deterministic demand, time-varying deterministic demand, and probabilistic demand.

In the next chapter, we turn to another, probably more important, type of coordination. That is, where the same item is stored at more than one location with stock flowing from one location to another.

Problems

10.1 Consider a family of four items with the following characteristics:

$$A = \$100 \quad r = 0.2/\text{year}$$

Item <i>i</i>	$D_i v_i$ (\$/Year)	a_i (\$)
1	100,000	5
2	20,000	5
3	1,000	21.5
4	300	5

- a. Use the procedure of Section 10.2.2 to find appropriate values of the family cycle time T and the integer m_i 's.
- b. Suppose that we restricted attention to the case where every item is included in each replenishment of the family. What does this say about the m_i 's? Now find the best value of T .

c. Using Equation 10A.2 (in this Appendix), find the cost difference in the answers to parts a and b.

10.2 The Ptomaine Tavern, a famous fast-lunch spot near an institute of higher learning, procures three basic cooking ingredients from the same supplier. H. Fishman, one of the co-owners, estimates that the fixed cost associated with placing an order is \$10. In addition, there is a fixed “aggravation” charge of \$1 for each SKU included in the order. Usage of each of the three ingredients is relatively stable with time. The usages and basic unit values are as follows:

Item ID	Usage Rate (Units/Week)	Unit Value (\$/Unit)
CO-1	300	1.00
CO-2	80	0.50
CO-3	10	0.40

- a. If a coordinated replenishment strategy is to be used, what should be the values of the m_i 's?
- b. When pressed by a consulting analyst for a value of the carrying charge r (needed to compute the family cycle time T), Fishman replies: “I don't know from nothing concerning a carrying charge! All I know is that I'm satisfied with placing an order every 2 weeks.” How would you use this answer to impute an approximate value of r ? Discuss why it is likely to be only an approximation.
- c. What quantities of the three items should the Ptomaine Tavern order?
- d. Suppose that the supplier offers Ptomaine a 3% discount on any order totaling at least \$1,350 in value (before the discount). Fishman's partner, L. Talks, likes the idea of saving on the purchase price. Fishman is not so sure about the advisability of tying up so much money in inventory. Should Ptomaine take advantage of the quantity discount offer?

10.3 Suppose you wished to estimate the annual costs of treating the members of a family independently in setting up run quantities. Under independent treatment, some major setups would be avoided, simply by the chance happening of two items of the same family being run one right after the other. How would you estimate the fraction of setups where this would happen?

10.4 A certain large manufacturing company has just hired a new member of its industrial engineering department, a Ms. V. G. Rickson. Rickson, knowing all about coordinated control, selects a family of two items with the following properties:

Item i	D_i (Units/Year)	v_i (\$/Unit)
1	10,000	0.50
2	1,000	0.40

From accounting records and discussions with operating personnel, she estimates that $A = \$5$, $a_1 = \$1$, $a_2 = \$4$, and $r = 0.2$ \$/\$/year.

- a. She computes the best values of T , m_1 , and m_2 . What are these values?

- b. The production supervisor, Mr. C. W. Donrath, is skeptical of the value of coordination and argues that independent control is less costly, at least for this family of two items. Is he correct?
- 10.5 For the A, a_i 's deterministic demand context, consider a special coordinated strategy—namely, where all $m_i = 1$.
- Determine the optimal value of T .
 - Determine the associated minimum total relevant costs per unit time.
 - Find the best independent and coordinated (the latter with all $m_i = 1$ as above) strategies for each of the following two examples:

Example 1

$$A = \$10 \quad r = 0.2/\text{year}$$

Item i	D_i (Units/Year)	v_i (\$/Unit)	a_i (\$)
1	800	1	2
2	400	0.5	2

Example 2

$$A = \$10 \quad r = 0.2/\text{year}$$

Item i	D_i (Units/Year)	v_i (\$/Unit)	a_i (\$)
1	900	1	2
2	20	0.5	4

- Determine as simple a relationship as possible that must be satisfied in order for the above special case of coordinated control (every $m_i = 1$) to be preferable to completely independent control. Ignore any system costs in your analysis. Normalize where possible to reduce the number of parameters.
- 10.6 The Steady–Milver Corporation produces ball bearings. It has a family of three items, which, run consecutively, do not take much time for changeovers. The characteristics of the items are as follows:

Item i	ID	D_i (Units/Year)	Raw Material (\$/Unit)
1	BB1	2,000	2.50
2	BB2	1,000	2.50
3	BB3	500	1.60

Item i	Value Added (\$/Unit)	Value after Production v_i (\$/Unit)	a_i (\$)
1	0.50	3.00	5
2	0.50	3.00	2
3	0.40	2.00	1

The initial setup cost for the family is \$30. Management has agreed on an r value of 0.10 \$/\$/year. Production rates are substantially larger than the demand rates.

- a. What are the preferred run quantities of the three items?
 - b. Raw material for product BB1 is acquired from a supplier distinct from that for the other two products. Suppose that the BB1 supplier offers an 8% discount on all units if an order of 700 or more is placed. Should Steady–Milver take the discount offer?
- 10.7 Brown (1967) has suggested a procedure for allocating a total order (of size W dollars) among a group of n items ($i = 1, 2, \dots, n$). A particular time interval (perhaps until the next order arrives) of duration T is considered. The allocation is made so that the probability that item i runs out during T is proportional to the fraction of the group’s sales that item i contributes. (Brown shows that this procedure, at least approximately, minimizes the total stock remaining at the end of the interval.) Suppose that it is reasonable to assume that demand in period T for item i is normally distributed with mean \hat{x}_i and standard deviation σ_{T_i} .
- a. Introducing whatever symbols are necessary, develop a routine for allocating the order among the items according to the above criterion.
 - b. Illustrate for the following 3 item example, in which $W = \$900$:

Item i	\hat{x}_i (Units)	σ_T (Units)	v_i (\$/Unit)	Initial (Before Allocation) Inventory I_i (Units)
1	100	40	1.00	50
2	300	70	2.00	100
3	250	100	1.20	250

- c. Repeat part b but with

$$I_1 = 250 \quad I_2 = 150 \quad I_3 = 0$$

- 10.8 In Miltenburg’s coordinated control system, the residual stock of a particular item is denoted by the symbol z , and in the case of periodic review, it turns out to be reasonable to approximate its distribution by a normal distribution with mean μ_z and standard deviation σ_z .
- a. Why is there a spike at $z = 0$ in the continuous review case but not in the periodic review case? Also, why can z be negative in the periodic review situation?

- b. Using the following equation and assuming that the demand x during $R+L$ is normally distributed with mean \hat{x}_{R+L} and standard deviation σ_{R+L} , find as simple an expression as possible that the reorder point s must satisfy to ensure a probability of no stockout equal to P_1 .

$$\begin{aligned} \Pr\{\text{Stockout}\} &= \Pr\{x \geq s + z\} \\ &= \int_{z_0=-\infty}^{\infty} f_z(z_0) dz_0 \int_{x=s+z}^{\infty} f_x(x_0) dx_0 \end{aligned}$$

Hint: Use the following result proved by Miltenburg (1982).

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_z} \exp[-(z_0 - \mu_z)^2 / 2\sigma_z^2] p_{u \geq} \left(\frac{s + z_0 - \hat{x}_{R+L}}{\sigma_{R+L}} \right) dz_0 = p_{u \geq} \left(\frac{s + \mu_z - \hat{x}_{R+L}}{c\sigma_{R+L}} \right)$$

where

$$c = \sqrt{1 + \sigma_z^2 / \sigma_{R+L}^2}$$

- c. Find s for the case where $P_1 = 0.96$, $\mu_z = 5.0$, $\sigma_z = 2.3$, $\hat{x}_{R+L} = 12.0$, and $\sigma_{R+L} = 4.2$. What s value would be used if the residual stock was ignored?
- 10.9 Build a spreadsheet to solve the algorithm of Section 10.4.2.
- 10.10 Three products are ordered by a retailer from a single vendor. The product data are given below, and the following information has been gathered. The fixed cost to place an order is \$15, but once an order is placed, an item can be added to the order for \$2. The carrying charge is 22% and the delivery lead time is 2 weeks. Assume a 50-week year.
- Use the algorithm of Section 10.4.2 to find the order frequencies for each product.
 - If the fill rate target is 99%, use the methods of Chapter 6 to find the order-up-to-level for each product, given the results in part a.

Product	Annual Demand (Units)	Unit Cost (\$)	Standard Deviation of Monthly Demand (Units)	EOQ (Units)	Time Supply (Weeks)
1	1,000	\$2.00	20	95	4.77
2	500	\$25.00	22	19	1.91
3	5,000	\$9.45	120	98	0.98

- 10.11 Ten products are ordered by a distributor from a single supplier. The specific product data are given in the table below, and the following general information has been gathered. A is \$25, and $a = \$4$. The carrying charge is 18% and the delivery lead time is 1 week.
- Use the algorithm of Section 10.4.2 to find the order frequencies for each product.
 - If the fill rate target is 98%, use the methods of Chapter 6 to find the order-up-to-level for each product, given the results in part a.

Product	Monthly Demand (Units)	Unit Cost (\$)	Standard Deviation of Monthly Demand (Units)
1	8	3.00	2.0
2	25	20.00	8.0
3	4	6.00	1.0
4	63	52.00	10.0
5	67	16.00	30.0
6	46	4.00	20.0
7	54	0.98	20.0
8	2	120.00	0.5
9	83	20.00	25.0
10	82	10.00	18.0

10.12 A drink bottler uses one bottling line to fill three different sized soft drink containers with the same soft drink. Other drinks are bottled on the same line as well. We shall consider only one product whose sizes are 10, 16, and 20 ounces, and whose demands are constant enough to be safely considered deterministic. The major setup cost to convert to this type of drink is estimated to be \$65. It costs \$22 to switch from one container size to another. The demand rates for this family of items are shown below. The carrying charge has been established as 0.24 \$/\$/year. Moreover, there is substantial excess capacity in the operation, so it is reasonable to ignore capacity considerations.

Item <i>i</i>	1	2	3
Description	10 Ounces	16 Ounces	20 Ounces
D_i (units/year)	120,000	95,000	80,000
v_i (\$/unit)	0.30	0.48	0.60

a. Find the appropriate run quantities for each item.
 b. Using Equation 10.5, find a lower bound on the total cost.

10.13 A retailer orders eight products from a single vendor. Assume that the demand for each product is deterministic, and is given below. The major fixed ordering cost is estimated to be \$8; that is, the cost for the first item on the purchase order is \$8. It costs \$1 to add additional products to the purchase order. The carrying charge has been established as 0.30 \$/\$/year.

Item	Demand (Units/Year)	v_i (\$/Unit)
1	600	2.50
2	200	12.65

Item	Demand (Units/Year)	v_i (\$/Unit)
3	350	25.36
4	450	18.52
5	850	62.50
6	900	85.20
7	525	3.65
8	1,000	1.98

- a. Find the appropriate run quantities for each item.
 b. Using Equation 10.5, find a lower bound on the total cost.
- 10.14 Consider four parts used in the assembly of a particular end-item. These four parts are purchased from the same supplier who offers a discount of 4% if the value of the total replenishment quantity is at least \$200. Assume that \$5 is the basic cost of placing an order with one item involved. The inclusion of each additional item costs \$1. $r = 0.24$ \$/\$/year and the D_i 's and v_{0i} 's are shown below. Find the best order quantities.

Item i	D_i (Units/Year)	v_{0i} (\$/Unit)
1	5,000	1.20
2	10,000	2.25
3	2,500	6.25
4	500	0.85

- 10.15 Consider nine parts ordered from a single supplier. The supplier offers a discount of 6% if the value of the total replenishment quantity is at least \$10,000. Assume that \$4.50 is the basic cost of placing an order with one item involved. The inclusion of each additional item costs \$0.50. $r = 0.22$ \$/\$/year and the D_i 's and v_{0i} 's are shown below. Find the best order quantities.

Item i	D_i (Units/Year)	v_{0i} (\$/Unit)
1	2,000	11.00
2	16,000	15.00
3	1,300	18.00
4	1,000	25.00
5	3,000	32.00
6	250	65.00

Item i	D_i (Units/Year)	v_{0i} (\$/Unit)
7	850	25.00
8	9,000	14.00
9	28,000	16.00

- 10.16 A distributor stocks six products that are supplied by the same manufacturer. The data for each product is given below. The manufacturer owns her own trucks and requires full truckload shipments. The truck can carry 400 units of the products, in any combination. Management has set a 97.25% fill rate target for each product.

Product	D_i	Y_i	\hat{x}_{L+R}	σ_{L+R}
1	1,000	42	83	20
2	2,000	83	167	50
3	560	23	46	10
4	850	35	70	20
5	4,200	175	350	120
6	1,000	42	83	32

At a given review instant, assume that the inventory positions of the products are 50, 175, 30, 60, 350, and 100, respectively. Should an order be placed, and if so, how much should be ordered of each product?

- 10.17 A manufacturer produces four products on one machine. No other products are assigned to this equipment. The data for the item are given in the following table. r is 0.24 \$/\$/year.

