INTERNATIONAL SERIES IN OPERATIONS RESEARCH AND MANAGEMENT SCIENCE

Postponement Strategies in Supply Chain Management

T. C. Edwin Cheng
Jian Li
C. L. Johnny Wan
Shouyang Wang

Springer

International Series in Operations Research & Management Science

Volume 143

Series Editor:
Frederick S. Hillier
Stanford University

For further volumes:
http://www.springer.com/series/6161

T.C. Edwin Cheng · Jian Li · C.L. Johnny Wan ·
Shouyang Wang

Postponement Strategies in Supply Chain Management

Springer

T.C. Edwin Cheng
Department of Logistics & Maritime Studies
The Hong Kong Polytechnic University
11 Yuk Choi Road
Hung Hom, Kowloon
Hong Kong SAR
lgtcheng@polyu.edu.hk

C.L. Johnny Wan
Department of Logistics & Maritime Studies
The Hong Kong Polytechnic University
11 Yuk Choi Road
Hung Hom, Kowloon
Hong Kong SAR
lgtwan@polyu.edu.hk

Jian Li
School of Economics & Management
Beijing University of Chemical Technology (BUCT)
15 East Road of North Third Ring Road
100029 Beijing
China, People's Republic
lijian@mail.buct.edu.cn

Shouyang Wang
Chinese Academy of Sciences
Academy of Mathematics & Systems Science
55 Zhongguancun East Road
100190 Beijing
China, People's Republic
sywang@amss.ac.cn

ISSN 0884-8289
ISBN 978-1-4419-5836-5 e-ISBN 978-1-4419-5837-2
DOI 10.1007/978-1-4419-5837-2
Springer New York Dordrecht Heidelberg London

Library of Congress Control Number: 2010921157

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Abstract

Postponement is a supply chain strategy that enables a supply chain to achieve both low cost and fast response by combining some common processes and delaying other product differentiation processes such as packaging and labeling. The point that separates the differentiation processes from the common processes is known as the point of product differentiation. Recent research studies have identified four common postponement strategies, namely pull, logistics, form and price postponements. They aim at balancing the costs and benefits of mass production and customization. In this book, four types of model are presented to evaluate the impacts of pull and form postponement strategies under various supply chain structures.

First, we develop two EOQ-based models to examine the impact of pull postponement. Then we develop some EPQ-based models to examine the impact of postponement. The third type of model is a stochastic model of a single end-product supply chain that consists of a supplier, a manufacturer and a number of customers. In the last type of model, we aim at conducting a simulation experiment of a two-end-product supply chain, for which customer demands are discrete and independent. Besides mathematical models, two case studies from industry are presented to support our theoretical results.

Keywords Supply chain management, Postponement strategy

Preface

Postponement strategy is one of the major supply chain management (SCM) practices that has a discernible impact on firms' competitive advantage and organizational performance. Postponement is a mass customization strategy that captures the advantages of both mass production and mass customization. Recent research studies have identified four common postponement strategies, namely *pull*, *logistics*, *form* and *price postponement*. The former three postponement strategies are linked to production and manufacturing, while the last one is a pure pricing strategy. They aim at balancing the costs and benefits of mass production and mass customization. Practical examples of postponement can be found in the high-tech industry, food industry and other industries that require high differentiation.

However, empirical studies have found that postponement may not be an evident SCM practice compared to the other practices. In addition, postponement has both positive and negative impacts on a supply chain. The advantages include following the JIT principles, reducing end-product inventory, making forecasting easier and pooling risk. The high cost of designing and manufacturing generic components is the main drawback of postponement. Thus, the evaluation of postponement strategy is an important research issue and there have been many qualitative and quantitative models for analyzing postponement under different scenarios.

The core of this book is to analyze how the pull postponement strategy and the form postponement strategy can be leveraged to yield substantial benefits to adopting firms in different competitive environments. This book is made up of seven chapters, the contents of which are outlined in the following.

In Chapter 1 we review the status of development of postponement. We begin with presenting a framework to link postponement with product variety, mass customization and quality. We then identify four types of postponement, followed by providing a review of the cost models for analyzing various postponement strategies. Finally, we present review of the literature pertinent to our model development.

In Chapter 2 we develop an EOQ-based model to examine the cost impact of the pull postponement strategy adopted by a supply chain that orders and keeps n end-products. We formulate a total average cost function for ordering and keeping the n end-products in a supply chain, in which their demands are known and deterministic. Using standard optimization techniques, we show that postponed customization of end-products will result in a lower total average cost and a lower EOQ. Furthermore, we develop an EOQ-based model with perishable items to evaluate the impact of item deterioration rate on inventory replenishment policies. Our theoretical analysis and computational results show that a postponement strategy for perishable items can yield a lower total average cost under certain circumstances.

In Chapter 3 we develop two EPQ-based models with and without stockout to examine the impact of postponement. We formulate the total average cost functions of the two scenarios for producing and keeping n end-products in a supply chain, in which their demands are known and deterministic. Using standard optimization techniques, we show that postponed customization of end-products results in a lower total average cost in certain circumstances. We also find that two key factors that influence postponement decisions are variance of the machine utilization rates and variance of the backorder costs.

In Chapter 4 we study the cost impact of the pull postponement strategy by comparing the total average cost function with the optimal or an approximately optimal total average cost of an (r,q) policy. This is a stochastic model of a single end-product supply chain that consists of a supplier, a manufacturer and a number of customers. We develop two distinct models to represent the inventory system of the manufacturer. We employ Markov chain analysis to determine the exact average inventory level and the exact average accumulated backorder per period at the steady state so that the total average cost can be evaluated analytically. Also, we design an algorithm to find a near optimal total average cost per period. Our results show that the postponement system is more cost effective when the lead-time is zero, while the (r,q) inventory system is more effective when the lead-time is greater than zero.

In Chapter 5 we conduct simulation experiments of a two-end-product supply chain, for which customer demands are discrete and independent. Customer demands follow a uniform, Poisson or normal distribution. Two simulation models, namely one is a postponement system while the other is a non-postponement system, are designed for comparing their performance and total cost after t periods. Given a set of (r,q) policies and a demand distribution, the postponement system outperforms the non-postponement system in terms of average order frequency, average on-hand inventory, average backorder and average fill-rate. Thus, this system provides some cost benefits when the net postponement cost is low.

In Chapter 6 we report on two case studies of applying postponement strategy in industry. The first case is a study of a Hong Kong based toaster manufacturing company, which has successfully implemented postponement strategy. We present a summary of how postponement strategy was implemented in its supply chain and elaborate on all the benefits arising from the implementation of postponement. We also discuss the implications of postponement for its supply chain. In the second case study we present an empirical analysis of the application of postponement strategy in Taiwanese information technology (IT) firms. We present the findings and discuss their managerial and practical implications.

In Chapter 7 we conclude the book and suggest some worthy topics for future research.

This book is intended for researchers in supply chain management interested in conducting in-depth studies on postponement strategy. The book is also intended for business practitioners seeking to understand the nature and law governing the working of postponement strategy and looking for guidance and decision support for the implementation of postponement strategy. Therefore, the book can be useful not only for researchers but also for practitioners and graduate students in operations management, management science, industrial engineering, and business administration.

We would like to thank many friends and colleagues for their help and support rendered to us in preparing this monograph. First, we thank Prof. Fangruo Chen of Columbia University, Prof. Xiuli Chao of the University of Michigan, Prof. Jeannette Song of Duke University, Prof. Gang Yu of the University of Texas at Austin, Prof. Hanqin Zhang, Prof. Ke Liu and Dr Jingan Li of the Academy of Mathematics and Systems Science of the Chinese Academy of Sciences for their helpful discussions, suggestions and valuable comments on our research in this area. We also thank many scholars who have made important contributions in this promising area, including Prof. Remko van Hoek of the Cranfield School of Management, Prof. Christopher S. Tang of the University of California at Los Angeles, Prof. Hau L. Lee of Stanford University, Prof. Jyh-Shen Chiou of National Chengchi University, Prof. Lei-Yu Wu of Van Nung Institute of Technology, and Prof. Jason C. Hsu of the University of California at Los Angeles, whose original research has inspired us to join this exciting field of research. Finally, we would like to thank the National Natural Science Foundation of China, the Research Grants Council of Hong Kong, the Chinese Universities Scientific Fund, the Natural Science Fund for Young Scholars of Beijing University of Chemical Technology, the Hong Kong Polytechnic University, the Academy of Mathematics and Systems Science of the Chinese Academy of Sciences, and Beijing University of Chemical Technology for their financial support to our research.

Hong Kong, China	T.C. Edwin Cheng
Beijing, China	Jian Li
Hong Kong, China	C.L. Johnny Wan
Beijing, China	Shouyang Wang

Contents

1 Introduction 1
1.1 From Product Variety to Postponement 1
1.1.1 Product Variety 1
1.1.2 Mass Customization 2
1.1.3 Postponement Strategy 3
1.2 Classification of Postponement 3
1.2.1 Pull Postponement 4
1.2.2 Logistics Postponement 6
1.2.3 Form Postponement 7
1.2.4 Price Postponement 8
1.2.5 Implications 8
1.2.6 Advantages and Disadvantages of Postponement 9
1.2.7 Prerequisites for Postponement Strategy Development 10
1.3 Cost Models for Analyzing Postponement Strategies 11
1.3.1 Stochastic Models 11
1.3.2 Heuristic Models 12
1.3.3 Descriptive Models 13
1.3.4 Performance Measures 14
1.4 A Literature Review for Model Development 14
1.4.1 EOQ and EPQ Models 15
1.4.2 Lot Size-Reorder Point Model 16
1.4.3 Markov Chain 16
1.5 Concluding Remarks 17

2 Analysis of Pull Postponement by EOQ-based Models 19
2.1 Postponement Strategy for Ordinary (Imperishable) Items 19
2.1.1 Proposed Model and Assumptions 19
2.1.2 Case 1: Same Backorder Cost 22
2.1.3 Case 2: Different Backorder Costs 26
2.1.4 A Numerical Example 30
2.2 Postponement Strategy for Perishable Items 32
2.2.1 Notation and Assumptions 33
2.2.2 Model Formulation 34

2.2.3 The Postponement and Independent Systems 38
2.2.4 Numerical Examples 39
2.3 Concluding Remarks 41

3 Analysis of Postponement Strategy by EPQ-based Models 43
3.1 Analysis of Postponement Strategy by an EPQ-based Model without Stockout 43
3.1.1 Proposed Model and Assumptions 43
3.1.2 2 Machines for 2 End-Products 46
3.1.3 *n* Machines for *n* End-Products 56
3.2 Analysis of Postponement Strategy by an EPQ-based Model with Planned Backorders 62
3.2.1 Proposed Model and Assumptions 63
3.2.2 Demands Are Met Continuously 65
3.2.3 Demands Are Met After Production Is Complete 71
3.3 Concluding Remarks 78

4 Evaluation of a Postponement System with an (r, q) Policy 81
4.1 The Proposed Models and Assumptions 81
4.2 System Dynamics for a Non-postponement System 83
4.3 The Algorithm for Finding a Near Optimal Total Average Cost of an (r, q) Policy 84
4.3.1 The Markov Chain Development 84
4.3.2 The Algorithm for Finding a Near Optimal Total Average Cost 99
4.4 System Dynamics for a Postponement System 102
4.5 Average Cost Comparison of the Two Systems When $L = 0$ 103
4.6 Average Cost Comparison of the Two Systems When $L \geq 1$ 104
4.6.1 An Overview of the Simulation Results 104
4.6.2 Impacts of Parameters on Average Cost 106
4.7 Concluding Remarks 107