Item	1	2	3	4
D_i (units/year)	2,000	500	8,000	5,000
v_i (\$/unit)	30	50	20	120
p_i (units/year)	50,000	2,000	20,000	24,000
A_i (\$)	75	120	110	60
K_i (hours)	5	6	8	4
v_i (\$/unit)	15	16	7.80	12.25

- From the results of Section 10.6.1, find the optimal order interval and quantities, assuming each product is produced each cycle.
 - Briefly comment on your answer for a real environment.
 - What is the ratio of the cost of this pure rotation schedule to the optimal?
- 10.18 National Electronics distributes a wide variety of electronic products. National has been experiencing cash flow problems; hence it is concerned about its investment in inventories. A management consultant, Mr. P. C. Snow, has been hired to suggest ordering decision

rules. As a preliminary step, he selects for study an important group of 10 products all bought from the same supplier, Harolds Corp. Based on an analysis of ordering costs, Mr. Snow estimates that the replenishment cost, A , in dollars, is given by

$$A = 46 + 5.5(V - 1,000)/1,000 \tag{10.18}$$

where V is the total dollar value of the order and a minimum order size of \$1,000 is required by Harolds. The National Electronics Controller, Mr. P. J. Schmeidler, tells Snow that the company uses an r value of 0.24 \$/\$/year. The sales forecast for each of the next 12 months, as well as the unit values, are shown in Exhibit 1. Snow ascertains that the requirements for any particular month must be on hand at the beginning of the month; hence he can restrict his attention to replenishments covering an integer number of months. In addition, at least initially, he decides to consider a deterministic version of the problem, namely, where lead times are known and the forecasts represent actual demands. Snow observes that because a marked seasonality exists, a simple EOQ strategy is inappropriate. He considers two possible types of decision rules:

EXHIBIT 1 Forecast (Units)

		2016				2017							
Item	Unit Value v_i \$/Unit	S	O	N	D	J	F	M	A	M	J	J	A
1	70.40	4	5	5	5	5	3	2	3	2	2	2	2
2	8.40	30	30	40	40	40	30	30	20	10	10	10	10
3	49.00	15	15	15	15	15	15	11	11	9	9	10	12
4	51.40	1	1	2	2	2	1	1	1	1	1	1	1
5	15.70	24	30	30	30	30	18	12	18	12	12	12	12
6	12.40	30	30	30	30	30	20	20	20	15	15	15	15
7	5.30	31	31	35	35	35	21	21	21	15	15	10	10
8	10.10	36	50	50	50	45	25	25	18	18	15	15	15
9	18.10	33	33	44	44	44	33	33	22	22	11	11	11
10	110.50	10	10	10	10	10	10	8	8	8	6	6	6

1. Order the same fixed time supply (e.g., 2 months coverage) of all items.
2. Use a modified version of the Silver–Meal heuristic (Section 5.6 of Chapter 5), a modification is necessary because of the nature of the setup cost A which can be shared among two or more items involved in the same order.

Snow argues that any reasonable decision rule would require a group replenishment to arrive at the beginning of September, the start of the peak demand period. Hence, he assumes 0 inventories of all the items just prior to the beginning of September 2016.

- a. For the fixed time supply (T months) strategy, he evaluates the order quantities, replenishment costs, and carrying costs for each of $T = 1, 2,$ and 3 months. Develop these quantities and select what appears to be the best T value.

- b. In an effort to appropriately modify the Silver–Meal heuristic, Snow reasons as follows. Suppose that a particular order involves n items and has a total value of V . Then, the setup cost A is given by Equation 10.18. Hence, an approximate apportioned cost for each item involved in the order is

$$\frac{46 + 5.5(V - 1,000)/1,000}{n}$$

The problem is that V depends on the order quantities, which, in turn, should depend upon the setup cost. This suggests the use of an iterative procedure. Historically, the average total value of an order involving n items has been approximately $\$600n$. Thus he starts with the Silver–Meal heuristic for each individual item using an A value of

$$\begin{aligned} A &= \frac{46 + 5.5(60V - 1,000)/1,000}{n} \\ &= 3.30 + 40.5/n \end{aligned} \quad (10.19)$$

The actual total value of the resulting group replenishment is determined from the individual order quantities. This implies a new value of A , the heuristic is reused, a new V is obtained, etc. until convergence takes place. Illustrate in detail these computations for the first replenishment (at the start of September), which must include all 10 items. Once a stable set of order quantities is established for a particular group replenishment, Snow moves on to the next time where at least one item requires another replenishment. The above iterative procedure involving A and V is used for that group of items. Complete, in this fashion, the 12 months of the modified heuristic computations. Determine the total replenishment and carrying costs and compare with your answer in part a.

- c. Rather than iterating as above, suppose instead that for each group replenishment the individual item A was established according to Equation 10.19 and the Silver–Meal results with this A value were used. Carry through the calculations on this basis and evaluate the total replenishment and carrying costs.
- d. Which approach do you think that Snow should recommend?

Appendix 10A: Derivation of Results in Section 10.2

If a time interval T between replenishments of the family and a set of m_i 's are used, where m_i , an integer, is the number of T intervals that the replenishment quantity of item i will last, then this replenishment quantity (Q_i) is given by

$$Q_i = D_i m_i T \quad (10A.1)$$

The typical sawtooth diagram of inventory is applicable so that the average inventory of item i (\bar{I}_i) is

$$\bar{I}_i = \frac{Q_i}{2} = \frac{D_i m_i T}{2}$$

A group setup cost (A) is incurred every T units of time, whereas the cost a_i is incurred only once in every $m_i T$ units of time. Therefore, the total relevant costs per unit time are given by

$$\text{TRC}(T, m_i's) = \frac{A + \sum_{i=1}^n a_i/m_i}{T} + \sum_{i=1}^n \frac{D_i m_i T v_i r}{2} \quad (10A.2)$$

(Note that we do not include the purchase cost in the total relevant costs in this case. However, we do need to include it when a quantity discount is available.)

Setting

$$\frac{\partial \text{TRC}}{\partial T} = 0$$

gives the best T for the particular set of m_i 's, that is,

$$-\frac{A + \sum_{i=1}^n a_i/m_i}{T^2} + \frac{r}{2} \sum_{i=1}^n m_i D_i v_i = 0$$

or

$$T^*(m_i's) = \sqrt{\frac{2(A + \sum a_i/m_i)}{r \sum m_i D_i v_i}} \quad (10A.3)$$

which is Equation 10.2.

Substitution of Equation 10A.3 back into Equation 10A.2 gives the best cost for a given set of m_i 's:

$$\text{TRC}^*(m_i's) = \frac{A + \sum_{i=1}^n a_i/m_i}{\sqrt{\frac{2(A + \sum a_i/m_i)}{r \sum m_i D_i v_i}}} + \sqrt{\frac{2(A + \sum a_i/m_i)}{r \sum m_i D_i v_i}} \frac{r \sum m_i D_i v_i}{2}$$

which simplifies to

$$\text{TRC}^*(m_i's) = \sqrt{2 \left(A + \sum a_i/m_i \right) r \sum m_i D_i v_i} \quad (10A.4)$$

(This equation was used to directly produce Equation 10.10.)

We wish to select the m_i 's to minimize $\text{TRC}^*(m_i's)$. From an inspection of Equation 10A.4, this is achieved by selecting the m_i 's to minimize

$$F(m_i's) = \left(A + \sum a_i/m_i \right) \sum m_i D_i v_i \quad (10A.5)$$

The minimization of Equation 10A.5 is no simple matter because of two facts: (1) the m_i 's interact (i.e., the effects of one m_i value depend on the values of the other m_i 's) and (2) the m_i 's must be integers (see Schweitzer and Silver 1983).

If we choose to ignore the integer constraints on the m_i 's and set partial derivatives of $F(m_i's)$ equal to zero (necessary conditions for a minimum), then

$$\frac{\partial F(m_i's)}{\partial m_j} = -\frac{a_j}{m_j^2} \sum_i m_i D_i v_i + D_j v_j \left(A + \sum a_i/m_i \right) = 0$$

or

$$m_j^2 = \frac{a_j \sum m_i D_i v_i}{D_j v_j (A + \sum a_i / m_i)} \quad j = 1, 2, \dots, n \quad (10A.6)$$

For $j \neq k$, we have

$$m_k^2 = \frac{a_k}{D_k v_k} \frac{\sum m_i D_i v_i}{(A + \sum a_i / m_i)}$$

Dividing gives

$$\frac{m_j^2}{m_k^2} = \frac{a_j}{D_j v_j} \frac{D_k v_k}{a_k}$$

or

$$\frac{m_j}{m_k} = \sqrt{\frac{a_j}{D_j v_j} \frac{D_k v_k}{a_k}} \quad j \neq k$$

We can see that if

$$\frac{a_j}{D_j v_j} < \frac{a_k}{D_k v_k}$$

then (the continuous solution) m_j is less than (the continuous solution) m_k . Therefore, the item i having the smallest value of $a_i/D_i v_i$ should have the lowest value of m_i , namely, 1. It is reasonable to assume that this will hold even when the m_j 's are restricted to being integers; of course, in this case, more than one item could have $m_j = 1$.

If the items are numbered such that item 1 has the smallest value of $a_i/D_i v_i$, then

$$m_1 = 1 \quad (10A.7)$$

and, from Equation 10A.6,

$$m_j = \sqrt{\frac{a_j}{D_j v_j}} \sqrt{\frac{\sum m_i D_i v_i}{(A + \sum a_i / m_i)}} \quad j = 2, 3, \dots, n \quad (10A.8)$$

Suppose that there is a solution to these equations that results in

$$\sqrt{\frac{\sum m_i D_i v_i}{(A + \sum a_i / m_i)}} = C \quad (10A.9)$$

Then, from Equation 10A.8, we have

$$m_j = C \sqrt{\frac{a_j}{D_j v_j}} \quad j = 2, 3, \dots, n \quad (10A.10)$$

Therefore,

$$\sum_{i=1}^n m_i D_i v_i = D_1 v_1 + \sum_{i=2}^n C \sqrt{\frac{a_i}{D_i v_i}} D_i v_i = D_1 v_1 + C \sum_{i=2}^n \sqrt{a_i D_i v_i} \quad (10A.11)$$

Similarly,

$$\sum_{i=1}^n a_i/m_i = a_1 + \frac{1}{C} \sum_{i=2}^n \sqrt{a_i D_i v_i} \quad (10A.12)$$

Substituting Equations 10A.11 and 10A.12 back into the left-hand side of Equation 10A.9 and squaring, we obtain

$$\frac{D_1 v_1 + C \sum_{i=2}^n \sqrt{a_i D_i v_i}}{A + a_1 + \frac{1}{C} \sum_{i=2}^n \sqrt{a_i D_i v_i}} = C^2$$

Cross-multiplication gives

$$D_1 v_1 + C \sum_{i=2}^n \sqrt{a_i D_i v_i} = C^2(A + a_1) + C \sum_{i=2}^n \sqrt{a_i D_i v_i}$$

or

$$C = \sqrt{\frac{D_1 v_1}{A + a_1}}$$

Substitution of this expression back into Equation 10A.10 gives

$$m_j = \sqrt{\frac{a_j}{D_j v_j} \frac{D_1 v_1}{A + a_1}} \quad j = 2, 3, \dots, n \quad (10A.13)$$

which is Equation 10.3.

To find a lower bound on the best possible total cost, we substitute $m_1 = 1$ and the generally noninteger m_j 's of Equation 10A.13 into Equation 10A.4. This leads, after considerable algebra, to the result of Equation 10.5.

References

- Agnihotri, S., U. Karmarkar, and P. Kubat. 1982. Stochastic allocation rules. *Operations Research* 30(3), 545–555.
- Allen, S. J. 1990. Production rate planning for two products sharing a single process facility: A real-world case study. *Production and Inventory Management Journal* 31(3), 24–29.
- Anupindi, R. and S. Tayur. 1998. Managing stochastic multi-product systems: Model, measures, and analysis. *Operations Research* 46(3 supp), S98–S111.
- Atkins, D. and P. Iyogun. 1988. Periodic versus “can-order” policies for coordinated multi-item inventory systems. *Management Science* 34(6), 791–796.
- Atkins, D. and D. Sun. 1995. 98% Effective lot sizing for series inventory systems with backlogging. *Operations Research* 43(2), 335–345.
- Atkins, D. R. 1991. The inventory joint replenishment problem with a general class of joint costs. *European Journal of Operational Research* 51, 310–312.
- Atkins, D. R. and P. Iyogun. 1987. A lower bound on a class of inventory/production problems. *Operations Research Letters* 6(2), 63–67.
- Axsäter, S. 1987. An extension of the extended basic period approach for economic lot scheduling problems. *Journal of Optimization Theory and Applications* 52, 179–189.
- Bahl, H. and L. Ritzman. 1984. An integrated model for master scheduling, lot sizing and capacity requirements planning. *Journal of the Operational Research Society* 35(5), 389–399.