5 Simulation of a Two-End-Product Postponement System 109
5.1 Proposed Model and Assumptions 110
5.1.1 Notation 111
5.1.2 Model Assumptions 111
5.2 Methodology 112
5.2.1 System Dynamics 112
5.2.2 The Simulation Model 114
5.2.3 Customer Demand Distribution 114
5.2.4 Order Quantity and Reorder Point 115
5.2.5 Summary of Parameters 115
5.2.6 Initial Conditions 115
5.3 Simulation Results for Non-cost Parameters 117

5.3.1 Uniform Distribution ... 117
5.3.2 Poisson Distribution ... 118
5.3.3 Normal Distribution I ... 119
5.3.4 Normal Distribution II ... 120
5.4 Simulation Results for Cost Parameters ... 121
5.5 Concluding Remarks ... 123

6 Application of Postponement: Examples from Industry ... 125
6.1 A Case Study from Hong Kong ... 125
6.1.1 An Overview of the Company ... 126
6.1.2 Implementation of Postponement ... 126
6.1.3 Benefits of Using Postponement ... 127
6.1.4 Implications ... 128
6.2 The Case of Taiwanese Information Technology Industry ... 129
6.2.1 The Hypothesis ... 129
6.2.2 Methodology ... 130
6.2.3 Results ... 131
6.2.4 Implications ... 131
6.3 Concluding Remarks ... 132

7 Conclusions, Implications and Future Research Directions ... 133
7.1 Conclusions ... 133
7.2 Implications and Further Research Directions ... 134

A Simulation Results (Uniform Distribution) ... 137

B Simulation Results (Poisson Distribution) ... 141

C Simulation Results (Normal Distribution I) ... 147

D Simulation Results (Normal Distribution II) ... 151

E Simulation Results for Cost Analysis ... 155

References ... 157

About the Authors ... 163

Index ... 165

List of Figures

1.1 A four-step supply chain process 5
1.2 A revised four-step supply chain process 6
2.1 A schematic diagram of the independent system 20
2.2 A schematic diagram of the postponement system 21
2.3 Demand over lead-time in a cycle for end-product i without postponement 23
2.4 Demand over lead-time in a cycle for joint ordering n end-products with the same backorder cost 24
2.5 Demand over lead-time in a cycle for joint ordering n end-products with different backorder costs 28
2.6 Graphical representation of inventory level 35
2.7 The impact of deterioration rate on cost 40
2.8 The impact of deterioration rate on the difference in cost between the two systems 40
3.1 A schematic diagram of the independent system 44
3.2 A schematic diagram of the postponement system 44
3.3 IP for end-product 1 and 2 in the independent system when the demand is met continuously 47
3.4 IP for end-product 1 and 2 in the postponement system when the demand is met continuously 49
3.5 IP for end-product 1 and 2 in the independent system when the demand is met after production is finished 52
3.6 IP for end-product 1 and 2 in the postponement system when the demand is met after production is finished 53
3.7 IP for product i in the non-postponement system when the demand is met continuously 66
3.8 IP for n products in the postponement system when the demand is met continuously 67
3.9 IP for product i in the non-postponement system when the demand is met after production is finished 72
3.10 IP for n products in the postponement system when the demand is met after production is finished 73
4.1 A schematic diagram of the postponement system 82

4.2 A schematic diagram of the non-postponement system 82
4.3 A plot of probability distribution diagram 96
5.1 Process flow of the non-postponement system 110
5.2 Process flow of the postponement system 110
5.3 Flow chart for the non-postponement system 113
5.4 Flow chart for the postponement system 114
6.1 A process flow chart of the toaster company 127
A.1 Difference in average inventory (Uniform Distribution) 139
A.2 Difference in average backorder (Uniform Distribution) 140
A.3 Improvement in average fill rate (Uniform Distribution) 140
B.1 Difference in average inventory (Poisson Distribution) 144
B.2 Difference in average backorder (Poisson Distribution) 145
B.3 Improvement in average fill rate (Poisson Distribution) 146
C.1 Difference in average inventory (Normal Distribution I) 149
C.2 Difference in average backorder (Normal Distribution I) 150
C.3 Improvement in average fill rate (Normal Distribution I) 150
D.1 Difference in average inventory (Normal Distribution II) 153
D.2 Difference in average backorder (Normal Distribution II) 154
D.3 Improvement in average fill rate (Normal Distribution II) 154

List of Tables

2.1 Parameters of 5 end-products . . . 31
2.2 The impact of deterioration rate on inventory replenishment policies . . . 39
2.3 The impact of deterioration rate on the difference in cost . . . 40
3.1 A summary of the eight hypotheses without stockout . . . 45
3.2 A summary of the four hypotheses with stockout . . . 63
3.3 Impact of backorder costs when the demands are met continuously . . . 76
3.4 Impact of machine utilization rates when demands are met continuously . . . 77
3.5 Impact of backorder costs when the demands are met after production is finished . . . 77
3.6 Impact of machine utilization rates when the demands are met after production is finished . . . 77
3.7 A summary of the findings of the eight hypotheses without stockout . . . 78
3.8 A summary of the findings of the four hypotheses with stockout . . . 79
4.1 Number of states required . . . 86
4.2 The 13 data sets and their required states . . . 94
4.3 The probability distributions (in terms of N) of the 13 data sets . . . 95
4.4 The probability distributions (in terms of R) of the 13 data sets . . . 96
4.5 Iterative procedure for $L = 1, a = 10$ . . . 103
4.6 A summary of the simulation results . . . 105
5.1 (r, q) policy in different distributions . . . 116
5.2 Customer demand sets (Uniform Distribution) . . . 117
5.3 Lot size-reorder points (r, q) for the non-postponement system and the postponement system (Uniform Distribution) . . . 117
5.4 Customer demand sets (Poisson Distribution) . . . 118
5.5 Lot size-reorder points (r, q) for the non-postponement system and the postponement system (Poisson Distribution) . . . 118
5.6 Customer demand sets (Normal Distribution I) . . . 119
5.7 Lot size-reorder points (r, q) for the non-postponement system and the postponement system (Normal Distribution I) . . . 119
5.8 Customer demand sets (Normal Distribution II) . . . 120
6.1 Lead-time breakdowns for toaster offering . . . 126
6.2 Change in inventory before and after postponement implementation . . . 128

6.3 The background statistics of the sample 131
7.1 Summary of book findings 135
A.1 Simulation results 1 (Uniform Distribution) 137
A.2 Simulation results 2 (Uniform Distribution) 138
B.1 Simulation results 1 (Poisson Distribution) 141
B.2 Simulation results 2 (Poisson Distribution) 142
C.1 Simulation results 1 (Normal Distribution I) 147
C.2 Simulation results 2 (Normal Distribution I) 148
D.1 Simulation results 1 (Normal Distribution II) 151
D.2 Simulation results 2 (Normal Distribution II) 152
E.1 Simulation results for cost analysis 156

Chapter 1
Introduction

Postponement has been discussed by numerous scholars and researchers from different perspectives since 1950s when Alderson [3] observed that products tend to differentiate when they near the time of purchase in order to reduce marketing costs. He named this concept the principle of postponement. Postponement, also known as late customization or delayed product differentiation, refers to delaying some product differentiation processes in a supply chain as late as possible until the supply chain is cost effective (Garg and Lee [43]). It gives rise to economies of scale and scope through product and process standardization and customization, respectively. The goal of postponement is to supply desirable products to customers at a relatively low cost and in a responsive way (Feitzinger and Lee [41]). Li et al. [73] selected postponement as one of five major SCM practices that have discernible impact on competitive advantage and organizational performance. In fact, postponement is an element of mass customization that is applied to cope with product variety in order to enhance customer quality. A comprehensive literature review of how service quality is improved by an adoption of postponement to deal with a wide product range is presented in this chapter.

This chapter is organized as follows. In Section 1.1 the relationships between product variety, mass customization and postponement are discussed. In Section 1.2 four types of postponement are addressed. The advantages and disadvantages of postponement strategies and the prerequisites for postponement strategy development are also discussed. In Section 1.3 a review of the cost models for analyzing various postponement strategies is provided. A literature review for our model development is presented in Section 1.4. In Section 1.5 we conclude the chapter.

1.1 From Product Variety to Postponement

1.1.1 Product Variety

MacDuffie et al. [75] defined product variety as the breath and depth of product lines. Strategically, broadening product variety gives a company a distinctive competitive advantage in quickly responding to ever-changing market environments and

T.C.E. Cheng et al., *Postponement Strategies in Supply Chain Management*,
International Series in Operations Research & Management Science 143,
DOI 10.1007/978-1-4419-5837-2_1,

customer tastes. Product proliferation can lead to greater flexibility, larger market shares, deeper market penetration, higher customer satisfaction and loyalty. However, it increases the number of set-up and inventory level, and demands more technical services, material handling, supervision, co-ordination and support in production (Yeh and Chu [128]). Since expanding product variety increases both values and costs, only value-added variety should be offered.

Product variety falls into two dimensions: perceived variety and actual variety. Customer value is only enhanced by an increase in perceived variety, regardless of the extent of actual variety offered (Porter [94] and Kahn [60]). Adding actual variety (but not perceived variety) not only increases customer confusion but also raises costs. Stalk [109] identified two forms of costs: those responding to volume or scale, and those driven by variety. The former reduces when volume increases, while the latter increases when manufacturing becomes complex as variety increases.

Since handling product variety is a complicated process, Ulrich et al. [115] identified 12 supply chain decision areas and classified them into strategic level and tactical level for coping with product variety. Strategic decisions include: (1) dimensions of variety offered, (2) distribution channel, (3) degree of vertical integration, (4) process technology, (5) position of decoupling point, and (6) product architecture. Tactical decisions cover: (1) number and combination of product attributes, (2) extent of parts sharing, (3) lot size policy, (4) inventory management policy, (5) production scheduling, and (6) promotion plans. Besides, information about product lines, customer behavior and tastes, market segmentation, suppliers, technological innovation, and strengths and weaknesses of competitors is vital for product variety strategy planning (Porter [94]).

1.1.2 Mass Customization

As mentioned above, an effective product variety strategy calls for the provision of highly perceived variety of products and services to individual customers. From a supply chain management's point of view, it aims at supplying customized products and services to each customer in a responsive and cost effective way. It is important to note that customization is a time consuming and costly strategy due to diseconomies of scale and scope. The associated costs include the logistics of managing variety, material handling, quality, production capacity and inventory (Martin et al. [76]). Mass customization (Feitzinger and Lee [41]) or customized standardization (Lampel and Mintzberg [63]), are aimed at balancing both standardization and customization in order to achieve both quick response and low cost. It is a guiding strategy for those multinational companies that manufacture global products with customization for local markets.

Lee [65], Feitzinger and Lee [41], and Yeh and Chu [128] mentioned three prerequisites for mass customization. The first is modular design of production processes. Processes can be moved, assembled and re-arranged to support various products manufacturing easily and at low costs. Moreover, modular design allows concurrent production, isolates potential problems within module, and most importantly, enables postponement. One of these process-re-engineering techniques is operations reversal (Lee and Tang [70]). It is an approach in which the sequence of two consecutive processes in a supply chain is reversed to reduce demand variability. The second requirement is products and parts standardization. Homogenous products are produced and they are capable of supporting multiple product functions and features with slight re-configurations. It reduces costs in part number administration, leads to inventory reduction and facilitates supplier management (Lee [64]). The last building block is a flexible supply chain, which involves the coordination and negotiation of marketing, R&D, manufacturing, distribution, finance and retailing, to support both generic and customized products responsively (Lee [64]).

1.1.3 Postponement Strategy

Garg and Lee [43] suggested applying a postponement strategy to deal with product variety. The concept was first discussed by Alderson [3]. In his principle of postponement, he argued that products differentiation in the point of purchase could reduce various marketing costs. In this regard, he suggested all changes in form and identity be delayed to the latest possible point in time and location. Christopher [29] referred to postponement as a vital element in an agile strategy, which adopts both a flexible manufacturing system (FMS) and standardization of products to achieve organization flexibility.

1.2 Classification of Postponement

Postponement, also known as late customization or delayed product differentiation, refers to delaying some product differentiation processes in a supply chain as late as possible until the supply chain is cost effective (Garg and Lee [43]). It implies economies of scope and scale can be achieved by product and process standardization. Economies of scale are made possible through standardization of components and processes to support a large variety of products. Economies of scope are achieved by producing various products at the same time (Pine [93]).

Under a postponement strategy, products from a product family share common parts and processes until their point of product differentiation. After the point of product differentiation, a “fan-out” occurs, as end products require different components and processes (Garg and Lee [43]). This strategy benefits from the “risk-pooling” effect, which suggests demand variability is reduced by considering aggregate demand instead of individual demand in a product category (Federgruen and Zipkin [40], Simchi-Levi et al. [105]). The major reason is that one extremely high demand may be offset by another extremely low demand after aggregation. Lower demand variability implies fewer safety stocks, lower inventory levels and more accurate resources planning. Apart from reducing demand variability, postponement is also used to tackle process and supply uncertainties.

There are four forms of postponement strategies, namely pull postponement, logistics postponement, form postponement (Lee [65]) and price postponement (van Mieghem and Dada [124]). The former three strategies are also referred to as production postponement (van Mieghem and Dada [124]).

1.2.1 Pull Postponement

Companies usually employ forecasting techniques to estimate customer demand. Products are produced in advance of customer orders, where production is planned to achieve optimal capacity and efficiency. Such a production strategy is called a make-to-stock (demand-push) strategy. One advantage of this strategy is that it has immediate stock availability (Schroeder [101], Browne et al. [19] and Arnold [5]) to respond to customer orders quickly. However, unavoidable overstock or stock outs occur due to demand forecast variations. The variation is amplified from downstream to upstream in a supply chain due to information distortion. This phenomenon is described as the "bullwhip effect" (Lee et al. [68]). Stock outs lead to losing customers while overstocked items become obsolescence at the end of the product life cycle. The costs associated with overstock and stock outs can be huge. In order to cope with these unwanted variations, a make-to-order (demand-pull) strategy is advocated.

In contrast to a make-to-stock strategy, a make-to-order strategy pulls production once a customer order is received. It can entirely eliminate unwanted inventory as production quantity is set after demand uncertainty is resolved (van Mieghem and Dada [124]). However, long order lead-time and wide production fluctuations are two major drawbacks of this strategy.

In practice, most companies operate between these two extremes in order to balance production capacity and demand (De Hann et al. [35]). A customer order decoupling point (Hoekstra and Romme [55], Browne et al. [19]) or a push-pull boundary (Brown et al. [17]), which indicates the extent of which a customer order penetrates into the goods flow, separates forecast-driven activities upstream from demand-driven activities downstream in a supply chain system (van Donk [117], van Hoek [118]). If a system cannot be decoupled, it should be operated only in either a fully forecast-driven or fully demand-driven mode (Hoekstra and Romme [55]). The concept of a customer order decoupling point was proposed by Bucklin [21] in 1965. Bucklin developed a postponement-speculation approach such that inventory is only built at each process in a distribution channel whenever the costs are less than the savings among channel members.

Pull postponement, also known as process postponement (Brown et al. [17]), refers to moving the decoupling point earlier in the supply chain such that fewer steps will be performed under forecast results (Lee [65]). There are two decoupling points: one for the physical supply chain and the other for the virtual supply chain (Christopher [29]). The virtual supply chain refers to the information-sharing network among supply chain partners. In pull postponement, customer demand is the key information for determining the push–pull boundary of a physical supply chain. In fact, it is an information strategy (Lee [65]).

For instance, assume there are four processes in a supply chain where process 4 receives customer order (see Fig. 1.1). If process 2 and process 3 update the customer order information simultaneously, the decoupling point for the virtual supply chain is in process 2. However, the pull postponement strategy can be carried out in either process 2, process 3 or process 4 but not in process 1 because it cannot receive customer order information for pull production. Process 1 can employ a push system only. In theory, the best pull postponement is implemented in process 2 as it is the earliest process that can be operated once an order is received. However, if the process lead-time for process 2 through process 4, L, is longer than the customer expected waiting time W, the decoupling point should be moved downstream to process 3 or 4. Otherwise, customers need to wait extra time for the product or they take their businesses to other suppliers.

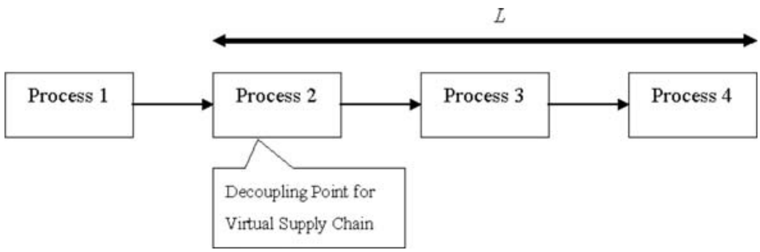

Fig. 1.1 A four-step supply chain process

1.2.1.1 Customer Order Decoupling Point and Point of Product Differentiation

Recall that a postponement strategy aims at moving the point of product differentiation as late as possible until it is cost effective. On the other hand, the purpose of a pull postponement strategy is to move the decoupling point earlier to minimize forecasting error. Thus, they seem to be unrelated and contradicting concepts. However, they are in fact related and vital for a postponement strategy to be carried out successfully. There is little literature discussing their relationships.

Rethink the four-process supply chain again in Fig. 1.1. This time the supply chain is analyzed from the product, instead of the process, perspective. Assume that there are two products, A and B, whose distinctive features are added after the completion of process 2. The revised diagram is shown in Fig. 1.2.

Assume further that information is shared and updated throughout the supply chain. If the decoupling point D lies in process 1, the customer waiting time would be L_1. It is a pure make-to-order system. In practice, production volume fluctuates widely due to demand variability and all orders are backlogged. In order to smooth the production schedule and make the system more responsive, D is moved

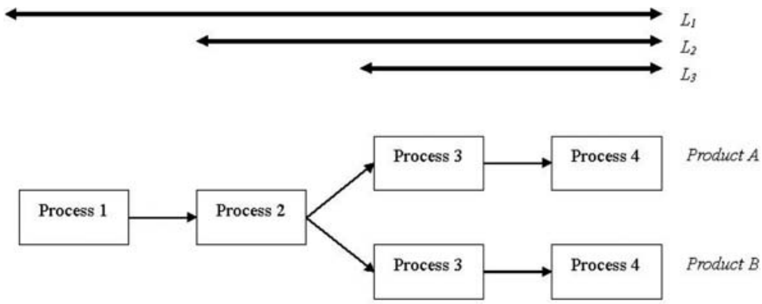

Fig. 1.2 A revised four-step supply chain process

to process 2 such that process 1 is make-to-stock. The customer waiting time is $L_2 < L_1$. Since inventory has been built up after process 1, there is little intention to stop passing process 2 in a series supply chain system until customer orders are resolved. In this sense, D will be further passed to process 3 and the customer waiting time is shortened to L_3. If D is further moved to process 4, demand variability increases as products start to differentiate after process 3. There are risks of overstock and stock outs of products A and B if D is set after the point of differentiation K. As a consequence, D should be set before K in order to implement postponement. D is determined by customer expected waiting time and availability of demand information in the system, while K is determined by operational factors such as manufacturability, product characteristics, costs and so on.

1.2.1.2 Applications of Pull Postponement

National Bicycle differentiates themselves from other traditional bicycle manufacturers by adopting a different decoupling point. They move the decoupling point from final assembly to frame welding such that they could offer more than 11 million options to customers quickly (Lee [65]). Xilinx employs a pull postponement strategy in semiconductor manufacturing (Brown et al. [17]). Dies are transformed to specialized integrated circuits (ICs) once customer orders are received. Xilinx management needs to monitor 100 types of work-in-process (WIP) inventory of dies, instead of 10,000 finished products. Corporate inventory drops by 23% after implementation for 1 year (Brown et al. [17]). Besides, Wings & Legs, a large poultry processor in the Netherlands, delays their packaging and labeling processes as their key clients demand tailor-made packages and labels for Wings & Legs' chicken legs (van Dijk et al. [116]). Benetton, an apparel manufacturer, postpones its color dyeing process until orders are received (Lee and Tang [70]). They all enjoy substantial savings through inventory reduction.

1.2.2 Logistics Postponement

Logistics postponement involves the re-designing of some of the processes in the supply chain so that some customization can be performed downstream closer to customers (Lee [65]). Bowersox and Closs [14] defined logistics postponement as an emerging strategy of form, time and place postponement. Similar to pull postponement, their meanings of form and time postponements refer to delaying some customization activities until customer orders are received. Place postponement refers to the positioning of inventories upstream or downstream. In general, logistics postponement is an extension of pull postponement.

Logistics postponement considers whether pull postponement can be implemented more effectively and efficiently by relocating some demand-pull processes closer to customer levels (Lee et al. [67] and Lee [65]). It is associated with the concepts of design for localization or design for customization, which takes operational and logistics services into account in the design process so as to serve different market segments (Lee et al. [67]). Packaging postponement and labeling postponement (Twede et al. [114]) or branding postponement (Ackerman [2]) can be subsets of logistics postponement when the packaging, labeling or branding processes are moved closer to customers.

1.2.2.1 Applications of Logistics Postponement

Hewlett-Packard produces generic printers at its factory and distributes them to local distribution centers, where power plugs with appropriate voltage and user manuals in the right language are packed. Since generic printers are lighter, more units could be shipped. Distribution cost is cut by half, and million dollars have been saved, although manufacturing costs are slightly higher due to the use of standardized components to support mass customization (Lee [65], Lee et al. [67], Feitzinger and Lee [41]).

WN, a European wine producer, demonstrates another real-life example of logistics postponement application. WN produces base wine in central bodega and defers bottling, packaging and labeling activities at the local level (van Hoek [118]). Ackerman [2] quotes a similar strategy adopted by Coca-Cola, in which concentrated syrup is shipped to retailers where it is mixed with carbonated water to form Coca-Cola in retailers' soda fountains.

Twede et al. [114] presented a logistics postponement application at Swedish furniture retailer IKEA. All products in IKEA retail stores are kept in semi-finished forms (flat packs) and are assembled by customers or deliverymen after home delivery. In this way, truckload capacities can be utilized and configurations can be easily made at customer locations. Kellogg Company, the world's leading cereal food producer, ships its products in bulk to local co-packers, where some constituents are added and packed (Brown et al. [18]).