- Bahl, H. C. and S. Zionts. 1987. Multi-item scheduling by Benders' decomposition. *Journal of the Operational Research Society* 38(2), 1141–1148.
- Baker, K. R., P. Dixon, M. J. Magazine, and E. A. Silver. 1978. An algorithm for the dynamic lot size problem with time varying production constraints. *Management Science* 24, 1710–1720.
- Balintfy, J. 1964. On a basic class of multi-item inventory problems. *Management Science* 10(2), 287–297.
- Bastian, M. 1986. Joint replenishments in multi-item inventory systems. *Journal of the Operational Research Society* 37(12), 1113–1120.
- Bausch, D. O., G. G. Brown, and D. Ronen. 1995. Consolidating and dispatching truck shipments of mobil heavy petroleum products. *Interfaces* 25(March–April), 1–17.
- Billington, P., J. Blackburn, J. Maes, R. Millen, and L. van Wassenhove. 1994. Multi-item lotsizing in capacitated multi-stage serial systems. *IIE Transactions* 26(2), 12–17.
- Billington, P. J. 1986. The capacitated multi-item dynamic lot-sizing problem. *IIE Transactions* 18(2), 217–219.
- Bitran, G. R. and H. Matsuo. 1986. Approximation formulations for the single-product capacitated lot size problem. *Operations Research* 34, 63–74.
- Blackburn, J. and R. Millen. 1984. Simultaneous lot-sizing and capacity planning in multi-stage assembly processes. *European Journal of Operational Research* 16, 84–93.
- Bolander, S. and S. Taylor. 1983. Time phased forward scheduling: A capacity dominated scheduling technique. *Production and Inventory Management Journal* 24(1), 83–97.
- Bolander, S. F. and S. G. Taylor. 1990. Process flow scheduling: Mixed-flow cases. *Production and Inventory Management Journal* 31(4), 1–6.
- Bourland, K. and C. Yano. 1994. The strategic use of capacity slack in the economic lot scheduling problem with random demand. *Management Science* 40(12), 1690–1704.
- Bourland, K. and C. Yano. 1997. A comparison of solutions approaches for the fixed-sequence economic lot scheduling problem. *IIE Transactions* 29, 103–108.
- Bowman, R. A. and J. A. Muckstadt. 1993. Stochastic analysis of cyclic schedules. *Operations Research* 41(5), 947–958.
- Bowman, R. A. and J. A. Muckstadt. 1995. Production control of cyclic schedules with demand and process variability. *Production and Operations Management* 4(2), 145–162.
- Brahimi, N., S. Dauzere-Peres, N. M. Najid, and A. Nordli. 2006. Single item lot sizing problems. *European Journal of Operational Research* 168(1), 1–16.
- Bramel, J., S. Goyal, and P. Zipkin. 2000. Coordination of production/distribution networks with unbalanced leadtimes. *Operations Research* 48(4), 570–577.
- Bretthauer, K., B. Shetty, S. Syam, and S. White. 1994. A model for resource constrained production and inventory management. *Decision Sciences* 25(4), 561–577.
- Brown, R. G. 1967. *Decision Rules for Inventory Management*. New York: Holt, Rinehart and Winston.
- Campbell, G. 1992. Using short-term dedication for scheduling multiple products on parallel machines. *Production and Operations Management* 1(3), 295–307.
- Carlson, M. and J. Miltenburg. 1988. Using the service point model to control large groups of items. *OMEGA* 16(5), 481–489.
- Carr, S. A., A. R. Gullu, P. L. Jackson, and J. A. Muckstadt. 1993. An Exact Analysis of a Production-Inventory Strategy for Industrial Suppliers. Cornell University Operations Research and Industrial Engineering.
- Cattrysse, D., M. Salomon, R. Kuik, and L. van Wassenhove. 1993. A dual ascent and column generation heuristic for the discrete lotsizing and scheduling problem with setup times. *Management Science* 39(4), 477–486.
- Chakravarty, A. K. 1984. Joint inventory replenishments with group discounts based on invoice value. *Management Science* 30(9), 1105–1112.
- Chakravarty, A. K. 1985. Multiproduct purchase scheduling with limited budget and/or group discounts. *Computer & Operations Research* 12(5), 493–505.
- Chakravarty, A. K. and G. E. Martin. 1988. An optimal joint buyer-seller discount pricing model. *Computers and Operations Research* 15(3), 271–281.

- Chen, C. and K. Min. 1994. A multi-product EOQ model with pricing consideration T.C.E. Cheng's model revisited. *Computers & Industrial Engineering* 26(4), 787–794.
- Coleman, B. J. and M. A. McKnew. 1995. A quick and effective method for capacitated lot sizing with startup and reservation costs. *Computers and Operations Research* 22(6), 641–653.
- Dagpunar, J. S. 1988. On the relationship between constrained and unconstrained multi-item replenishment problems. *International Journal of Operations & Production Management* 9(4), 77–81.
- Davis, S. G. 1995. An improved algorithm for solving the economic lot size problem (ELSP). *International Journal of Production Research* 33(4), 1007–1026.
- De Bodt, M., L. Gelders, and L. Van Wassenhove. 1984. Lot sizing under dynamic demand conditions: A review. *Engineering Costs and Production Economics* 8, 165–187.
- Dekker, R., R. E. Wildeman, and F. A. van der Duyn Schouten. 1997. A review of multi-component maintenance models with economic dependence. *Mathematical Methods of Operations Research* 45(3), 411–435.
- Delporte, C. and L. J. Thomas. 1977. Lot sizing and sequencing for N periods on one facility. *Management Science* 23(10), 1070–1079.
- DeMatta, R. and M. Guignard. 1995. The performance of rolling production schedules in a process industry. *IIE Transactions* 27, 564–573.
- Diaby, M., H. Bahl, M. Karwan, and S. Zionts. 1992. A Lagrangian relaxation approach for very-large-scale capacitated lot-sizing. *Management Science* 38(9), 1329–1340.
- Dixon, P. S. 1979. *Multi-Item Lot-Sizing with Limited Capacity*. PhD thesis. University of Waterloo, Waterloo, Canada.
- Dixon, P. S. and C. L. Poh. 1990. Heuristic procedures for multi-item inventory planning with limited storage. *IIE Transactions* 22(2), 112–123.
- Dixon, P. S. and E. A. Silver. 1981. A heuristic solution procedure for the multi-item, single-level, limited capacity, lot-sizing problem. *Journal of Operations Management* 2(1), 23–39.
- Dobson, G. 1987. The economic lot-scheduling problem: Achieving feasibility using time-varying lot sizes. *Operations Research* 35, 764–771.
- Dobson, G. 1992. The cyclic lot scheduling problem with sequence dependent setups. *Operations Research* 40, 736–749.
- Dobson, G. and C. A. Yano. 1994. Cyclic scheduling to minimize inventory in a batch flow line. *European Journal of Operational Research* 75, 441–461.
- Duenyas, I. and M. P. Van Oyen. 1995. Stochastic scheduling of parallel queues with set-up costs. *Queueing Systems* 19(4), 421–444.
- Duenyas, I. and M. P. Van Oyen. 1996. Heuristic scheduling of parallel heterogeneous queues with set-ups. *Management Science* 42(6), 814–829.
- Elmaghraby, S. 1978. The economic lot scheduling problem (ELSP): Review and extensions. *Management Science* 24(6), 587–598.
- Eppen, G. D. and R. K. Martin. 1987. Solving multi-item capacitated lot-sizing problems using variable redefinition. *Operations Research* 35(6), 832–848.
- Erenguc, S. S. 1988. Multiproduct dynamic lot-sizing model with coordinated replenishments. *Naval Research Logistics* 35, 1–22.
- Ernst, R. and D. Pyke. 1992. Component part stocking policies. *Naval Research Logistics* 39, 509–529.
- Ettl, M., G. E. Feigin, G. Y. Lin, and D. D. Yao. 2000. A supply network model with base-stock control and service requirements. *Operations Research* 48(2), 216–232.
- Eynan, A. and D. H. Kropp. 1998. Periodic review and joint replenishment in stochastic demand environments. *IIE Transactions* 30(11), 1025–1033.
- Federgruen, A., H. Groenevelt, and H. C. Tijms. 1984. Coordinated replenishments in a multi-item inventory system with compound Poisson demand. *Management Science* 30(3), 344–357.
- Federgruen, A. and Z. Katalan. 1994. Approximating queue size and waiting time distributions in general polling systems. *Queueing Systems* 18, 353–386.
- Federgruen, A. and Z. Katalan. 1996a. Customer waiting time distributions under base-stock policies in single facility multi-item production systems. *Naval Research Logistics* 43, 533–548.

- Federgruen, A. and Z. Katalan. 1996b. The impact of setup times on the performance of multi-class service and production systems. *Operations Research* 44(6), 989–1001.
- Federgruen, A. and Z. Katalan. 1996c. The stochastic economic lot scheduling problem: Cyclical base-stock policies with idle times. *Management Science* 42(6), 783–796.
- Federgruen, A. and Z. Katalan. 1998. Determining production schedules under base-stock policies in single facility multi-item production systems. *Operations Research* 46(6), 883.
- Florian, M. and M. Klein. 1971. Deterministic production planning with concave costs and capacity constraints. *Management Science* 18, 12–20.
- Fogarty, D. W. and R. L. Barringer. 1987. Joint order release decisions under dependent demand. *Production and Inventory Management Journal* 28(1), 55–61.
- Gallego, G. 1990. An extension to the class of easy economic lot scheduling problems. *IIE Transactions* 22(2), 189–190.
- Gallego, G. 1992. A minmax distribution free procedure for the (Q, R) inventory model. *Operations Research Letters* 11(1), 55–60.
- Gallego, G. 1993. Reduced production rates in the economic lot scheduling problem. *International Journal of Production Research* 31(5), 1035–1046.
- Gallego, G. 1994. When is a base stock policy optimal in recovering disrupted cyclic schedules? *Naval Research Logistics* 41, 317–333.
- Gallego, G. and I. Moon. 1992. The effect of externalizing setups in the economic lot scheduling problem. *Operations Research* 40(3), 614–619.
- Gallego, G. and I. Moon. 1995. Strategic investment to reduce setup times in the economic lot scheduling problem. *Naval Research Logistics* 42(5), 773–790.
- Gallego, G. and I. Moon. 1996. How to avoid stockouts when producing several items on a single facility? What to do if you can't? *Computers and Operations Research* 23(1), 1–12.
- Gallego, G. and M. Queyranne. 1995. Inventory coordination and pricing decisions: Analysis of a simple class of heuristics. In A. Sciomachen (Ed.), *Optimization in Industry 3*, New York: John Wiley & Sons.
- Gallego, G., M. Queyranne, and D. Simchi-Levi. 1996. Single resource multi-item inventory systems. *Operations Research* 44(4), 580–595.
- Gallego, G. and D. X. Shaw. 1997. Complexity of the ELSP with general cyclic schedules. *IIE Transactions* 29, 109–113.
- Gardiner, S. and J. Blackstone. 1995. Setups and effective capacity: The impact of lot sizing techniques in an MRP environment. *Production Planning & Control* 6(1), 26–38.
- Gascon, A., R. C. Leachman, and P. Lefrancois. 1994. Multi-item, single-machine scheduling problem with stochastic demands: A comparison of heuristics. *International Journal of Production Research* 32(3), 583–596.
- Gilbert, K. and M. Madan. 1991. A heuristic for a class of production planning and scheduling problems. *IIE Transactions* 23(3), 282–289.
- Golany, B. and A. Lev-er. 1992. Comparative analysis of multi-item joint replenishment inventory models. *International Journal of Production Research* 30(8), 1791–1801.
- Goncalves, J., R. Leachman, A. Gascon, and Z. Xiong. 1994. A heuristic scheduling policy for multi-item, multi-machine production systems with time-varying, stochastic demands. *Management Science* 40(11), 1455–1468.
- Goyal, S. K. 1974. A decision rule for producing Christmas greeting cards. *Operations Research* 22(4), 795–801.
- Goyal, S. K. 1988. Economic ordering policy for jointly replenished items. *International Journal of Production Research* 26(7), 1237–1240.
- Goyal, S. K. and S. G. Deshmukh. 1993. A note on “the economic ordering quantity for jointly replenishing items”. *International Journal of Production Research* 31(12), 2959–2961.
- Goyal, S. K. and A. T. Satir. 1989. Joint replenishment inventory control: Deterministic and stochastic models. *European Journal of Operational Research* 38, 2–13.
- Gross, D., C. Harris, and P. Robers. 1972. Bridging the gap between mathematical inventory theory and the construction of a workable model—A case study. *International Journal of Production Research* 10(3), 201–214.

- Güder, F., J. Zydiak, and S. Chaudhry. 1995. Non-stationary ordering policies for multi-item inventory systems subject to a single resource constraint. *Journal of the Operational Research Society* 46, 1145–1152.
- Güder, F. and J. L. Zydiak. 1999. Ordering policies for multi-item inventory systems subject to multiple resource constraints. *Computers and Operations Research* 26(6), 583–597.
- Gunther, H. O. 1987. Planning lot sizes and capacity requirements in a single stage production system. *European Journal of Operational Research* 31, 223–231.
- Hahm, J. and C. Yano. 1992. The economic lot and delivery scheduling problem: The single item case. *International Journal of Production Economics* 28, 235–252.
- Hahm, J. and C. A. Yano. 1995a. The economic lot and delivery scheduling problem: Models for nested schedules. *IIE Transactions* 27, 126–139.
- Hahm, J. and C. A. Yano. 1995b. The economic lot and delivery scheduling problem: The common cycle case. *IIE Transactions* 27, 113–125.
- Haksever, C. and J. Moussourakis. 2005. A model for optimizing multi-product inventory systems with multiple constraints. *International Journal of Production Economics* 97(1), 18–30.
- Hall, N. G. 1988. A multi-item EOQ model with inventory cycle balancing. *Naval Research Logistics* 35, 319–325.
- Hall, R. W. and M. Racer. 1995. Transportation with common carrier and private fleets: System assignment and shipment frequency optimization. *IIE Transactions* 27, 217–225.
- Hariga, M. 1994. Two new heuristic procedures for the joint replenishment problem. *Journal of the Operational Research Society* 45(4), 463–471.
- Hariri, A. M. A., M. O. Abou-El-Ata, and S. K. Goyal. 1995. The resource constrained multi-item inventory problem with price discount: A geometric programming approach. *Production Planning & Control* 6(4), 374–377.
- Higginson, J. K. and J. H. Bookbinder. 1994. Policy recommendations for a shipment-consolidation program. *Journal of Business Logistics* 15(1), 87–112.
- Higginson, J. K. and J. H. Bookbinder. 1995. Markovian decision processes in shipment consolidation. *Transportation Science* 29(3), 242–255.
- Hill, A. V. and A. S. Raturi. 1992. A model for determining tactical parameters for materials requirements planning systems. *Journal of Operational Research Society* 43(6), 605–620.
- Hindi, K. S. 1995. Efficient solution of the single-item, capacitated lot-sizing problem with start-up and reservation costs. *Journal of the Operational Research Society* 46, 1223–1236.
- Hong, J. and J. C. Hayya. 1992. An optimal algorithm for integrating raw materials inventory in a single-product manufacturing system. *European Journal of Operational Research* 59, 313–318.
- Hsu, W. 1983. On the general feasibility test of scheduling lot sizes for several products on one machine. *Management Science* 29, 93–105.
- Hwang, H., D. B. Kim, and Y. D. Kim. 1993. Multiproduct economic lot size models with investment costs for setup, reduction and quality improvement. *International Journal of Production Research* 31(3), 691–703.
- Ignall, E. 1969. Optimal continuous review policies for two product inventory systems with joint setup costs. *Management Science* 15(5), 278–283.
- Iyogun, P. 1991. Heuristic methods for the multi-product dynamic lot size problem. *Journal of the Operational Research Society* 42(10), 889–894.
- Jackson, P. 1988. Stock allocation in a two-echelon distribution system or “what to do until your ship comes in”. *Management Science* 34(7), 880–895.
- Jackson, P., W. Maxwell, and J. Muckstadt. 1985. The joint replenishment problem with a powers-of-two restriction. *IIE Transaction* 17(1), 25–32.
- Johnston, F. R. 1980. An interactive stock control system with a strategic management role. *Journal of the Operational Research Society* 31(12), 1069–1084.
- Joneja, D. 1990. The joint replenishment problem: New heuristics and worst case performance bounds. *Operations Research* 38(4), 711–723.
- Kao, E. 1979. A multi-product dynamic lot-size model with individual and joint set-up costs. *Operations Research* 27(2), 279–289.

- Karmarkar, U., S. Kekre, and S. Kekre. 1987. The dynamic lot-sizing problem with startup and reservation costs. *Operations Research* 35, 389–398.
- Karmarkar, U. S. and L. Schrage. 1985. The deterministic dynamic product cycling problem. *Operations Research* 33, 326–345.
- Karmarkar, U. S. and J. Yoo. 1994. The stochastic dynamic product cycling problem. *European Journal of Operational Research* 73, 360–373.
- Karni, R. and Y. Roll. 1982. A heuristic algorithm for the multi-item lot-sizing problem with capacity constraints. *IIE Transactions* 14(4), 249–256.
- Kaspi, M. and M. J. Rosenblatt. 1991. On the economic ordering quantity for jointly replenished items. *International Journal of Production Research* 29(1), 107–114.
- Katz, P., A. Sadrian, and P. Tendick. 1994. Telephone companies analyze price quotations with Bellcore's PDSS software. *Interfaces* 24(1), 50–63.
- Khouja, M. and S. Goyal. 2008. A review of the joint replenishment problem literature: 1989–2005. *European Journal of Operational Research* 186(1), 1–16.
- Kirca, O. 1995. A primal-dual algorithm for the dynamic lotsizing problem with joint set-up costs. *Naval Research Logistics* 42, 791–806.
- Klein, C. M. and J. A. Ventura. 1995. An optimal method for a deterministic joint replenishment inventory policy in discrete time. *Journal of the Operational Research Society* 46, 649–657.
- Lambert, A. M. and H. Luss. 1982. Production planning with time-dependent capacity bounds. *European Journal of Operational Research* 9, 275–280.
- Lambrecht, M. and H. Vanderveken. 1979. Heuristic procedures for the single operation, multi-item loading problem. *AIE Transactions* 11(4), 319–326.
- Leachman, R. and A. Gascon. 1988. A heuristic scheduling policy for multi-item, single-machine production systems with time-varying, stochastic demands. *Management Science* 34(3), 377–390.
- Leachman, R., Z. Xiong, A. Gascon, and K. Park. 1991. An improvement to the dynamic cycle lengths heuristic for scheduling the multi-item, single-machine. *Management Science* 37(9), 1201–1205.
- Lee, W. J. 1994. Optimal order quantities and prices with storage space and inventory investment limitations. *Computers in Industrial Engineering* 26(3), 481–488.
- Loerch, A. G. and J. A. Muckstadt. 1994. An approach to production planning and scheduling in cyclically scheduled manufacturing systems. *International Journal of Production Research* 32(4), 851–871.
- Lotfi, V. and Y. Yoon. 1994. An algorithm for the single-item capacitated lot-sizing problem with concave production and holding costs. *Journal of Operational Research Society* 45(8), 934–941.
- Low, R. A. and J. F. Waddington. 1967. The determination of the optimum joint replenishment policy for a group of discount-connected stock lines. *Operational Research Quarterly* 18(4), 443–457.
- Lu, L. 1995. A one-vendor multi-buyer integrated inventory model. *European Journal of Operational Research* 81, 312–323.
- Maes, J., J. O. McClain, and L. N. Van Wassenhove. 1991. Multilevel capacitated lotsizing complexity and LP-based heuristics. *European Journal of Operational Research* 53, 131–148.
- Maes, J. and L. Van Wassenhove. 1986. Multi item single level capacitated dynamic lotsizing heuristics: A computational comparison. *IIE Transactions* 18(2), 114–123.
- Maes, J. and L. Van Wassenhove. 1988. Multi-item, single level capacitated dynamic lot-sizing heuristics: A general review. *Journal of the Operational Research Society* 39(11), 991–1004.
- Markowitz, D. M., M. I. Reiman, and L. M. Wein. 2000. The stochastic economic lot scheduling problem: Heavy traffic analysis of dynamic cyclic policies. *Operations Research* 48(1), 136–154.
- Martel, A., M. Diaby, and F. Boctor. 1995. Multiple items procurement under stochastic nonstationary demands. *European Journal of Operational Research* 87, 74–92.
- Maxwell, W. 1964. The scheduling of economic lot sizes. *Naval Research Logistics Quarterly* 11(2–3), 89–124.
- Maxwell, W. L. and H. Singh. 1983. The effect of restricting cycle times in the economic lot scheduling problem. *IIE Transactions* 15(3), 235–241.
- McGee, V. E. and D. F. Pyke. 1996. Periodic production scheduling at a fastener manufacturer. *International Journal of Production Economics* 46–47, 65–87.

- Mehrez, A. and D. Benn-Arieh. 1991. All-unit discounts, multi-item inventory model with stochastic demand, service level constraints and finite horizon. *International Journal of Production Research* 29(8), 1615–1628.
- Mercan, H. M. and S. S. Erenguc. 1993. A multi-family dynamic lot sizing problem with coordinated replenishments: A heuristic procedure. *International Journal of Production Research* 31(1), 173–189.
- Millar, H. and M. Yang. 1994. Lagrangian heuristics for the capacitated multi-item lot-sizing problem with backordering. *International Journal of Production Economics* 34, 1–15.
- Miltenburg, G. J. 1982. *The Co-ordinated Control of a Family of Discount-Related Items*. PhD thesis. University of Waterloo, Waterloo, Canada.
- Miltenburg, G. J. 1985. Allocating a replenishment order among a family of items. *IIE Transactions* 17(3), 261–267.
- Miltenburg, G. J. 1987. Co-ordinated control of a family of discount-related items. *INFOR* 25(2), 97–116.
- Miltenburg, J. G. and E. A. Silver. 1984a. Accounting for residual stock in continuous review coordinated control of a family of items. *International Journal of Production Research* 22, 607–628.
- Miltenburg, J. G. and E. A. Silver. 1984b. The diffusion process and residual stock in periodic review coordinated control of families of items. *International Journal of Production Research* 22, 629–646.
- Miltenburg, J. G. and E. A. Silver. 1984c. The evaluation of discounts in co-ordinated control inventory models. *Engineering Costs and Production Economics* 8, 15–25.
- Miltenburg, J. G. and E. A. Silver. 1988. A microcomputer inventory control package for controlling families of items. *Engineering Costs and Production Economics* 15, 201–209.
- Moon, I. 1994. Multiproduct economic lot size models with investment costs for setup reduction and quality improvement: Review and extensions. *International Journal of Production Research* 32(12), 2795–2801.
- Moon, I., G. Gallego, and D. Simchi-Levi. 1991. Controllable production rates in a family production context. *International Journal of Production Research* 29(12), 2459–2470.
- Muckstadt, J. A. and R. O. Roundy. 1993. Analysis of multistage production systems. In S. C. Graves, A. H. G. Rinnooy Kan, and P. H. Zipkin (Eds.), *Logistics of Production and Inventory*, Volume 4, pp. 59–131. Amsterdam: North-Holland.
- Naddor, E. 1975. Optimal heuristic decisions for the s, S inventory policy. *Management Science* 21(9), 1071–1072.
- Nielsen, C. and C. Larsen. 2005. An analytical study of the $q(s, s)$ policy applied to the joint replenishment problem. *European Journal of Operational Research* 163(3), 721–732.
- Pantumsinchai, P. 1992. A comparison of three joint ordering inventory policies. *Decision Sciences* 23, 111–127.
- Perez, A. P. and P. Zipkin. 1997. Dynamic scheduling rules for a multiproduct make-to-stock queue. *Operations Research* 45(6), 919–930.
- Pochet, Y. and L. A. Wolsey. 1991. Solving multi-item lot-sizing problems using strong cutting planes. *Management Science* 37(1), 53–67.
- Potamianos, J., A. J. Orman, and A. K. Shahani. 1996. Modeling for a dynamic inventory-production control system. *European Journal of Operational Research* 96, 645–658.
- Prentis, E. L. and B. M. Khumawala. 1989. MRP lot sizing with variable production/purchasing costs: Formulation and solution. *International Journal of Production Research* 27(6), 965–984.
- Qui, J. and R. Loulou. 1995. Multiproduct production/inventory control under random demands. *IEEE Transactions on Automatic Control* 40, 350–356.
- Rajagopalan, S. 1992. A note on an efficient zero-one formulation of the multilevel lot-sizing problem. *Decision Sciences* 23(4), 1023–1025.
- Ramudhin, A. and H. D. Ratliff. 1995. Generating daily production schedules in process industries. *IIE Transactions* 27, 646–656.
- Rosenblatt, M. and M. Kaspi. 1985. A dynamic programming algorithm for joint replenishment under general order cost functions. *Management Science* 31(3), 369–373.
- Rosenblatt, M. J. and N. Finger. 1983. An application of a grouping procedure to a multi-item production system. *International Journal of Production Research* 21(2), 223–229.
- Roundy, R. 1985. 98%-effective integer-ratio lot-sizing for one-warehouse multi-retailer systems. *Management Science* 31(11), 1416–1430.