1.2.3 Form Postponement

Form postponement, also called product postponement (Brown et al. [17]), opts for a fundamental change of the product structure by using standardized components and processes to achieve high customization (Lee [65] and Brown et al. [17]). Bowersox and Closs [14] introduced form postponement as a postponement of the final manufacturing or processing activities. Their concept is akin to Lee's [65] definition of pull postponement. In order not to mix up the two concepts, the version of Lee [65] and Brown et al. [17] is adopted. In fact, form postponement is an enabling strategy to supplement pull postponement (Lee [65]).

1.2.3.1 Applications of Form Postponement

Brown et al. [17] applied form postponement in a semiconductor company (Xilinx), where it re-designs the IC so that it could be re-configured by software easily and quickly for customized features and functions. It is particularly useful in programmable devices because a nearly infinite number of products can be produced by using program configuration.

1.2.4 Price Postponement

Van Mieghem and Dada [124] defined price postponement from economic and marketing perspectives. They described price postponement as a strategy aimed at deferring the pricing decision until customer demand is known. Selling price is negotiated with customers after they place their orders. Based on their findings, one advantage of the price postponement strategy is that it makes investment and production decisions insensitive to demand uncertainty, since profit margin can be covered by setting various selling prices after demand is known. Another advantage is its ease of implementation. Unlike the above three postponement that require re-engineering techniques such as operations reversal and standardization of product and process, price postponement is a managerial decision that is determined by marketers.

1.2.4.1 Applications of Price Postponement

Bank of China (BOC) Hong Kong applied a price postponement strategy for its initial public offering in Hong Kong in July 2002. In the face of high demand uncertainty, BOC set an offering price range, between HK$6.93 and HK$9.5 per share, for investors to subscribe to its shares that were worth HK$6.93 per share (*South China Morning Post*, 23 July 2002 [108]). Since the public offering was over-subscribed by 26 times, BOC finally allotted at least 500 shares to each investor at a price of HK$8.5 (*South China Morning Post*, 22 July 2002 [107]).

1.2.5 Implications

Practically, postponement strategies can be combined, integrated or partly applied to a supply chain in order to achieve different objectives. Recall the model shown in Fig. 1.2, if the aggregate demand of products *A* and *B* is known and standardized components are ready in process 2 (point of differentiation) for customization in process 3 and process 4, both form postponement and pull postponement strategies are applied. Furthermore, if process 3 and process 4 are performed locally for products *A* and *B*, instead of in central production facilities, logistics postponement is adopted. In selling, if product prices are set after customer orders are received for product *A* and product *B* respectively, a price postponement is employed. This example is flexible to be applied in all supply chain systems.

Not all products can be made shortly after customer demands are known. If they are not available within a promising time, customers may turn down the business. Therefore, there is a need to keep some inventory for these products, while a make-to-order approach is used to those products that can be produced within customer expected waiting time. This mixed strategy is referred to as a partial postponement strategy or a hybrid postponement strategy (Brown et al. [17], Graman and Magazine [50]), in which the lead-times after the point of differentiation and customer expected waiting time are key determinants. Xilinx keeps both semi-finished dies and finished IC inventory in dealing with long back-end lead-time of some products (Brown et al. [17]).

1.2.6 Advantages and Disadvantages of Postponement

1.2.6.1 Advantages of Postponement

To some extent, the philosophy of postponement follows the JIT principle, as both emphasize to have the right product in the right place at the right time (Cheng and Podolsky [27], and Heskett [52]). In fact, postponement offers substantial advantages for a supply chain to improve in terms of time, quality and cost. Graman and Magazine's study [50] found that although postponement results a reduction in inventory in terms of quantity, the service level is unchanged. In general, less inventory held makes inventory management easier and more responsive. On the other hand, perceived product quality is enhanced by small design changes (Lee [64]). Besides, standardized components can reduce the risk of obsolescence since configurations become easier in the form of WIP inventories, instead of end products (Brown et al. [17]).

Postponement makes forecasting easier at a generic level than at the level of finished forms because demand variability is reduced by aggregation (Christopher [29], and Ernst and Kamrad [38]). It is particularly obvious under a multi-echelon supply chain system in which the demand of the current stage is equal to the demands of the previous stages. Moreover, it supports various production alternatives such as engineering to order, purchasing to order, make to order, manufacture/assemble to order, packaging and labeling to order, shipment to order and adjust to order by shifting the customer order decoupling point (van Hoek [118], Olhager [89], and Hoekstra and Romme [54]). Its modularity characteristic not only reduces the cost of assembly (Chiou et al. [28]) but also enables outsource capability and speeds up new product development (Brown et al. [17], and Ernst and Kamrad [38]). The higher degree of modularity, the more outsourcing opportunities a company can pursue. Thus, fixed investment can be reduced drastically. In particular, logistics postponement can initiate the use of third party logistics (3PL) to handle local value-added activities and product delivery (van Hoek [120]). Successful outsourcing examples are Dell Computer, Nike, Reebok and General Motors (Tully [113]).

1.2.6.2 Disadvantages of Postponement

Notwithstanding postponement brings significant benefits to a supply chain, the cost of re-engineering and developing the supply chain cannot be neglected. Since a postponement strategy aims at delaying customized activities as late as possible until customer orders are received, more work-in-process inventories need to be built up before the point of differentiation (Brown et al. [17]). To a certain extent, standardized components increase variable costs because they need to support various product features (Lee [65], Ma et al. [74]). Power plugs, for example, should be re-designed so that a switch is added in order to support different voltage supplies in different countries. The modification makes variable costs higher. However, variable costs can be reduced by shifting the assemble processes to local facilities so that there is always one suitable type of power plug available for the product. In this scenario, other factors such as transportation cost, setup cost, training cost and local material cost should be weighted against variable cost savings.

Moreover, there is always a trade-off between mass production and customization in carrying out postponement as the former gains economies of scale while the later gains higher customer values. Economies of scale are lost after the point of differentiation due to customization (Zinn and Bowersox [130]). This effect is more pronounced in logistics postponement since customized processes are performed separately in local facilities with different product lines. To a large extent, it is associated with a high risk of creating quality problems as production is moved from central facilities to local facilities or even at the retailer level (Ackerman [2]).

1.2.7 Prerequisites for Postponement Strategy Development

Postponement is not a panacea for all industries, as it may not be possible or economical for companies to re-design common processes and components such that specific end products are produced from a group of generic products. Recall that Lee [65], Feitzinger and Lee [41], and Yeh and Chu [128] mentioned there are three building blocks for carrying out mass customization. Lee [65] described four postponement enablers, namely modularity, design for postponement, supply chain collaboration and associated costs. Besides, there are additional prerequisites that are conducive to postponement strategy development. All prerequisites are summarized below:

(i) Mass customization principles should be embedded throughout the supply chain system such that the three building blocks are present to support and facilitate postponement implementation.
(ii) Products can be categorized into various product families such that each product family shares common characteristics in terms of product design, standardized parts, common production processes and the same production location.

In other words, a point of product differentiation can be defined. Bills of materials (BOM) help one to look for opportunities to group common parts at a lower level of the product structure so that products in the same product family share more common production processes (Kennedy et al. [62]).

(iii) Operations should be closely linked with product and process design (Ma et al. [74]). That is, the philosophy of design for flexible manufacturing is essential for effective postponement strategy development. The concept of postponement should be embedded in the design process such that not only cost, quality, flexibility and serviceability but also distribution, service, maintenance, marketing, manufacturing capabilities, inventory management and supplier management are considered (Barkan [11] and Calvin and Miller [23]).

(iv) Postponement is particularly powerful if the supply chain network operates on a global scale in which the positioning of inventory, production mode and structure, and distribution facilities become critical success factors for both cost reductions and customer value creation. Global efficiency can be achieved through mass production while local responsiveness can be enhanced by customization (van Hoek et al. [123]).

(v) Lead-times after the point of postponement should be justified with respect to customer expected waiting time for the product to avoid backorders and lost sales (Lee [64]). Under certain circumstances, partial postponement may be used to reduce lead-times.

(vi) Information technology and greater supplier involvement are vital factors to streamline the supply chain process (Brown et al. [17]).

1.3 Cost Models for Analyzing Postponement Strategies

In general, models for analyzing postponement strategies can be classified into four types. They are deterministic models, stochastic models, heuristic models and descriptive models based on case studies (Beamon [12]).

1.3.1 Stochastic Models

We define a stochastic model as an inventory model where demand in any period is random (Hillier and Lieberman [53]). Lee and Tang [69] formulated a total relevant cost model to analyze the effectiveness of a designed strategy to be applied in a N-stage manufacturing system that produces two products, whose demands follow normal distribution. The first k operations are common for the two products. Their cost model consists of four cost factors, including total average investment cost, total processing cost, total WIP inventory cost and total buffer inventory cost. As an extension, Garg and Tang [44] considered a production system that has two product differentiation points. Their cost model showed that these two points also yield a lower inventory saving. Other cost models for analyzing the point of product differentiation can be found in Garg and Lee [43] and Lee [64].

Ernst and Kamrad [38] developed a total cost model to analyze four supply chain structures, namely rigid, modularized, postponed and flexible, of an ice-cream supply chain that serves two different markets. Demand of a particular flavour follows a probability density function, and the total cost under evaluation includes fixed cost, variable cost, holding cost and backorder cost. Surprisingly, they concluded that postponement is the worst choice among four. However, this conclusion is drawn directly from numerical analysis of an ice-cream supply chain. A more generalized model should be developed in order to provide a more concrete and reliable framework for evaluating supply chain effectiveness under various scenarios.

Aviv and Federgruen [6] modeled a two-phase production system in which common products are produced in the first phase and product differentiation is delayed to the second phase. Their objective is to minimize an expected long-run average discounted cost of j products, each of them follows multivariate distribution with arbitrary correlations. First, they found a lower-bound of the average cost for a single-stage single-product production. Then they extended it to a two-stage postponement system with a heuristic strategy. A numerical analysis and a study of Hewlett-Packard case are provided. They found that postponement results in substantial savings when coping with a large degree of product variety, and less correlated or high seasonality product demands.

Ma et al. [74] analyzed component commonality and postponement in a multi-stage multi-product assembly system. A set of common base-stock levels at all stocking points is found to minimize the total expected inventory cost. In their model, both lead-time and service level are considered. They showed that those processes with long procurement lead-times are rearranged after the point of product differentiation so as to meet the service level . Also, component commonality is more preferred to be implemented in earlier stages. In fact, the relationships between part commonality and aggregate safety stock are discussed by Collier [32], McClain et al. [78] and Baker et al. [9]. They all supported the view that less aggregate safety stock is needed for maintaining a constant service level .

1.3.2 Heuristic Models

Heuristic models are models that employ rule of thumb to approach the best solution (Ballou [10]). Brown et al. [18] presented an application of an enterprise resources planning (ERP) system in Kellogg Company. The system is called Kellogg Planning System (KPS). It is a rule-based heuristic program, which helps Kellogg Company to control its operations, production, inventory and distribution for breakfast cereal and other food products. The objective function is to minimize production costs, packing costs, inventory costs, shipping costs and penalty costs for overstocks and under stocks. Corry and Kozan [33] developed a push/pull hybrid production system (HIHPS) of a foundry from which a single assembly stage is demand-pulled. Simulated annealing algorithm is applied to determine the optimal buffer reorder point and replenishment level of the components that can minimize a total cost function. Another HIHPS system can be found in Cochran and Kim [31]. Graman and

Magazine [50] studied a manufacturing system that keeps both WIP and finished goods of a group of products so as to fulfill a given service level . They used a Monte Carlo integration method to determine the stock levels of both inventories. Besides those dynamic systems, Zinn [129] used a percent saving ratio to evaluate the safety stock savings resulted from postponement. He showed that a large product line yields a high safety stock saving when the demand and standard deviation of demand for each product are independent and approximately equal respectively. A simulation study by Johnson and Anderson also revealed that postponement improves fill rate and service level when demand for each product is at the same level [59].

1.3.3 Descriptive Models

Bucklin [21] proposed a postponement-speculation model to test six hypotheses that associated delivery time, product type, product cost and demand variability with the choice of postponement. He claimed that the point of postponement-speculation appears in a distribution channel whenever there are systemwide net savings resulted from postponement. It is one of the earliest research studies on postponement.

Van Hoek [118] made use of a case study approach to compare two logistics postponement strategies implemented in European wine producer WN in terms of transport cost, production cost, material cost, inventory holding cost and bottling cost. The first alternative is to delay final distribution to customer and the second one is to defer the bottling, packaging and labeling activities to the local level. He found that there are cost savings in transportation and inventory holding for both postponement alternatives. In addition, he pointed out that product characteristics, such as product value, volume and weight, affect the choice of postponement. In other papers [120, 121], he studied the role of third party logistics providers (3PL) in carrying out parts of the postponed supply chain activities. He anticipated that 3PL would play a crucial role in those postponed activities that in turn accelerate the adoption of postponement.

Zinn and Bowersox [130] conducted a discriminant analysis to explore factors affecting the choice of five postponement strategies, namely labeling, packaging, assembly, manufacturing and time postponements. They found that cost saving, product value and demand uncertainty are key drivers of postponement. Chiou et al. [28] employed factor analysis and path analysis to explore factors affecting postponement and the causal relationships between demand characteristics and postponement among Taiwanese IT firms.[1] They found that modular product design, component cost and product life cycle induce different choices of postponement. Also, Huang and Lo [56] described a postponed PC supply chain in Taiwan that sourced components globally and assembled locally. Jahre [57] applied postponement in a household waste collection process, which provides insights into postponement being applied in the context of reverse logistics to deal with

[1] The case study is presented in Chapter 6.

environmental issues. In his model, the waste-sorting process (separation of tin cans, plastic bottles and paper for recycling) can be either at the consumer level or delayed until the waste reaches the collection center.

1.3.4 Performance Measures

In summary, a set of performance measures need to be adopted to evaluate any analytical model formulated to study a particular postponement strategy. Basically, performance measures can be grouped into three categories, namely qualitative, quantitative and time measurements (Beamon [12]).

- Qualitative factors
  - Service level (Ma et al. [74], Graman and Magazine [50]), product quality (Martin et al. [76]) and information and material flow integration (Martin et al. [76] and Nicoll [85])
- Quantitative factors
  - Cost minimization

    Fixed costs, variable costs, inventory costs, distribution costs, per-unit penalty for stock out and overstock (Bucklin [21], Ma et al. [74], Aviv and Federgruen [6], Ernst and Kamrad [38], Lee and Tang [69])
  - Value maximization

    Sales, profits (Beamon [12])
- Time
  - Fill rate (Graman and Magazine [50]), order and production lead-time (Ma et al. [74]) and customer order response time (Lee [64])

1.4 A Literature Review for Model Development

To the best of our knowledge, there is only a small number of postponement models that are based on deterministic demand. Recent deterministic models include, among others, those by Wan [125], and Li et al. [71, 72]. There are new research opportunities in studying postponement by deterministic models such as economic order quantity (EOQ) and economic production quantity (EPQ) models. Both models can be used to derive a total cost function for analyzing postponement. They can help to analyze whether or not combining certain supply chain processes can reduce the total cost. One of the potential savings is from joint ordering [119] or joint production of a group of products whose demands are deterministic. EOQ and EPQ models, which relate to joint ordering and joint production, are reviewed and they provide many useful ideas and insights for our model development in the sequel.

1.4.1 EOQ and EPQ Models

The EOQ model is used to answer two questions: when and how many to order (Zipkin [131] and Hadley and Whitin [51]). Silver [104] developed a simple cycle policy based on the EOQ model to decide when to jointly replenish N groups of products, in which one group is replenished every T years while the other $N - 1$ groups is reordered every $k_n T$ years, where k_n is the number of integer multiples of T for the replenishment of item n. The idea originated from Shu [103] and Nocturne's revised version of Shu's optimal ordering frequency (Nocturne [86]). Their models only considered two groups of products. In addition to dealing with the issue of when to order, equal attention should be paid to the other issue: how many to order? An integrated EOQ model was presented by Cheng [26], and Chen and Min [25]. They considered profit maximization when multi-products are ordered jointly, subject to storage space and inventory investment constraints. They employed the Karush-Kuhn-Tucker (KKT) conditions to solve the problem. They assumed that there are no backorders.

On the other hand, there are numerous articles addressing multi-product production with EPQ. However, none of them focuses on comparing the total average cost between a postponement system and a non-postponement system. EPQ models for multi-products can be classified into two types: single machine and multi-machines. The single machine EPQ model follows a rotation cycle policy, by which end-products are produced in sequence in each production cycle [83]. Eilon [37] classified the production of several products by a single machine, in which product demand is known and only one product is produced at a time, as a multi-product batch scheduling problem. Instead of grouping the production, he split the batch into subbatches and compared the total cost per day. Goyal [46] studied a similar problem. He used a search procedure to determine the EPQs of two items. The classical problem is found in Tersine [112] and Nahmias [83]. Recall the logic behind a postponement strategy is to group certain processes together so as to lower the total cost. It is more appropriate to consider grouping the production of end-products instead of splitting their production across a production cycle.

Apart from the single machine scheduling problem, the multi-machine multi-item EPQ models fall into the other end of the spectrum. In dealing with this kind of problem, it is not uncommon to group a large number of end-products into different product families based on processing similarities and economic considerations so that each product family shares a single lot size [22]. It can greatly reduce manufacturing complexity. Byrne [22], and O'Grady and Byrne [87, 88] employed a simulation approach to find good production lot sizes for a number of product families that share common machines in order to minimize a total cost function. In their model, each product family is assumed to have a single lot size. A more dynamic model was presented by Bertrand [13]. He developed a total cost model for a production system that produces multi-items in multi-work centers. His model is based on batch size optimization and queuing theory and it accounts for both finished goods and WIP inventories. Good solutions are found by using the Newton-Raphson method. Besides, Goyal et al. [49] presented a realistic problem in determining the EPQ at each production stage for multiple items across a multi-stage production system.

However, due to problem complexity, their model can only be solved by a heuristic approach that yields sub-optimal solutions.

1.4.2 Lot Size-Reorder Point Model

An (r,q) inventory policy, also known as a lot size-reorder point model (Nahmias [83], and Hadley and Whitin [51]), attempts to optimize a total cost function by continuously reviewing the inventory level in order to fulfill stochastic demands. In this system, r and q are independent decision variables. The operation of an (r,q) inventory policy is that: when the inventory position drops to a reorder point r, an order of fixed quantity q is placed (Federgruen and Zheng [39]). Nahmias [83] constructed an expected average annual cost function that includes fixed setup cost, inventory holding cost and shortage cost. His cost function used estimated average inventory as he found that the true average inventory is complicated to derive. He assumed demand is normally distributed and an iterative procedure is employed to solve for optimal values of r and q, starting with $q = \text{EOQ}$. It is similar to the model developed by Hadley and Whitin [51]. Moreover, Hadley and Whitin [51] developed an exact cost function for solving a case in which demand follows a Poisson distribution, lead-time is constant and all unfilled demands are backordered. A searching method is required for solving their model. Generally speaking, there is no '*reliable and straightforward method*' for solving an optimal (r,q) policy in a perfect manner (Browne and Zipkin [20]). As a result, several algorithms are developed. Federgruen and Zheng [39] developed an algorithm for solving a model similar to that of Hadley and Whitin's [51]. Matheus and Gelders [77] considered an (r,q) inventory system that supports a number of customers whose demand pattern is compound Poisson. By varying the reorder point while keeping a constant order quantity, they could use the model to cope with different desired service levels. Moinzadeh and Nahmias [81] employed two (r,q) policies to handle a single product, from which one ordering strategy is for emergency purpose. On the other hand, Moinzadeh and Lee [80] considered a two shipments policy in which an order q is partially shipped in two different time units. Again, they used EOQ as a lower bound of the order quantity and they formulated a searching algorithm for the reorder point r, given the value q. They considered both Poisson and normal demands. For more dynamic systems, Badinelli [8], Axsäter [7] and Ng et al. [84] proposed some (r,q) policies that could be applied to an inventory system with more than one levels or facilities. Axsäter's model [7] assumed that the inventory positions at all retailers are uniformly distributed, while Ng et al.'s model [84] assumed that they followed a Poisson distribution.

1.4.3 Markov Chain

Browne and Zipkin [20] developed an (r,q) policy in which the demand is a time-homogeneous Markov process. They designed an algorithm to evaluate the policy.

In their model, they assumed the inventory position is uniformly distributed in the interval $(r, r+q)$. Parlar and Perry's Markovian model is used to tackle supply uncertainty [92]. They checked the availability of the supply before placing an order of quantity q when inventory level drops to a reorder point r. If there is enough stock from the supplier, replenishment is made in zero lead-time. This state is called an "ON" state. If the stock is not available, then the order will arrive after T periods. It is called an "OFF" state. They tried to find an optimal (r, q, T) policy for minimizing a long-run average cost function based on the renewal reward theorem. Melchiors [79] applied Markov chain in comparing two can-order policies suggested by two other authors. He considered twelve products whose demands are Poisson and used simulation in comparison. Despite of coping with inventory management problems, Markov chain analysis is widely used in modelling machine breakdowns (Abboud [1]), soil conditions testing (Taha [111]), accounts receivable system (Render et al. [98]) and so on.

1.5 Concluding Remarks

Undoubtedly, manufacturing global products with customization for local markets plays an important role in striving for a distinctive competitive advantage for players in a global supply chain. Postponement enables companies to achieve higher manufacturing flexibility and product quality at lower costs. Pull postponement makes use of customer demand information to determine the push–pull boundary in a supply chain that allows common processes to be completed before those that differentiate product characteristics. Logistics postponement aims at deciding whether pull postponement should be carried out at local facilities instead of in the central production line. Form postponement is an enabling strategy for pull postponement as it opts for the use of standardized components and processes to achieve customization. Price postponement is an economic strategy that resorts to postponing the setting of product price. Practically, they can be combined and applied simultaneously to achieve optimization in a supply chain.

As mentioned, postponement is not a panacea to all situations. Examples cited in this chapter are on a situational basis. As a matter of fact, more generalized models and frameworks need to be developed. They can offer better insights and supportive evidence for postponement implementation in different areas. In this book, four types of model are presented to evaluate the impacts of pull and form postponement strategies under various supply chain structures. First, we develop two EOQ-based models to examine the impact of pull postponement in Chapter 2. Then we develop some EPQ-based models to examine the impact of postponement in Chapter 3. In Chapter 4 we propose a stochastic model of a single end-product supply chain that consists of a supplier, a manufacturer and a number of customers. In Chapter 5 we aim at conducting a simulation experiment of a two-end-product supply chain, for which customer demands are discrete and independent. Besides mathematical models, two case studies from industry are presented to support our theoretical results in Chapter 6. In Chapter 7 we conclude the book and suggest some worthy topics for future research.