- Roundy, R. 1986. A 98%-effective lot-sizing rule for a multi-product, multi-stage production/inventory system. *Mathematics of Operations Research* 11, 699–727.
- Roundy, R. 1989. Rounding off to powers of two in continuous relaxations of capacitated lot sizing problems. *Management Science* 35, 1433–1442.
- Russell, R. M. and L. J. Krajewski. 1992. Coordinated replenishments from a common supplier. *Decision Sciences* 23, 610–632.
- Sadrian, A. and Y. Yoon. 1992. Business volume discount: A new perspective on discount pricing strategy. *Journal of Purchasing and Materials Management* 28, 43–46.
- Sadrian, A. and Y. Yoon. 1994. A procurement decision support system in business volume discount environments. *Operations Research* 42, 14–23.
- Salomon, M. 1991. Deterministic lotsizing models for production planning. In *Lecture Notes in Economics and Mathematical Systems*, Volume 355. Berlin: Springer-Verlag.
- Sandbothe, R. A. 1991. The capacitated dynamic lot-sizing problem with startup and reservation costs: A forward algorithm solution. *Journal of Operations Management* 10(2), 255–266.
- Schneider, H. and D. Rinks. 1989. Optimal policy surfaces for a multi-item inventory problem. *European Journal of Operational Research* 39, 180–191.
- Schneider, M. J. 1982. Optimization of purchasing from a multi-item vendor. In *EURO V and the 25th International Conference of the Institute of Management Sciences*, Lausanne, Switzerland.
- Schultz, H. and S. G. Johansen. 1999. Can-order policies for coordinated inventory replenishment with Erlang distributed times between ordering. *European Journal of Operational Research* 113(1), 30–41.
- Schweitzer, P. J. and E. A. Silver. 1983. Mathematical pitfalls in the one machine multiproduct economic lot scheduling problem. *Operations Research* 31(2), 401–405.
- Shaoxiang, C. and M. Lambrecht. 1996. $X - Y$ band and modified (s, S) policy. *Operations Research* 44(6), 1013–1019.
- Silver, E. A. 1974. A control system for coordinated inventory replenishment. *International Journal of Production Research* 12(6), 647–671.
- Silver, E. A. 1975. Modifying the economic order quantity (EOQ) to handle coordinated replenishments of two or more items. *Production and Inventory Management Journal* 16(3), 26–38.
- Silver, E. A. 1981. Operations research in inventory management: A review and critique. *Operations Research* 29(4), 628–645.
- Silver, E. A. 1995. Dealing with a shelf life constraint in cyclic scheduling by adjusting both cycle time and production rate. *International Journal of Production Research* 33(3), 623–629.
- Sivazlian, B. D. and Y.-C. Wei. 1990. Approximation methods in the optimization of a stationary (s, S) inventory problem. *Operations Research Letters* 9(2), 105–113.
- Song, J.-S. 1998. On the order fill rate in a multi-item, base-stock inventory system. *Operations Research* 46(6), 831–845.
- Sox, C. R. and J. A. Muckstadt. 1995. Optimization-based planning for the stochastic lot scheduling problem. *IIE Transactions* 29(5), 349–358.
- Sox, C. R. and J. A. Muckstadt. 1996. Multi-item, multi-period production planning with uncertain demand. *IIE Transactions* 28(11), 891–900.
- Taylor, G. D., H. A. Taha, and K. M. Chowning. 1997. A heuristic model for the sequence-dependent lot scheduling problem. *Production Planning and Control* 8(3), 213–225.
- Taylor, S. and S. Bolander. 1986. Scheduling product families. *Production and Inventory Management Journal* 27(3), 47–56.
- Taylor, S. and S. Bolander. 1991. Process flow scheduling principles. *Production and Inventory Management Journal* 32(1), 67–71.
- Taylor, S. G. and S. F. Bolander. 1990. Process flow scheduling: Basic cases. *Production and Inventory Management Journal* 31(3), 1–4.
- Taylor, S. G. and S. F. Bolander. 1994. *Process Flow Scheduling*. Chicago: APICS.
- Taylor, S. G. and S. F. Bolander. 1996. Can process flow scheduling help you? *APICS—The Performance Advantage March* 6, 44–48.
- Tempelmeier, H. and M. Derstroff. 1996. A Lagrangian-based heuristic for dynamic multilevel multiitem constrained lotsizing with setup times. *Management Science* 42(5), 738–757.

- Tempelmeier, H. and S. Helber. 1994. A heuristic for dynamic multi-item multi-level capacitated lot sizing for general product structures. *European Journal of Operational Research* 75(2), 296–311.
- ter Haseborg, F. 1982. On the optimality of joint ordering policies in a multi-product dynamic lot size model with individual and joint set-up costs. *European Journal of Operational Research* 9, 47–55.
- Thizy, J. M. 1991. Analysis of Lagrangian decomposition for the multi-item capacitated lot-sizing problem. *INFOR* 29(4), 271–283.
- Thizy, J. M. and L. N. Van Wassenhove. 1985. Lagrangian relaxation for the multi-item capacitated lot-sizing problem: A heuristic implementation. *IIE Transactions* 17(4), 308–313.
- Thompson, R. M. and E. A. Silver. 1975. A coordinated inventory control system for compound Poisson demand and zero lead time. *International Journal of Production Research* 13(6), 581–602.
- Toklu, B. and J. M. Wilson. 1992. A heuristic for multi-level lot-sizing problems with a bottleneck. *International Journal of Production Research* 30(4), 787–798.
- Trigeiro, W. W., L. J. Thomas, and J. O. McClain. 1989. Capacitated lot sizing with setup times. *Management Science* 35(3), 353–366.
- Tyworth, J. E., J. L. Cavinato, and C. J. Langley. 1991. *Traffic Management: Planning, Operations, and Control*. Reading, MA: Addison-Wesley.
- Van der Duyn Schouten, F. A., M. J. G. van Eijs, and R. Heuts. 1994. The value of supplier information to improve management of a retailer's inventory. *Decision Sciences* 25(1), 1–14.
- Van Eijs, M. J. G. 1993. A note on the joint replenishment problem under constant demand. *Journal of the Operational Research Society* 44(2), 185–191.
- Van Eijs, M. J. G. 1994. Multi-item inventory systems with joint ordering and transportation decisions. *International Journal of Production Economics* 35, 285–292.
- Van Eijs, M. J. G., R. Heuts, and J. P. C. Kleijnen. 1992. Analysis and comparison of two strategies for multi-item inventory systems with joint replenishment costs. *European Journal of Operational Research* 59, 405–412.
- Van Nunen, J. A. E. E. and J. Wessels. 1978. Multi-item lot size determination and scheduling under capacity constraints. *European Journal of Operational Research* 2(1), 36–41.
- Van Wassenhove, L. N. and M. A. De Bodt. 1983. Capacitated lot sizing for injection molding: A case study. *Journal of the Operational Research Society* 34(6), 489–501.
- Van Wassenhove, L. N. and P. Vanderhenst. 1983. Planning production in a bottleneck department: A case study. *European Journal of Operational Research* 12, 127–137.
- Veatch, M. H. and L. M. Wein. 1996. Scheduling a make-to-stock queue: Index policies and hedging points. *Operations Research* 44(4), 634–647.
- Ventura, J. A. and C. M. Klein. 1988. A note on multi-item inventory systems with limited capacity. *Operations Research Letters* 7(2), 71–75.
- Viswanathan, S. and S. K. Goyal. 1997. Optimal cycle time and production rate in a family production context with shelf life considerations. *International Journal of Production Research* 35(6), 1703–1712.
- Wein, L. 1992. Dynamic scheduling of a multi-class make-to-stock queue. *Operations Research* 40, 724–735.
- Winands, E. M., I. J. Adan, and G. Van Houtum. 2011. The stochastic economic lot scheduling problem: A survey. *European Journal of Operational Research* 210(1), 1–9.
- Xu, J., L. L. Lu, and F. Glover. 2000. The deterministic multi-item dynamic lot size problem with joint business volume discount. *Annals of Operations Research* 96(1–4), 317–337.
- Zheng, Y. 1994. Optimal control policy for stochastic inventory systems with Markovian discount opportunities. *Operations Research* 42(4), 721–738.
- Zipkin, P. 1991a. Evaluation of base-stock policies in multiechelon inventory systems with compound-Poisson demands. *Naval Research Logistics* 38, 397–412.
- Zipkin, P. H. 1991b. Computing optimal lot sizes in the economic lot scheduling problem. *Operations Research* 39(1), 56–63.
- Zoller, K. 1990. Coordination of replenishment activities. *International Journal of Production Research* 28(6), 1123–1136.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Chapter 11

Multiechelon Inventory Management

In Chapters 4 through 9, we were primarily focused on single-item, single-location inventory systems. In Chapter 10, we extended the discussion to multiple items, and the issues associated with coordinating their ordering or production. In this chapter, we extend our treatment to another level of complexity—multiple locations. Specifically, we examine inventory control across the entire supply chain. As we will see, the solution approaches are somewhat more complex, but a student with solid understanding of the single-location models treated elsewhere in the book should be well prepared for this complexity. In this chapter, we focus on the multilocation setting where there is a single decision maker controlling inventory. We further investigate the important situation where decisions cross organizational boundaries in Chapter 12.

We begin in Section 11.1 by introducing supply chain management and the many issues involved with managing a supply chain. Section 11.2 discusses broader strategic issues that help to frame the discussion of the multiechelon models and insights we develop thereafter. In Section 11.3, we deal with the conceptually easiest situation of deterministic demand. Nevertheless, particularly when the demand rate varies with time, exact analysis quickly becomes quite complicated. Section 11.4 introduces the complexities of probabilistic demand in a multiechelon framework. We describe what is known as a *base stock control* system, a serial system, two arborescent (tree-structure) systems, and a “push” system. Next, in Section 11.5, for the first time, we deal with the so-called *product recovery situation*. Specifically, now, not all units demanded are consumed, but rather some units are returned, perhaps in a damaged state. As a result, both repair and purchase decisions must be made. Finally, we present additional insights in Section 11.6. As we have done throughout this book, in this chapter, we refer, sometimes very briefly, to research papers that apply to the relevant topics. These are intended for further insight and research, and can be safely ignored by many readers.

11.1 Multiechelon Inventory Management

Supply chain management (SCM) is the term used to describe the management of materials and information across the entire supply chain, from suppliers to component producers to final

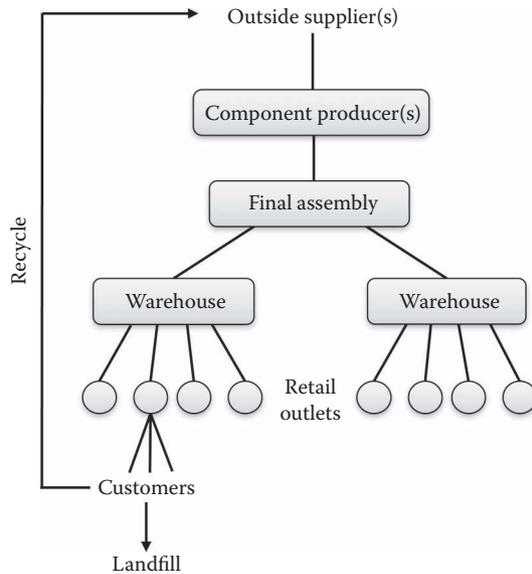


Figure 11.1 A schematic of a supply chain.

assemblers to distribution (warehouses and retailers), and ultimately to the consumer. In fact, it often includes after-sales service and returns or recycling. Figure 11.1 is a schematic of a supply chain. Supply chain management has generated much interest in recent years because of the realization that actions taken by one member of the chain can influence the performance of all others in the chain. Firms are increasingly thinking in terms of competing as part of a supply chain against other supply chains, rather than as a single firm against other individual firms. Also, as firms successfully streamline their own operations, the next opportunity for improvement is through better coordination with their suppliers and customers.

Before addressing opportunities for better coordination with suppliers and customers (in Chapter 12), we focus in this chapter on how a firm can manage inventory across multiple locations under their control. As suggested by Figure 11.1, it is often the case that a single firm makes inventory decisions at multiple locations such as their final assembly locations, warehouses, and retail outlets. Specifically, we will introduce models of multiechelon inventory systems that can be used to optimize the deployment of inventory in a supply chain and to evaluate a change in the supply chain “givens.” We will look at the case of an item being stocked at more than one location with resupply being made between at least some of the locations. A three-echelon illustration is shown in Figure 11.2. Retail outlets (the first echelon) are replenished from branch warehouses (the second echelon), which are supplied from a central warehouse (the third echelon). The latter, in turn, is replenished from outside sources. Inventory management in this system is complex because demand at the central warehouse is dependent on the demand (and stocking decisions) at the branches. And demand at the branches is dependent on the demand (and stocking decisions) at the retail outlets. More generally, we refer to this as a dependent demand situation, in contrast with earlier chapters of the book where demands for different stock-keeping units were considered to be independent.

Multistage manufacturing situations (raw materials, components, subassemblies, assemblies) are conceptually very similar to multiechelon inventory systems. However, in this chapter, we

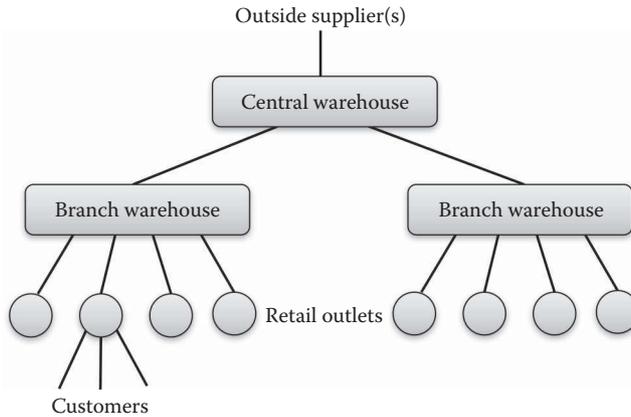


Figure 11.2 A multiechelon inventory situation.

restrict our attention to pure inventory (distribution) situations, by ignoring any finite replenishment capacities that are usually present in the production context. Nevertheless, some of the results of this chapter can be applied, with due caution, within production environments and, in fact, were developed with that type of application in mind. In addition, we will return in Chapter 15 to a detailed treatment of multistage manufacturing systems in which the concept of dependent demand will play a crucial role.