Chapter 2
Analysis of Pull Postponement by EOQ-based Models

A number of quantitative models for analyzing postponement based upon cost and time evaluation have been discussed in the literature. Most of them assumed that the product demand is uncertain. However, if the demand is deterministic, e.g., because there is a long-term supply contract between the manufacturer and the retailers, the benefits due to economies of scope and risk pooling do not exit. Thus, evaluation of postponement structures under scenarios with deterministic demand is also an important issue.

It is natural that the economic order quantity (EOQ) model can be used to derive a total cost function for analyzing postponement. In this chapter we develop an EOQ-based model to examine the cost impact of the pull postponement strategy adopted by a supply chain that orders and keeps n end-products. We formulate a total average cost function for ordering and keeping the n end-products in a supply chain, in which their demands are known and deterministic. Using standard optimization techniques, we show that postponed customization of end-products will result in a lower total average cost and a lower EOQ. Furthermore, we develop an EOQ-based model with perishable items to evaluate the impact of item deterioration rate on inventory replenishment policies. Our theoretical analysis and computational results show that a postponement strategy for perishable items can yield a lower total average cost under certain circumstances.

This chapter is organized as follows. In Section 2.1 postponement strategy for ordinary (imperishable) items are discussed. In Section 2.2 postponement strategy for perishable items are addressed. We conclude the chapter in Section 2.3.

2.1 Postponement Strategy for Ordinary (Imperishable) Items

2.1.1 Proposed Model and Assumptions

In order to examine the effects of pull postponement on the total average cost and EOQ, we formulate two models to describe a supply chain.[1] It is assumed that the

[1] The following discussion in this section is largely based on the ideas and results presented in Wan [125].

T.C.E. Cheng et al., *Postponement Strategies in Supply Chain Management*,
International Series in Operations Research & Management Science 143,
DOI 10.1007/978-1-4419-5837-2_2,

chain supplies n ordinary (imperishable) end-products that are manufactured from the same type of raw material, say, plastics. The end-products belong to the same product category, but they have slight differences, say, color or size. In the first model, their ordering decisions are independent of one another, so there are n EOQ decisions. We called it an independent system. A schematic diagram is shown in Fig. 2.1.

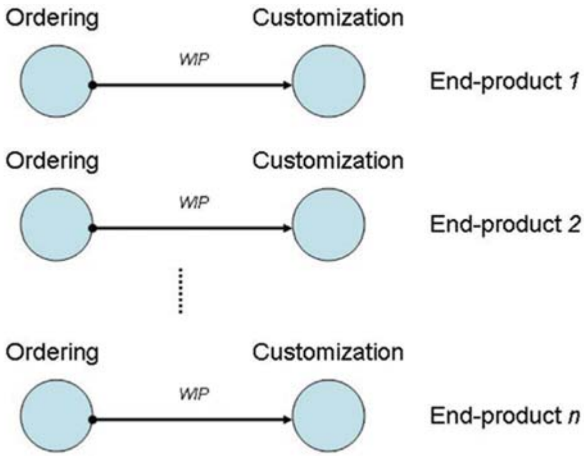

Fig. 2.1 A schematic diagram of the independent system

However, if customization can be postponed after ordering, then the ordering decisions can be combined so that a single EOQ decision is made. This is considered in the second model, which can be viewed as a special case of pull postponement strategy enabled by form postponement. In this model, the point of product differentiation is in the ordering process. A schematic diagram of this postponement system is shown in Fig. 2.2.

Our objective is to apply the EOQ model to investigate whether the combined system is more cost-effective than the independent system by comparing their total average cost functions and EOQs. In addition, we consider two cases of backordering cost. In the first case, we assume the planned backorder cost is the same for all end-products. In the second case, we assume different backorder costs for different end-products so as to generalize our model. In sum, we wish to test the following four hypotheses in this section.

H1: Postponement leads to a lower total average cost for a supply chain with n end-products in which there are planned backorders, where the backorder costs are identical for all end-products.

H2: Postponement leads to a lower economic order quantity (EOQ) for a supply chain with n end-products in which there are planned backorders, where the backorder costs are identical for all end-products.

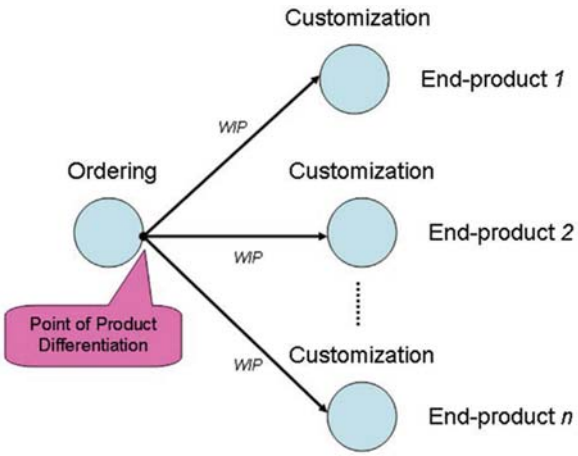

Fig. 2.2 A schematic diagram of the postponement system

H3: Postponement leads to a lower total average cost for a supply chain with n end-products in which there are planned backorders, where the backorder costs are not the same.

H4: Postponement leads to a lower economic order quantity (EOQ) for a supply chain with n end-products in which there are planned backorders, where the backorder costs are not the same.

We adopt Zipkin's EOQ model and his notation throughout this chapter (Zipkin [131]). Definitions of the notation and the general assumptions of this chapter are presented below.

2.1.1.1 Notation

- i = end-product ($i = 1, 2, \cdots, n$),
- λ_i = demand rate for end-product i, $\lambda_i > 0$,
- c = common variable cost, $c > 0$,
- k = common fixed ordering cost, $k > 0$,
- h = common unit holding cost per unit time, $h > 0$,
- b_i = unit backorder cost for end-product i, $b_i > 0$,
- r_i = reorder point for end-product i, $r_i \geq 0$,
- q_i = order quantity for end-product i, $q_i > 0$,
- q_i^* = economic order quantity (EOQ) for end-product i, $q_i^* > 0$,
- D_i = demand during lead-time for end-product i, $D_i > 0$,
- L_i = total cycle time for end-product i, $L_i > 0$,
- L_i' = backorder lead-time for end-product i, $L_i' \geq 0$,
- $v_i = r_i - D_i$, planned backorder quantity for end-product i, $v_i \leq 0$,
- $C(q_i)$ = total average cost for ordering and keeping end-product i with reorder quantity q_i, $C(q_i) > 0$,

- $C(q_i^*)$ = total average cost for ordering and keeping end-product i with EOQ q_i^*, $C(q_i^*) > 0$,
- TC = total average cost for ordering and keeping end-product 1 to end-product n with order quantities $q_1, q_2, \cdots, q_n$, $TC > 0$,
- TC^* = total average cost for ordering and keeping end-product 1 to end-product n with EOQs $q_1^*, q_2^*, \cdots, q_n^*$, $TC^* > 0$,
- TCP = total average cost per unit time for ordering and keeping end-products $1, 2, \cdots, n$ in a postponement system,
- TCP^* = the optimal total average cost per unit time for ordering and keeping end-products $1, 2, \cdots, n$ in a postponement system.

2.1.1.2 General Assumptions

(i) The number of end-products is $n \geq 2$.
(ii) Orders arrive without delay.
(iii) Each order quantity (q_i) is of the same size for end-product i.
(iv) All end-products are produced from the same type of raw material.
(v) The common variable cost (c), fixed ordering cost (k) and holding cost (h) for all end-products are the same.
(vi) The lead-time and the cost for the customization process is negligible.

2.1.2 Case 1: Same Backorder Cost

By the well-known EOQ formula, we obtain the total average cost for ordering and keeping end-product i as follows [131].

$$C(v_i, q_i) = c\lambda_i + \frac{k\lambda_i}{q_i} + \frac{h(q_i + v_i)^2}{2q_i} + \frac{bv_i^2}{2q_i}. \tag{2.1.1}$$

The first two terms are the variable and fixed components of the average ordering cost. The third term is the average inventory holding cost and the last term is the average backorder cost, where $v_i(\leq 0)$ is the planned backorder quantity for end-product i.

If all end-products are ordered independently (i.e., without postponement), the total average cost for ordering and keeping these n end-products is

$$\begin{aligned} TC &= \sum_{i=1}^{n} C(v_i, q_i) \\ &= c\sum_{i=1}^{n}\lambda_i + k\sum_{i=1}^{n}\frac{\lambda_i}{q_i} + \frac{h}{2}\sum_{i=1}^{n}\frac{(q_i + v_i)^2}{q_i} + \frac{b}{2}\sum_{i=1}^{n}\frac{v_i^2}{q_i}. \end{aligned} \tag{2.1.2}$$

A graphical explanation is illustrated in Fig. 2.3.

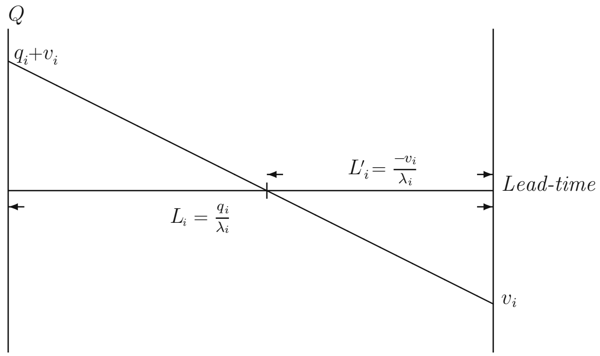

Fig. 2.3 Demand over lead-time in a cycle for end-product i without postponement

Minimizing *TC* in (2.1.2), we obtain the EOQ and optimal backorder quantity for end-product *i*, respectively, as follows [131].

$$q_i^* = \sqrt{\frac{2k\lambda_i}{h}}\sqrt{\frac{1}{\omega}}, \tag{2.1.3}$$

and

$$v_i^* = -(1-\omega)q_i^*, \tag{2.1.4}$$

where $\omega = \frac{b}{b+h}$.

The optimal total average cost is given by

$$\begin{aligned} TC^* &= \sum_{i=1}^{n} C(v_i^*, q_i^*) \\ &= c\sum_{i=1}^{n}\lambda_i + \sum_{i=1}^{n}\sqrt{2kh\omega\lambda_i}. \end{aligned} \tag{2.1.5}$$

However, if ordered jointly in quantity $(q_1 + q_2 + \cdots + q_n)$ in a postponement system (i.e., postponing the customization process), then the total average cost for ordering and keeping these n end-products becomes

$$\begin{aligned} TCP &= C(v_1+v_2+\cdots+v_n, q_1+q_2+\cdots+q_n) \\ &= c\sum_{i=1}^{n}\lambda_i + k\left[\frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{n}q_i}\right] + \frac{h}{2}\left[\frac{[\sum_{i=1}^{n}(q_i+v_i)]^2}{\sum_{i=1}^{n}q_i}\right] + \frac{b}{2}\left[\frac{(\sum_{i=1}^{n}v_i)^2}{\sum_{i=1}^{n}q_i}\right]. \end{aligned} \tag{2.1.6}$$

A graphical explanation is illustrated in Fig. 2.4.