It seems intuitive that using the multiechelon inventory techniques we describe below provides significant benefit over using the methods of Chapters 6 and 7 in a multilevel setting. (See, e.g., Muckstadt and Thomas 1980.) Moreover, these multiechelon techniques have been successfully implemented. See, for example, Cohen et al. (1990) who apply multiechelon methods in a low demand, spare parts environment. Farasyn et al. (2011) describe the benefits of using inventory optimization tools, including multiechelon approaches at a large consumer goods firm. For very slow-moving items, from a single-level perspective, each retailer might hold, say, one unit, and yet sell one or fewer units per year. Because multiechelon models examine the entire system, they might recommend holding just a few parts at the warehouse level, and none at the retailers. System-wide savings can be enormous. With faster-moving parts, of course, it is likely that more units would be held at the retail level, but multiechelon methods help determine how many to hold at the retailers, and how many to hold at the warehouse.

11.2 Structure and Coordination

In this section, we provide a framework for supply chain management issues by briefly discussing structural and coordination decisions. Structural decisions are those that focus on longer-term, strategic issues and parallel the longer-term management levers described in Chapter 2. These decisions are typically made infrequently because they involve high costs to implement and change. Some structural decisions include

- Where to locate factories, warehouses, and retail sites?
- How many of these facilities to have?

- What capacity should each of these facilities have?
- When, and by how much, should capacity be expanded or contracted?
- Which facilities should produce and distribute which products?
- What modes of transportation should be used for which products, and under which circumstances?

The result of these structural decisions is a network of facilities designed to produce and distribute the products under consideration. Other parameters, such as lead times and some costs, used in our models are determined as well. The models we develop can be used to evaluate alternative structures. For example, customer service personnel often argue for having numerous stocking points close to customers. However, this leads to duplication of effort, extra carrying costs, and possibly extra transportation costs. The required inventory capacities and related costs of different options can be evaluated by the procedures of this chapter.*

Coordination decisions are usually taken after the structural decisions are made. Whereas structural decisions tend to be based on long-term, deterministic approximations, coordination decisions take the structure of the multiechelon network as given and focus on the short term. Therefore, they often must account for probabilistic demand and lead times. Coordination decisions include

- Should inventory stocking and replenishment decisions be made centrally or in a decentralized fashion?
- Should inventory be held at central warehouses or should these simply be used as break-bulk facilities?
- Where should inventory be deployed? In other words, should most inventory be held at a central location, or should it be pushed “forward” to the retail level?
- How should a limited and insufficient amount of stock be allocated to different locations that need it?

In addition, the four questions noted in Chapter 6 (how important is the item; periodic or continuous review; what form should the inventory policy take; and what specific cost and service goals should be set) fall in the coordination category as well. Clearly, the decisions are quite complex, and we cannot cover all of them in this chapter. However, we will present some models and insights that will help decision makers address some of the fundamental issues in supply chain management.

As a final comment, we note that costs associated with structural decisions primarily include fixed operating costs, but also some variable operating costs and certain transportation costs. Costs associated with coordination decisions include variable operating costs, inventory costs (holding and shortage), emergency transportation costs, and replenishment costs. General reviews by Clark (1972); Goyal and Gunasekaran (1990); Axsäter (1993a); Federgruen (1993); Diks et al. (1996); Thomas and Griffin (1996); and Gunasekaran et al. (2004) should be helpful for further information.

* Research relevant to structural decisions includes Arntzen et al. (1995); Cohen et al. (1995); Graves and Willems (2003b); Meixell and Gargeya (2005); and Graves and Willems (2005).

11.3 Deterministic Demand

In this section, we deal with the relatively simple situation in which the external demand rates are known with certainty. This is admittedly an idealization, but it is important to study for several reasons. First, the models will reveal the basic interactions among replenishment quantities at the different echelons. Second, as in earlier chapters, the deterministic case can serve as a basis for establishing replenishment quantities for the probabilistic case. Finally, the production environment can occasionally be approximated by deterministic demand and production rates. We also assume that replenishment lead times are known and fixed.

11.3.1 Sequential Stocking Points with Level Demand

Consider the simplest of multiechelon situations, namely, where the stocking points are serially connected—for example, one central warehouse, one retailer warehouse, and one retail outlet. Another interesting production interpretation is depicted in Figure 11.3. An item progresses through a series of operations, none of which are of an assembly or splitting nature; that is, each operation has only one predecessor and only one successor operation. For exposition purposes, we restrict ourselves to the most simple case of but two stages, denoted by a warehouse (*W*) and a retailer (*R*). Let us introduce some preliminary notation:

- D = deterministic, constant demand rate at the retailer, in units/unit time (normally 1 year)
- A_W = fixed (setup) cost associated with a replenishment at the warehouse, in dollars
- A_R = fixed (setup) cost associated with a replenishment at the retailer, in dollars
- v_W = unit variable cost or value of the item at the warehouse, in \$/unit
- v_R = unit variable cost or value of the item at the retailer, in \$/unit
- r = carrying charge, in \$/\$/unit time
- Q_W = replenishment quantity at the warehouse, in units
- Q_R = replenishment quantity at the retailer, in units

The two controllable (or decision) variables are the replenishment sizes Q_W and Q_R . Figure 11.4 shows the behavior of the two levels of inventory with the passage of time for the particular case where $Q_W = 3Q_R$. A little reflection shows that at least for the case of deterministic demand it never would make sense to have Q_W anything but an integer multiple of Q_R . Therefore, we can think of two alternative decision variables Q_R and n where

$$Q_W = nQ_R \quad n = 1, 2, \dots \tag{11.1}$$

Note that from Figure 11.4 the inventory at the warehouse does not follow the usual sawtooth pattern even though the end usage is deterministic and constant with time. The reason is that the

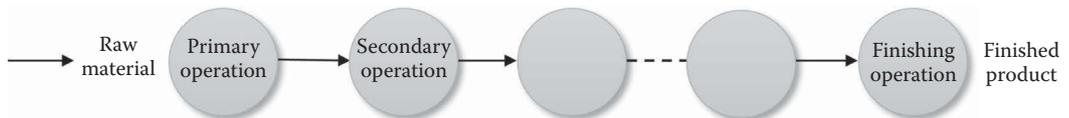


Figure 11.3 A serial production process.

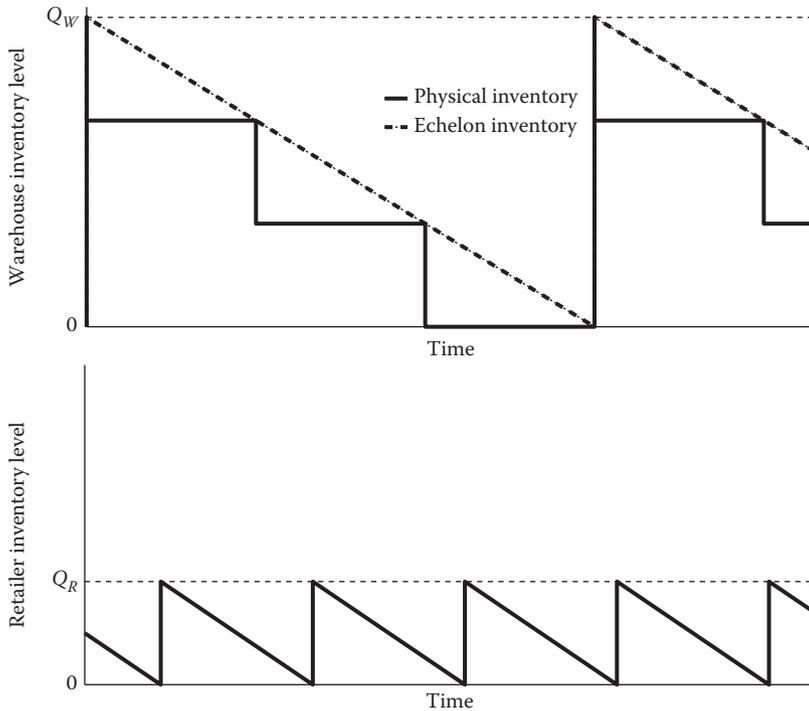


Figure 11.4 Behavior of the inventory levels in a deterministic two-stage process.

withdrawals from the warehouse inventory are of size Q_R . We could carry out an analysis using the conventional definitions of inventories, but the determination of average inventory levels becomes complicated. Instead, it is easier to use a different concept, known as *echelon stock*, introduced by Clark and Scarf (1960). They define the echelon stock of echelon j (in a general multiechelon system) as the number of units in the system that are at, or have passed through, echelon j but have as yet not been specifically committed to outside customers. When backorders are permitted, the echelon stock can be negative. With this definition and uniform end-item demand, each echelon stock has a sawtooth pattern with time, as illustrated in Figure 11.4; thus, it is simple to compute the average value of an echelon stock. However, we cannot simply multiply each average echelon stock by the standard v_j term and sum to obtain total inventory carrying costs. The reason is that the same physical units of stock can appear in more than one echelon inventory. For example, in our two-stage process, the actual finished inventory is counted in both the finished echelon inventory and the warehouse echelon inventory. The way around this dilemma is to value any specific echelon inventory at *only the value added* at that particular echelon.* When the decision

* In the deterministic demand case, echelon stock policies dominate installation stock policies, see Axsäter and Juntti (1996, 1997). (Installation stocks are simply the stocks at the given location, without regard for downstream stocks.) But in certain cases with probabilistic demand, the opposite might be true, particularly if the lead time to the warehouse is short relative to the lead time to the retailers. Moreover, when there is more than one retail outlet and transshipments are not possible, basing a replenishment action at the wholesale level on its echelon stock position is not likely to be optimal when the inventories at the branches are markedly unbalanced.

being made is whether to store inventory at an upstream location or at a downstream location that it supplies, the relevant holding cost is the incremental cost of moving the product to the retailer. This incremental cost is exactly the echelon holding cost. (Problem 11.2 is concerned with showing that this is equivalent to using the usual definitions of inventory levels and stock valuations.) Thus, in our two-stage example, the warehouse echelon inventory is valued at $v'_W = v_W$, while the retailer echelon inventory is valued at only $v'_R = v_R - v_W$. More generally, in a production assembly context, the echelon valuation v' at a particular stage i is given by

$$v'_i = v_i - \sum v_j \tag{11.2}$$

where the summation is over all immediate predecessors, j .

Returning to our two-stage serial situation, the total relevant (setup plus carrying) costs per unit time are given by

$$\text{TRC}(Q_W, Q_R) = \frac{A_W D}{Q_W} + \bar{I}'_W v'_W r + \frac{A_R D}{Q_R} + \bar{I}'_R v'_R r \tag{11.3}$$

where

- \bar{I}'_W = average value of the warehouse echelon inventory, in units
- \bar{I}'_R = average value of the retailer echelon inventory, in units

Substituting from Equation 11.1 and noting that the echelon stocks follow sawtooth patterns,

$$\begin{aligned} \text{TRC}(n, Q_R) &= \frac{A_W D}{n Q_R} + n \frac{Q'_R v'_W r}{2} + \frac{A_R D}{Q_R} + \frac{Q'_R v'_R r}{2} \\ &= \frac{D}{Q_R} \left(A_R + \frac{A_W}{n} \right) + \frac{Q_R r}{2} (n v'_W + v'_R) \end{aligned} \tag{11.4}$$

We must find the values of n (an integer) and Q_R that minimize this expression. As proved in the Appendix of this chapter, this is achieved by the following procedure:

Step 1 Compute

$$n^* = \sqrt{\frac{A_W v'_R}{A_R v'_W}} \tag{11.5}$$

If n^* is exactly an integer, go to Step 4 with $n = n^*$. Also, if $n^* < 1$, go to Step 4 with $n = 1$.

Otherwise, proceed to Step 2.

Step 2 Ascertain the two integer values, n_1 and n_2 , that surround n^* .

Step 3 Evaluate

$$F(n_1) = \left[A_R + \frac{A_W}{n_1} \right] [n_1 v'_W + v'_R]$$

and

$$F(n_2) = \left[A_R + \frac{A_W}{n_2} \right] [n_2 v'_W + v'_R] \tag{11.6}$$

If $F(n_1) \leq F(n_2)$, use $n = n_1$.

If $F(n_1) > F(n_2)$, use $n = n_2$.

Step 4 Evaluate

$$Q_R = \sqrt{\frac{2[A_R + A_W/n]D}{[nv'_W + v'_R]r}} \quad (11.7)$$

Step 5 Calculate

$$Q_W = nQ_R$$

11.3.1.1 Numerical Illustration of the Logic

Let us consider a particular liquid product, which a firm buys in bulk, then breaks down and repackages. So, in this case, the warehouse corresponds to the inventory prior to the repackaging operation, and the retailer corresponds to the inventory after the repackaging operation. The demand for this item can be assumed to be essentially deterministic and level at a rate of 1,000 liters per year. The unit value of the bulk material (v'_W or v_W) is \$1/liter, while the value added by the transforming (break and package) operation $v'_R = (v_R - v_W)$ is \$4/liter. The fixed component of the purchase charge (A_W) is \$10, while the setup cost for the break and repackage operation (A_R) is \$15. Finally, the estimated carrying charge is 0.24 \$/\$/year.

Step 1

$$\begin{aligned} n^* &= \sqrt{\frac{(10)(4)}{(15)(1)}} \\ &= 1.63 \end{aligned}$$

Step 2

$$\begin{aligned} n_1 &= 1 \\ n_2 &= 2 \end{aligned}$$

Step 3

$$\begin{aligned} F(1) &= \left[15 + \frac{10}{1}\right][1 + 4] = 125 \\ F(2) &= \left[15 + \frac{10}{2}\right][(2)(1) + 4] = 120 \end{aligned}$$

that is, $F(1) > F(2)$. Thus, use $n = 2$.

Step 4

$$Q_R = \sqrt{\frac{2[15 + 10/2]1,000}{[(2)(1) + 4]0.24}} \approx 167 \text{ liters}$$

Step 5

$$Q_W = (2)(167) = 334 \text{ liters}$$

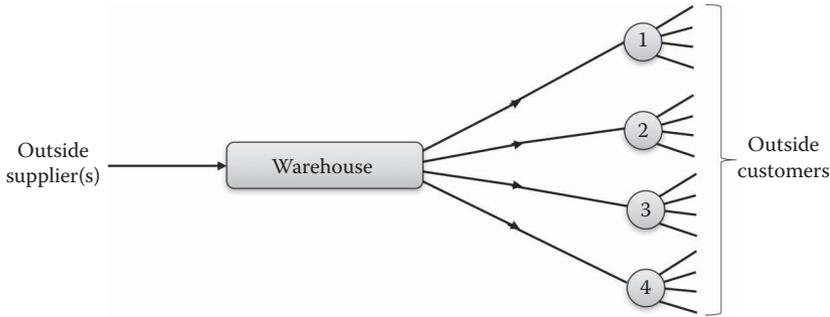


Figure 11.5 A one-warehouse, four-retailer system.

In other words, we purchase 334 liters* at a time; one-half of these or 167 liters are immediately broken and repackaged. When these 167 (finished) liters are depleted, a second break and repackaging run of 167 liters is made. When these are depleted, we start a new cycle by again purchasing 334 liters of raw material.

11.3.1.2 Additional Research

See Roundy (1985) for research directly related to this problem. In other related research, Rosenblatt and Lee (1985) study a supplier with a single customer having level demand. The customer uses the EOQ for placing orders, and the supplier provides a discount of the form $a - bQ$, where Q is the order quantity of the customer. The supplier's order quantity is kQ , where k is an integer. Their model finds b and k so as to maximize the supplier's profit. See also Lee and Rosenblatt (1986); Moily (1986); Chakravarty and Martin (1988); Goyal and Gupta (1989); Dobson and Yano (1994); Kohli and Park (1994); Atkins and Sun (1995); Weng (1995); Aderohunmu et al. (1995); Desai (1996); Gurnani (1996); Lam and Wong (1996); and Chen and Zheng (1997b).

11.3.2 Other Results for the Case of Level Demand

For the serial situation described in the previous subsection, Szendrovits (1975) proposes a different type of control procedure—namely, where the same lot size is used at all operations but subbatches can be moved between operations. This allows an overlap between operations, and therefore reduces the manufacturing cycle time. He shows that, under certain conditions, this type of control procedure can outperform that proposed in the preceding subsection.

A number of authors have looked at systems more general in nature than the serial. Schwarz (1973) deals with a one-warehouse, n -retailer situation (depicted in Figure 11.5 for $n = 4$). For $n \geq 3$, he shows that the form of the optimal policy can be very complex; in particular, it requires that the order quantity at one or more of the locations vary with time even though all relevant demand and cost factors are time-invariant. He rightfully argues for restricting attention to a simpler class of strategies (where each location's order quantity does not change with time) and develops an effective heuristic for finding quite good solutions.

* Of course, it may make sense to round this quantity to a convenient shipment size such as 350 liters. In such a case, Q_R would also be adjusted (to 175 liters).

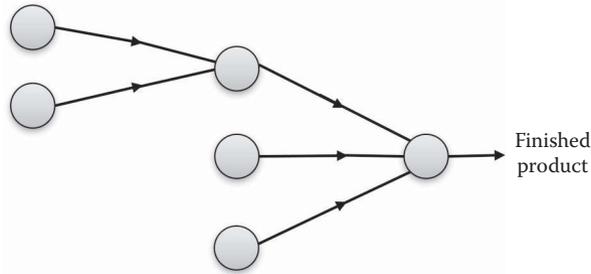


Figure 11.6 A “pure” assembly system.

Schwarz and Schrage (1975) adopt essentially the same approach for another type of deterministic situation—namely, “pure” assembly* (each node feeds into, at most, one other node as portrayed in Figure 11.6). These authors make use of a “myopic” strategy where each node and its successor are treated in isolation by much the same procedure as for the two-stage serial case discussed in the preceding section. Graves and Schwarz (1977) perform a similar analysis for arborescent systems, which are just the opposite of a “pure” assembly system—now each node has, at most, one node feeding into it.

Williams (1981, 1983) treats the general situation of an assembly network feeding into an arborescent system (assembly production followed by distribution). Maxwell and Muckstadt (1985) address the same problem but impose several restrictions (based on practical considerations) on the nature of the policy to be used, specifically:

1. The policy is nested—that is, a replenishment cannot occur at an operation unless one also occurs at immediate successor operations.
2. The policy is stationary—that is, the time between replenishments at each stage is a constant.
3. A base planning period (e.g., a shift, day, week) exists, and all reorder intervals must be an integer multiple of this base period.
4. The time between replenishments at any particular stage must be 1, 2, 4, 8, . . . times the interval between replenishments at any immediately following stage. This is the “powers of 2” restriction encountered in Section 10.4.2 of Chapter 10.

The resulting mathematical formulation results in a large, nonlinear, integer programming problem, but the authors suggest an efficient, specialized solution procedure. See also Muckstadt and Roundy (1987) and Iyogun (1992).

11.3.3 Multiechelon Stocking Points with Time-Varying Demand

Here, we return to the demand situation assumed in Chapter 5 where end-item usage is known *but varies from period to period* [$D(j)$ in period j]. Also, carrying costs are incurred only on period-ending inventories. However, now we have a multiechelon assembly or distribution system. Our

* The term *assembly* implies a production context. However, the procedures discussed here ignore any capacity constraints.

discussion is based on the work of Blackburn and Millen (1982) specifically for an assembly structure. Moreover, for illustrative purposes, we concentrate on the simple two-stage serial process used in Section 11.3.1. (The more general case can be found in the Blackburn and Millen reference.)

One approach would be simply to use one of the procedures of Chapter 5 sequentially echelon by echelon. For example, the Silver–Meal heuristic could be utilized to schedule replenishments for the retailer (or the finished item in a production context). This would imply a pattern of requirements for the warehouse (or the primary processing stage in production), which would then be used as input to the Silver–Meal heuristic to plan the replenishments of the warehouse.* Although this approach is simple, it ignores the cost interdependency of the two echelons. Specifically, in choosing the replenishment strategy at the retailer the method does not take account of the cost implications at the warehouse.

Recognizing the above deficiency, we wish to develop a procedure that can still be applied sequentially (to keep implementation effort at a reasonable level) but that captures the essence of the cost interdependencies. An examination of the level demand case provides considerable insight. In particular, we repeat Equation 11.4 here:

$$\text{TRC}(Q_W, Q_R) = \frac{D}{Q_R} \left(A_R + \frac{A_W}{n} \right) + \frac{Q_R r}{2} (n v'_W + v'_R) \tag{11.8}$$

where

$$Q_W = n Q_R \tag{11.9}$$

Equation 11.8 is analogous to a *single* echelon problem (the selection of Q_R) if the adjusted fixed cost of a replenishment is

$$\hat{A}_R = A_R + \frac{A_W}{n} \tag{11.10}$$

and the adjusted unit variable cost of the item is

$$\hat{v}_R = n v'_W + v'_R \tag{11.11}$$

Note that the term A_W/n reflects that there is a warehouse setup at only every n th retailer setup. We can select Q_R (see Equation 11.7), including properly taking account of the cost impact at the warehouse, if we have a good preestimate of n . We could use Steps 1–3 of the algorithm in Section 11.3.1, but Blackburn and Millen have found that simply using Equation 11.5 and ensuring that n is at least unity works well (particularly in more complex assembly structures).

Thus we have

$$n = \max \left[\sqrt{\frac{A_W v'_R}{A_R v'_W}}, 1 \right] \tag{11.12}$$

Returning to the time-varying demand situation, we compute n using Equation 11.12; then, employ the n value to obtain adjusted setup and unit variable costs for the retailer according to Equations 11.10 and 11.11. The Silver–Meal heuristic can then be applied to the retailer using \hat{A}_R

* Because of batching at the retailer, the resulting requirements pattern at the warehouse level is likely to contain a number of periods with no demand. As mentioned in Section 5.6.7 of Chapter 5, the unmodified Silver–Meal heuristic should be used with caution under such circumstances.

and \hat{v}_R . The resulting replenishments again imply a requirements pattern for the warehouse. Subsequently, the Silver–Meal, or another, lot-sizing procedure is used at the warehouse with A_W and v_W . We next show a numerical illustration. Problem 11.4 also presents a full numerical example.

11.3.3.1 Numerical Illustration

To illustrate the computation of \hat{A}_R and \hat{v}_R , consider a two-stage process where $A_W = \$30$, $A_r = \$20$, $v_W = \$2/\text{unit}$, and $v_r = \$7/\text{unit}$. First, from Equation 11.2, we have

$$\begin{aligned}v'_R &= v_R - v_W = \$5/\text{unit} \\v'_W &= v_W = \$2/\text{unit}\end{aligned}$$

Then, using Equation 11.12, we obtain

$$\begin{aligned}n &= \max \left[\sqrt{\frac{(30)(5)}{(20)(2)}}, 1 \right] \\&= \max [1.94, 1] \\n &= 1.94\end{aligned}$$

Finally, Equations 11.10 and 11.11 give us

$$\hat{A}_R = 20 + \frac{30}{1.94} = \$35.46$$

and

$$\hat{v}_R = 1.94(2.00) + 5.00 = \$8.88/\text{unit}$$

For related research in this area, see Zangwill (1969); Love (1972); Crowston et al. (1972, 1973a,b); Crowston and Wagner (1973); Lambrecht et al. (1981); Graves (1981); Afentakis et al. (1984); Chakravarty (1984); De Bodt et al. (1984); Blackburn and Millen (1985); Axsäter (1986); Park and Kim (1987); Afentakis (1987); Jackson et al. (1988); Joneja (1990); Rajagopalan (1992); El-Najdawi and Kleindorfer (1993); Iyogun and Atkins (1993); Lu and Posner (1994); Billington et al. (1994); Clark and Armentano (1995); and Jans and Degraeve (2007). See Simpson and Selcuk Erenguc (1996); Erengüç et al. (1999); Jans and Degraeve (2008); Buschkühl et al. (2010), and Clark et al. (2011) for reviews and discussions of this literature. Also see the references in Section 10.6.2 of Chapter 10.

11.4 Probabilistic Demand

Probabilistic demand raises several new issues and creates extreme modeling complexities in a multiechelon inventory situation. For discussion purposes, we examine the relatively simple three-echelon distribution system depicted in Figure 11.7.

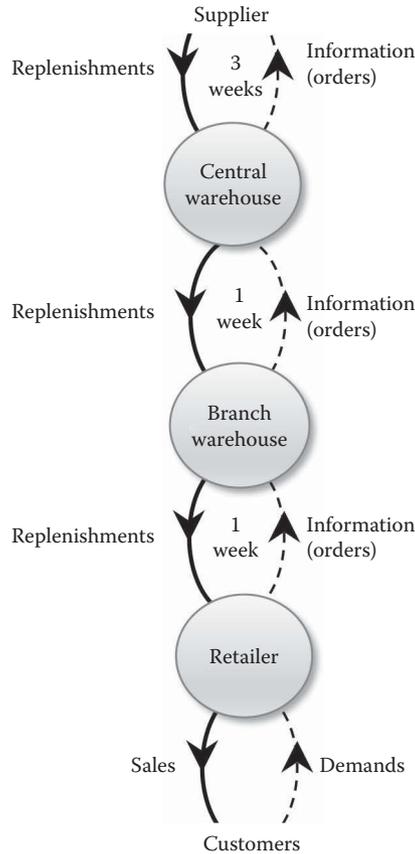


Figure 11.7 A multiechelon situation with single-stage information flow.