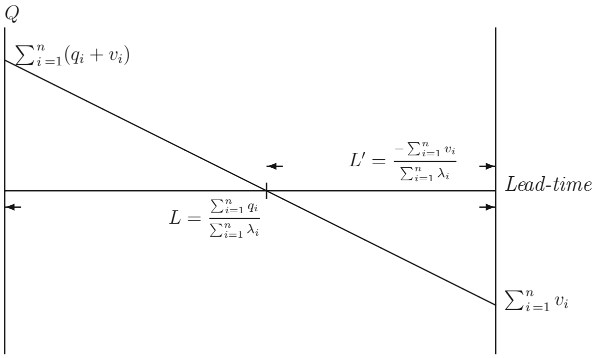

Fig. 2.4 Demand over lead-time in a cycle for joint ordering n end-products with the same backorder cost

Thus, the difference in the total average cost of the postponement system and independent systems is Z, given by (2.1.6)−(2.1.2), as follows.

$$Z = k\left[\frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{n}q_i} - \sum_{i=1}^{n}\frac{\lambda_i}{q_i}\right] + \frac{h}{2}\left[\frac{[\sum_{i=1}^{n}(q_i+v_i)]^2}{\sum_{i=1}^{n}q_i} - \sum_{i=1}^{n}\frac{(q_i+v_i)^2}{q_i}\right] + \frac{b}{2}\left[\frac{(\sum_{i=1}^{n}v_i)^2}{\sum_{i=1}^{n}q_i} - \sum_{i=1}^{n}\frac{v_i^2}{q_i}\right].$$

The first and the second terms represent, respectively, the differences in the average ordering cost and average inventory holding cost, while the last term is the difference in the average backorder cost between the two systems. We will show that $Z < 0$ so as to prove that postponement will be a better strategy to adopt.

We let the first term be Z_1, i.e.,

$$\begin{aligned} Z_1 &= k\left[\frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{n}q_i} - \sum_{i=1}^{n}\frac{\lambda_i}{q_i}\right] \\ &= k\left[\lambda_1\left(\frac{1}{\sum_{i=1}^{n}q_i} - \frac{1}{q_1}\right) + \lambda_2\left(\frac{1}{\sum_{i=1}^{n}q_i} - \frac{1}{q_2}\right) + \cdots + \lambda_n\left(\frac{1}{\sum_{i=1}^{n}q_i} - \frac{1}{q_n}\right)\right] \\ &< 0 \;\; (i \geq 2, k > 0, q_1, q_2, \cdots, q_n > 0 \text{ and } \lambda_1, \lambda_2, \cdots, \lambda_n > 0). \end{aligned}$$

Let the second term be Z_2, i.e.,

$$Z_2 = \frac{h}{2}\left[\frac{[\sum_{i=1}^{n}(q_i+v_i)]^2}{\sum_{i=1}^{n}q_i} - \sum_{i=1}^{n}\frac{(q_i+v_i)^2}{q_i}\right].$$

Further, let $a_i = \frac{q_i+v_i}{\sqrt{q_i}}, b_i = \sqrt{q_i}$. By the Cauchy-Schwarz inequality,

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \le \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right)$$
$$\left[\sum_{i=1}^{n} (q_i + v_i)\right]^2 \le \left[\sum_{i=1}^{n} \frac{(q_i + v_i)^2}{q_i}\right]\left(\sum_{i=1}^{n} q_i\right)$$
$$\frac{[\sum_{i=1}^{n} (q_i + v_i)]^2}{\sum_{i=1}^{n} q_i} \le \sum_{i=1}^{n} \frac{(q_i + v_i)^2}{q_i}.$$

Since $h > 0, Z_2 \le 0$.

Let the last term be Z_3, i.e.,

$$Z_3 = \frac{b}{2}\left[\frac{(\sum_{i=1}^{n} v_i)^2}{\sum_{i=1}^{n} q_i} - \sum_{i=1}^{n} \frac{v_i^2}{q_i}\right].$$

Further, let $a_i = \frac{v_i}{\sqrt{q_i}}, b_i = \sqrt{q_i}$. By the Cauchy-Schwarz inequality,

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \le \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right)$$
$$\left(\sum_{i=1}^{n} v_i\right)^2 \le \left(\sum_{i=1}^{n} \frac{v_i^2}{q_i}\right)\left(\sum_{i=1}^{n} q_i\right)$$
$$\frac{(\sum_{i=1}^{n} v_i)^2}{\sum_{i=1}^{n} q_i} \le \sum_{i=1}^{n} \frac{v_i^2}{q_i}.$$

Since $b > 0, Z_3 \le 0$. Hence, $Z < 0$. It follows that the pull postponement strategy will result in a lower total average cost. Furthermore, by letting $Q = \sum_{i=1}^{n} q_i$ and substituting it into (2.1.3), we obtain the optimal order quantity for TCP of (2.1.6) as

$$Q^* = \sqrt{\frac{2k \sum_{i=1}^{n} \lambda_i}{h}}\sqrt{\frac{1}{\omega}}. \tag{2.1.7}$$

In other words,

$$\begin{aligned}
[(q_1 + q_2 + \cdots + q_n)^*]^2 &= \frac{2k \sum_{i=1}^{n} \lambda_i}{h\omega} \\
&= \frac{2k\lambda_1}{h\omega} + \frac{2k\lambda_2}{h\omega} + \cdots + \frac{2k\lambda_n}{h\omega} \\
&= (q_1^*)^2 + (q_2^*)^2 + \cdots + (q_n^*)^2 \\
&< (q_1^* + q_2^* + \cdots + q_n^*)^2.
\end{aligned}$$

It is interesting to note that the relationship of the economic order quantity between a postponement system and a non-postponement system follows the Pythagoras' Theorem. Further, by letting $V = \sum_{i=1}^{n} v_i$ and substituting it into (2.1.4), $V^* = -(1-\omega)Q^*$. By substituting V^* and Q^* (from (2.1.7)) into (2.1.6), the optimal total average cost is

$$\begin{aligned} TCP^* &= C(V^*, Q^*) \\ &= c\sum_{i=1}^{n}\lambda_i + \sqrt{2kh\omega\sum_{i=1}^{n}\lambda_i}. \end{aligned} \tag{2.1.8}$$

The difference in the optimal total average cost between the postponement system and the independent system, given by (2.1.8)−(2.1.5), is as follows

$$\begin{aligned} Z^* &= C(V^*, Q^*) - \sum_{i=1}^{n} C(v_i^*, q_i^*) \\ &= \sqrt{2kh\omega\sum_{i=1}^{n}\lambda_i} - \sum_{i=1}^{n}\sqrt{2kh\omega\lambda_i}. \end{aligned}$$

Since $\sum_{i=1}^{n}\lambda_i - (\sum_{i=1}^{n}\sqrt{\lambda_i})^2 \le 0$, we have $\sqrt{\sum_{i=1}^{n}\lambda_i} - \sum_{i=1}^{n}\sqrt{\lambda_i} \le 0$ and $Z^* \le 0$. It implies that the total average cost can be further lowered by using a smaller EOQ for the same set of variables. So, the hypotheses *H1* and *H2* are supported.

2.1.3 Case 2: Different Backorder Costs

By the EOQ formula, we obtain the total average cost for ordering and keeping end-product i as follows [131].

$$C(v_i, q_i) = c\lambda_i + \frac{k\lambda_i}{q_i} + \frac{h(q_i + v_i)^2}{2q_i} + \frac{b_i v_i^2}{2q_i}. \tag{2.1.9}$$

The EOQ, optimal backorder quantity and optimal total average cost for ordering and keeping end-product i are respectively [131]

$$\begin{aligned} q_i^* &= \sqrt{\frac{2k\lambda_i}{h}}\sqrt{\frac{1}{\omega_i}}, \\ v_i^* &= -(1-\omega_i)q_i^*, \end{aligned}$$

and

$$C(v_i^*, q_i^*) = c\lambda_i + \sqrt{2kh\lambda_i\omega_i},$$

where $\omega_i = \frac{b_i}{b_i+h}$.

If all ordering decisions are considered independently (i.e., without postponement), the total average cost is

$$\begin{aligned} TC &= \sum_{i=1}^{n} C(v_i, q_i) \\ &= c\sum_{i=1}^{n}\lambda_i + k\left(\sum_{i=1}^{n}\frac{\lambda_i}{q_i}\right) + \frac{h}{2}\left[\sum_{i=1}^{n}\frac{(q_i+v_i)^2}{q_i}\right] + \sum_{i=1}^{n}\frac{b_i v_i^2}{2q_i}. \end{aligned} \quad (2.1.10)$$

Minimizing TC in (2.1.10), we obtain the optimal total average cost as follows.

$$\begin{aligned} TC^* &= \sum_{i=1}^{n} C(v_i^*, q_i^*) \\ &= c\sum_{i=1}^{n}\lambda_i + \sum_{i=1}^{n}\sqrt{2kh\lambda_i\omega_i}. \end{aligned} \quad (2.1.11)$$

The total average cost for ordering and keeping these n end-products with postponement is

$$\begin{aligned} TCP &= C(v_1+v_2+\cdots+v_n, q_1+q_2+\cdots+q_n) \\ &= c\sum_{i=1}^{n}\lambda_i + k\left(\frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{n}q_i}\right) + \frac{h}{2}\left[\frac{[\sum_{i=1}^{n}(q_i+v_i)]^2}{\sum_{i=1}^{n}q_i}\right] + \sum_{i=1}^{n} b_i\bar{B}_i, \end{aligned} \quad (2.1.12)$$

where $\bar{B}_i$ is the average backorder quantity for end-product i (see Fig. 2.5 for graphical explanation).

Since

$$\begin{aligned} b_i\bar{B}_i &= b_i\left(\frac{-\frac{1}{2}v_i L'}{L}\right) \\ &= \frac{\frac{b_i v_i}{2}\left(\frac{\sum_{i=1}^{n}v_i}{\sum_{i=1}^{n}\lambda_i}\right)}{\left(\frac{\sum_{i=1}^{n}q_i}{\sum_{i=1}^{n}\lambda_i}\right)} \end{aligned}$$

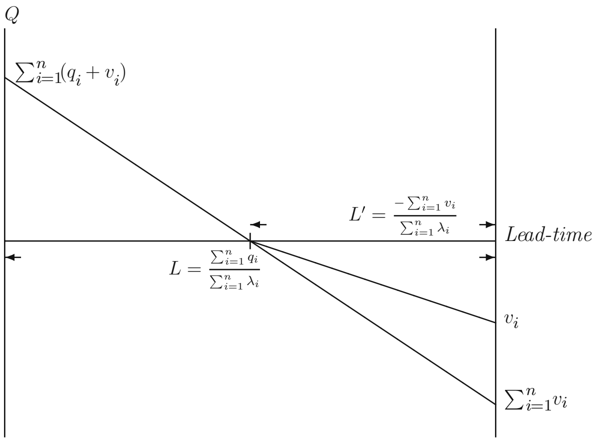

Fig. 2.5 Demand over lead-time in a cycle for joint ordering n end-products with different back-order costs

$$= \frac{b_i v_i (\sum_{i=1}^{n} v_i)}{2\sum_{i=1}^{n} q_i},$$

we have,

$$\sum_{i=1}^{n} b_i \bar{B}_i = \sum_{i=1}^{n} \frac{b_i v_i (\sum_{i=1}^{n} v_i)}{2\sum_{i=1}^{n} q_i}$$
$$= \frac{(\sum_{i=1}^{n} b_i v_i)(\sum_{i=1}^{n} v_i)}{2\sum_{i=1}^{n} q_i}.$$

Therefore, (2.1.12) can be expressed as

$$TCP = c\sum_{i=1}^{n} \lambda_i + k\left(\frac{\sum_{i=1}^{n} \lambda_i}{\sum_{i=1}^{n} q_i}\right) + \frac{h}{2}\left[\frac{[\sum_{i=1}^{n}(q_i+v_i)]^2}{\sum_{i=1}^{n} q_i}\right] + \frac{1}{2}\left[\frac{(\sum_{i=1}^{n} b_i v_i)(\sum_{i=1}^{n} v_i)}{\sum_{i=1}^{n} q_i}\right]. \tag{2.1.13}$$

If $b_1 = b_2 = \cdots = b_n = b$, then (2.1.13) reduces to (2.1.6).