The replenishment lead times between adjacent levels (in each case assuming that the feeding level has adequate stock) are as follows:

- Branch to retailer—1 week
- Central to branch—1 week
- Supplier to central—3 weeks

Straightforward use of the methods of Chapters 4 through 6 would dictate that each stocking level (retailer, branch warehouse, and central warehouse) would independently make replenishment decisions based on its own:

1. Cost factors and service considerations
2. Predicted demand—presumably forecasts based on historical demand that it has observed from the next stocking point downstream
3. Replenishment lead time from the next stocking point upstream

Such a system has three serious flaws. First, the lead time observed at, say, the retailer is dependent on whether the branch warehouse has sufficient stock to fill the order. If stock is available, the

lead time is simply the transportation time plus any order processing time. If the branch is out of stock, the lead time could be as long as the branch's lead time plus the transportation time to the retailer. Second, as discussed in Section 11.3.1, it ignores the cost implications at one echelon of using certain ordering logic at another level. Third, even if end-customer demand is fairly smooth, the orders placed farther up the line become progressively larger and less frequent.

Other complicating factors include

1. How do we define service in a multiechelon situation? Normally, service is measured only at the lowest echelon. In a multiechelon system, a stockout at one of the higher echelons has only a secondary effect on service: it may lengthen the lead time for a lower echelon, which ultimately may cause customer disservice. We can see that safety stock at a particular echelon has effects on stockouts at *other* echelons. Thus, we have to be careful to avoid unnecessary duplication of safety stocks.
2. Suppose that the branch warehouse places orders of size Q_b on the central warehouse. What happens when such an order is placed and the on-hand stock at the central location is less than Q_b ? Is a partial shipment made or does the system wait until the entire order can be shipped?
3. What about the possibility of an emergency shipment directly from the central warehouse to the retailer?
4. In more complicated multiechelon structures, transshipments between points at the same echelon may be possible, for example, between retailers in Figure 11.5.
5. Again, for more complicated structures when one central facility supplies several different stocking points at the next echelon, the central facility is likely to adopt a rationing policy when it faces multiple requests with insufficient stock to meet them all. Section 11.4.3 will specifically address this issue.

Researchers have developed several fundamental approaches to dealing with these complexities. One is to approximate the complex problem by simpler, and more easily solved versions. The METRIC model and extensions, to be discussed in Section 11.5.1, are examples. A second approach is to restrict the inventory policy to, say, the nested procedure discussed in Section 11.3.1. The model discussed in Section 11.4.2 employs this approach. In either case, however, the basic problem of establishing the linkages among echelons is critical. Three fundamental approaches have been developed. The first is most widely used and finds or approximates the distribution of the lead time observed by a downstream facility. See Section 11.4.2 for example. The second finds, or approximates, the distribution of the total number of orders outstanding at all retailers. These are then disaggregated to find the approximate number of orders outstanding at each retailer. The model described in Section 11.5.1 follows this approach. The third tracks each demand unit and matches it with a supply unit, calculating costs from the information obtained. Svoronos and Zipkin (1988); Axsäter (1993b); and Chen and Zheng (1997a) are papers that use this approach.

Despite all of the above complexities, Muckstadt and Thomas (1980) argue that it is indeed still worthwhile to use multiechelon control procedures. These authors show substantial benefits in a military supply context, and Lawrence (1977) does the same for a commercial distribution system. Arguing for the other side, Hausman and Erkip (1994) demonstrate that appropriately modified single echelon models lead to small cost penalties for a data set of low-volume, high-cost items. It is clear, however, that in many instances multiechelon techniques are valuable and can lead to large improvements in cost and service.

Two dimensions will be useful as we proceed: local versus global information, and centralized versus decentralized control. Local information implies that each location sees demand only in the form of orders that arrive from the locations it directly supplies. Also, it has visibility of only its own inventory status, costs, and so on. Global information implies that the decision maker has visibility of the demand, costs, and inventory status of all the locations in the system. Centralized control implies that attempts are made to jointly optimize the entire system, usually by the decisions of one individual or group. Centralized control is often identified with “push” systems, because a central decision maker pushes stock to the locations that need it most. Decentralized control implies that decisions are made independently by separate locations. Decentralized control is often identified with “pull” systems because independent decision makers pull stock from their suppliers. (See Pyke and Cohen 1990.)

The best solutions are obtained by using global information and centralized control because the decisions are made with visibility to the entire system using information for all locations. Unexpectedly large replenishment orders from downstream are eliminated. However, these solutions require cooperation and coordination across multiple parties within operations, across functions, and, in some cases, across firms. We address coordination across firms in Chapter 12.

If a firm is faced with decentralized control and local information, one good approach is for each location to use the methods of Chapters 4 through 7, probably with the probabilistic lead time formulas of Section 6.10 of Chapter 6. Benefits can be gained from sharing information even if control remains decentralized. Section 11.4.1 discusses a system that assumes decentralized control and global information.

11.4.1 Base Stock Control System

The base stock system is a response to the difficulties of each echelon deciding when to reorder based only on demand from the next lower echelon. The key change is to make end-item demand information available for decision making at all stocking points, as illustrated in Figure 11.8. This requires the use of an effective communication system that provides timely and accurate information; it also requires a high level of trust across the supply chain so that firms are willing to share potentially sensitive information. Each stocking point makes replenishments based on actual end-item customer demands rather than on replenishment orders from the next level downstream. With this modification, the procedures of Chapters 4 through 6 are now more appropriate to use.

The most common type of base stock system, in the research literature as well as in practice, is one in which an order-up-to policy is used. Order-up-to policy systems were introduced in Chapter 6 where decisions for any particular stocking point were based only on its stock position and its direct demand process. Here, we present a more general (s, S) system, more appropriate in a multiechelon situation. For each stocking point treated independently, an order quantity, Q , is established (using end-item demand forecasts) by one of the methods of Chapters 4 and 5. Next, a reorder point s is established by one of the procedures of Chapter 6, using end-item demand forecasts over the replenishment lead time appropriate to the echelon under consideration. (For example, the central warehouse would use a lead time of 3 weeks.) Then, the order-up-to-level S , also called the base stock level, is determined through the relation

$$S = s + Q \quad (11.13)$$

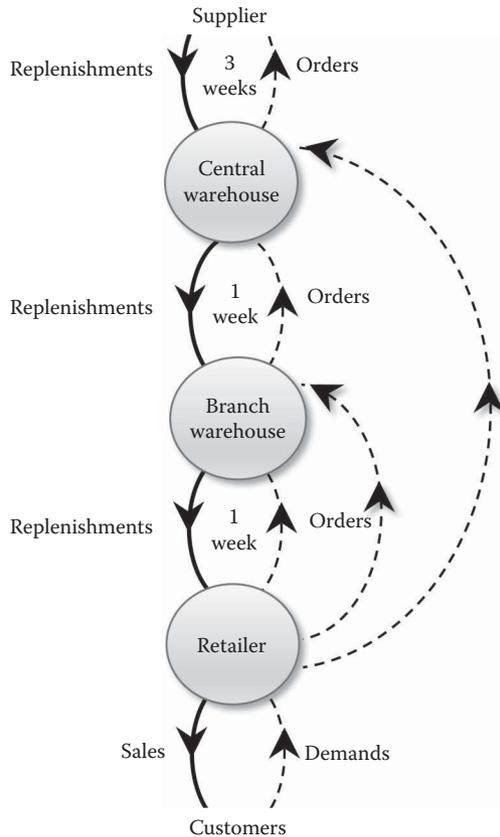


Figure 11.8 Information and stock flow in a base stock system.

In terms of physical operation, the echelon inventory position at each level is monitored according to the following relation:

$$\text{Echelon inventory position} = (\text{Echelon stock}) + (\text{On order}) \quad (11.14)$$

where, as defined earlier, the echelon stock of echelon j is the number of units in the system that are at, or have passed through, echelon j but have as yet not been specifically committed to outside customers.

The “on-order” term in Equation 11.14 refers to an order placed by echelon j on the next *higher* echelon. To illustrate, suppose that at a point in time the physical stocks of our example (Figure 11.8) are:

Branch warehouse—50 units
Retail outlet—20 units

Furthermore, suppose that 5 units of known customer demand have not yet been satisfied. Moreover, assume that there is no order outstanding from the branch on the central warehouse and

that 10 units are in transit between the branch and the retail outlet. Then, the echelon inventory position for the branch warehouse level would be

$$\text{Inventory position} = (50 + 10 + 20 - 5) + (0) = 75 \text{ units}$$

The echelon inventory position is reviewed after each transaction or on a periodic basis. (In the latter case, the replenishment lead time for safety stock calculations at each level must be increased by the review interval.) Once the echelon inventory position is known, it is compared with the reorder point s . Whenever it is at or lower than s , enough is ordered from the preceding echelon to raise the position to the base stock level S .

To repeat, ordering decisions at any stocking point in the base stock system are made as a result of end-item demand, not as a result of orders from the next level downstream. There is much less variability in the former than in the latter; hence, significantly lower safety stocks (reduced carrying costs) are achieved by the base stock system.

11.4.2 Serial Situation

For the special case of a serial system (such as in Figure 11.3), De Bodt and Graves (1985) propose a procedure for computing the replenishment quantities and reorder points that explicitly takes account of the effects of the stock at one echelon on the lead time at the next lower echelon. Their model assumes that demand information is available to all locations (global information), and that decisions are made centrally (centralized control). Again, for illustrative purposes and consistency with Section 11.3.1, we restrict our description to the case of two stages and call them *warehouse* and *retailer*. The assumptions underlying the decision rules include

1. External demand occurs only at the retailer and is a stationary process. Conceptually, the method can certainly be applied to a process that changes slowly with time where we have estimates of the mean and standard deviation of demand over suitable durations of time. We assume normally distributed forecast errors.
2. There is a deterministic replenishment lead time associated with each stage (L_W and L_R). Furthermore, the lead time L_R only begins when there is sufficient warehouse stock available to fulfill a retailer replenishment.
3. The policy used is of the (s, Q) form, that is, continuous review with four parameters:
 - s_W = reorder point (based on the echelon inventory position) at the warehouse
 - Q_W = order quantity at the warehouse
 - s_R = reorder point at the retailer
 - Q_R = order quantity at the retailer

Furthermore, for our purposes, we assume that Q_R and $Q_W = nQ_R$ have been predetermined by the procedure of Section 11.3.1. (De Bodt and Graves suggest an iterative scheme to simultaneously find the Q 's and s 's, similar to what was shown in Section 7.3 of Chapter 7.) Note that the restriction, $Q_W = nQ_R$, implies that the policies are nested: when the warehouse orders, the retailer also orders.

The decision rules for choosing s_R and s_W are:

1. Select s_R such that

$$s_R = \hat{x}_{L_R} + k_R \sigma_{L_R} \tag{11.15}$$

where

- \hat{x}_{L_R} = expected (forecast) demand over a retailer lead time
- σ_{L_R} = standard deviation of forecast errors over the same interval
- k_W = retailer safety factor, which satisfies

$$p_{u \geq}(k_R) = \frac{Q_R(v_R - v_W)r}{B_2 v_R D} \tag{11.16}$$

with D being the annual demand rate for the item, r the carrying charge in \$/\$/year, and B_2 the fraction of unit value charged per unit short. Note the similarity to Equation 6.22 of Chapter 6.

2. Select s_W such that

$$s_W = \hat{x}_{L_W+L_R} + k_W \sigma_{L_W+L_R} \tag{11.17}$$

where

- $\hat{x}_{L_W+L_R}$ = expected demand over a warehouse lead time, plus a retailer lead time
- $\sigma_{L_W+L_R}$ = standard deviation of forecast errors over the same interval
- k_W = warehouse safety factor, which satisfies

$$p_{u \geq}(k_W) = \frac{Q_R[v_R + (n - 1)v_W]r}{B_2 v_R D} \tag{11.18}$$

The v expression in the numerator of Equation 11.16 is equivalent with an echelon valuation. The $L_W + L_R$ combination occurs in Equation 11.17 because on every n th replenishment at the retailer, we have an associated warehouse replenishment, and the latter should be initiated when there is sufficient echelon stock to protect against shortages over a period of length $L_W + L_R$ (recall assumption 2). The derivation of Equation 11.18 (see De Bodt and Graves 1985) uses echelon stocks and hinges on distinguishing this special replenishment from the other $n - 1$ retailer replenishments. Note that the right-hand side of Equation 11.16 is always smaller than that of Equation 11.18. Hence, the safety factor at the retailer is always larger than the safety factor at the warehouse level.

11.4.2.1 Numerical Illustration

We use the same example as in Section 11.3.1, the liquid product that the firm buys in bulk (the warehouse inventory), then breaks down and repackages (the retailer inventory). Suppose that the purchasing lead time is 6 weeks and the lead time for the *retailer*—that is, for repackaging—is 1 week. Also, assume that forecast errors are normally distributed with the σ for 1 week being 6.4 units, and that $\sigma_t = \sqrt{t}\sigma_1$. From Section 12.3.1, we know that the annual demand rate is 1,000 liters, $v_W = \$1/\text{liter}$, $v_r = \$5/\text{liter}$, $r = 0.24$ \$/\$/year, $n = 2$, and $Q_R = 167$ liters. Finally, suppose that the fractional charge per unit short, B_2 , is set at 0.35. From Equation 11.16, we obtain

$$\begin{aligned} p_{u \geq}(k_R) &= \frac{(167)(5.00 - 1.00)(0.24)}{(0.35)(5.00)(1,000)} \\ &= 0.0916 \end{aligned}$$

Then, Table II.1 in Appendix II gives

$$k_R = 1.33$$