The difference in the total average cost between the two systems is Z, given by (2.1.13)−(2.1.10), as follows.

$$Z = k\left(\frac{\sum_{i=1}^{n} \lambda_i}{\sum_{i=1}^{n} q_i} - \sum_{i=1}^{n} \frac{\lambda_i}{q_i}\right) + \frac{h}{2}\left[\frac{[\sum_{i=1}^{n}(q_i+v_i)]^2}{\sum_{i=1}^{n} q_i} - \sum_{i=1}^{n} \frac{(q_i+v_i)^2}{q_i}\right]$$

$$+\frac{1}{2}\left[\frac{(\sum_{i=1}^{n} b_i v_i)(\sum_{i=1}^{n} v_i)}{\sum_{i=1}^{n} q_i} - \sum_{i=1}^{n} \frac{b_i v_i^2}{q_i}\right]. \tag{2.1.14}$$

The first two terms are Z_1 and Z_2 in Case 1, respectively, which have been shown to be less than or equal to zero. We let the last term, which represents the difference in the average backorder cost between the two systems, be Z_4. By numerical trials, Z_4 can be positive or negative. For example, if $i = 2, b_1 = 10, b_2 = 5, v_1 = -5, v_2 = -6, q_1 = 12, q_2 = 10$, then $Z_4 = -0.583 < 0$. If $i = 2, b_1 = 1, b_2 = 5, v_1 = -5, v_2 = -6, q_1 = 12, q_2 = 10$, then $Z_4 = 0.208 > 0$. In other words, postponement will result in a lower total average cost only if $Z = Z_1 + Z_2 + Z_4 < 0$, and vice versa. It implies that the pull postponement strategy is advantageous when (i) the backorder costs are identical for all end-products (Case 1), (ii) there is a saving in the average backorder cost, $Z_4 \leq 0$, or (iii) the average ordering and inventory holding cost saving is greater than the average backorder cost, i.e., $Z_1 + Z_2 < -Z_4$. Provided that postponement does not lead to a lower total average cost, we wish to compare their optimal total average costs. Assuming that $q_1, q_2, \cdots, q_n$ and $v_1, v_2, \cdots, v_n$ are given, we can re-write (2.1.13) as

$$TCP = c\lambda + \frac{k\lambda}{Q} + \frac{h(Q+V)^2}{2Q} + \frac{V[(b_1 - b_n)v_1 + (b_2 - b_n)v_2 + \cdots + (b_{n-1} - b_n)v_{n-1} + b_n V]}{2Q},$$

where

(i) $\lambda = \sum_{i=1}^{n} \lambda_i$,
(ii) $Q = \sum_{i=1}^{n} q_n$,
(iii) $V = \sum_{i=1}^{n} v_n$,
(iv) v_n is re-written as $V - v_{n-1} - \cdots - v_1$.

Without loss of generality, we can assume that $b_1 \geq b_2 \geq \cdots \geq b_n$, so that $b_1 - b_n \geq 0, b_2 - b_n \geq 0, \cdots, b_{n-1} - b_n \geq 0$. Therefore, $(b_1 - b_n)v_1 V \geq 0, (b_2 - b_n)v_2 V \geq 0, \cdots, (b_{n-1} - b_n)v_{n-1} V \geq 0$. Thus, for any given Q and V, we can take $v_1 = v_2 = \cdots = v_{n-1} = 0$ (i.e., $V = v_n$) to obtain a minimum total average cost TCP. The total average cost equation is revised as

$$TCP = c\lambda + \frac{k\lambda}{Q} + \frac{h(Q+v_n)^2}{2Q} + \frac{b_n v_n^2}{2Q}. \tag{2.1.15}$$

By setting $\frac{\partial TCP}{\partial v_n} = 0$ and $\frac{\partial TCP}{\partial Q} = 0$, we can find the optimal v_n^* and Q^* that minimize TCP. The optimal solution is summarized below.

$$Q^* = \sqrt{\frac{2k\lambda}{h}}\sqrt{\frac{b_n + h}{b_n}}$$
$$= \sqrt{\frac{2k\lambda}{h}}\sqrt{\frac{1}{\omega_n}},$$
$$v_n^* = -\sqrt{\frac{2kh\lambda}{b_n(b_n + h)}},$$

and

$$TCP^* = c\lambda + \sqrt{2kh\lambda}\sqrt{\frac{b_n}{b_n + h}}$$
$$= c\lambda + \sqrt{2kh\lambda\omega_n}. \tag{2.1.16}$$

The difference between the optimal total average cost for the postponement system and that for the independent system is given by (2.1.16)−(2.1.11), i.e.,

$$Z^* = C[(v_1 + v_2 + \cdots + v_n)^*, (q_1 + q_2 + \cdots + q_n)^*] - \sum_{i=1}^{n} C(v_i^*, q_i^*)$$
$$= \sqrt{2kh\lambda\omega_n} - \sum_{i=1}^{n}\sqrt{2kh\lambda_i\omega_i}.$$

Letting $A = \sqrt{\lambda\omega_n}$ and $B = \sum_{i=1}^{n}\sqrt{\lambda_i\omega_i}$, we see that

$$A^2 - B^2 = (\lambda_1 + \lambda_2 + \cdots + \lambda_n)\omega_n - (\lambda_1\omega_1 + \lambda_2\omega_2 + \cdots + \lambda_n\omega_n)$$
$$= \lambda_1(\omega_n - \omega_1) + \lambda_2(\omega_n - \omega_2) + \cdots + \lambda_{n-1}(\omega_n - \omega_{n-1})$$
$$\leq 0\ (\omega_1 \geq \omega_2 \geq \cdots \geq \omega_n).$$

Since $A^2 - B^2 \leq 0$, we have $A - B \leq 0$ and $Z^* \leq 0$. The result indicates that the postponement system has a lower optimal average total cost. So, the hypotheses *H3* and *H4* are supported.

2.1.4 A Numerical Example

We give a numerical example to illustrate how postponement can yield savings in the total average cost in practice. In this example, we assume there are 5 end-products and the values of their various parameters are shown in the Table 2.1.

Table 2.1 Parameters of 5 end-products

i	λ_i	q_i	c	h	k	b_i	v_i
1	200	180	2	3	5	10	–5
2	80	100	2	3	5	8	–25
3	150	120	2	3	5	7	–30
4	100	70	2	3	5	6	–20
5	60	60	2	3	5	5	–10

From (2.1.10), the total average cost without postponement is equal to

$$\begin{aligned}TC &= 2(200+80+150+100+60) + 5\left(\frac{200}{180}+\frac{80}{100}+\frac{150}{120}+\frac{100}{70}+\frac{60}{60}\right)\\&+\frac{3}{2}\left(\frac{175^2}{180}+\frac{75^2}{100}+\frac{90^2}{120}+\frac{50^2}{70}+\frac{50^2}{60}\right)\\&+\left[\frac{(10)(-5)^2}{2(180)}+\frac{(8)(-25)^2}{2(100)}+\frac{(7)(-30)^2}{2(120)}+\frac{(6)(-20)^2}{2(70)}+\frac{(5)(-10)^2}{2(60)}\right]\\&= \$1838.11.\end{aligned}$$

From (2.1.13), the total average cost with postponement is equal to

$$\begin{aligned}TCP &= 2(200+80+150+100+60) + 5\left(\frac{200+80+150+100+60}{180+100+120+70+60}\right)\\&+\frac{3}{2}\left[\frac{(175+75+90+50+50)^2}{180+100+120+70+60}\right]\\&+\frac{(-90)[10(-5)+8(-25)+7(-30)+6(-20)+5(-10)]}{2(180+100+120+70+60)}\\&= \$1786.98.\end{aligned}$$

Thus, cost saving $= \frac{1838.11-1789.98}{1838.11} = 2.78\%$.

Applying the EOQ derived in the previous section, from (2.1.11), the optimal total average cost without postponement is equal to

$$\begin{aligned}TC^* &= 2(200+80+150+100+60) + \sqrt{6000\left(\frac{10}{10+3}\right)} + \sqrt{2400\left(\frac{8}{8+3}\right)}\\&+\sqrt{4500\left(\frac{7}{7+3}\right)} + \sqrt{3000\left(\frac{6}{6+3}\right)} + \sqrt{1800\left(\frac{5}{5+3}\right)}\\&= \$1424.10.\end{aligned}$$

From (2.1.16), the optimal total average cost with postponement is equal to

$$\begin{aligned} TCP^* &= 2(200+80+150+100+60)+\sqrt{30(200+80+150+100+60)\left(\frac{5}{5+3}\right)} \\ &= \$1285.18. \end{aligned}$$

Thus, cost saving $= \frac{1424.10-1285.18}{1424.1} = 9.76\%$.

2.2 Postponement Strategy for Perishable Items

The above models assumed that the inventoried items can be stored indefinitely to meet future demands.[2] However, certain types of products either deteriorate or become obsolete in the course of time. Perishable products are commonly found in commerce and industry, for example, fruits, fresh fish, perfumes, alcohol, gasoline, photographic films, etc. For these kinds of products, traditional inventory models are no longer applicable. An early study of perishable inventory systems was carried out by Whitin [126]. Since then, considerable effort has been expended on this line of research. Comprehensive surveys of related research can be found in Nahmias [82], Raafat [97], and Goyal and Giri [47], where relevant literature published before the 1980s, in the 1980s, and in the 1990s was reviewed, respectively. Recent studies before 2004 can be found in Song et al. [106].

There are many articles addressing perishable products with EOQ-based models. The analysis of perishable inventory problems began with Ghare and Schrader [45]. They developed a simple EOQ model to deal with products that experience exponential decay. Then there were many following studies that extended the basic model. For example, Covert and Phlilp [34] extended the model to consider item deterioration that follows the Weibull distribution. Other authors such as Tadikamalla [110], Shah [102], and Raafat [95, 96] discussed EOQ models under more general conditions.

One of the focuses of the research on perishable products is interaction and coordination in supply chains (Song et al. [106]). For example, Goyal and Gunasekaran [48] developed an integrated production- inventory-marketing model for determining the economic production quantity and economic order quantity for raw materials in a multi-stage production system. Yan and Cheng [127] studied a production-inventory model for perishable products, where they assumed that the production, demand and deterioration rate are all time-dependent. They gave the conditions for a feasible point to be optimal. Arcelus et al. [4] modeled a profit-maximizing retail promotion strategy for a retailer confronted with a vendor's trade promotion offer of credit and/or pricediscount on the purchase of regular or

[2] The following discussion in this section is largely based on the ideas and results presented in Li et al. [71].

perishable products. Kanchanasuntorn and Techanitisawad [61] investigated the effect of product deterioration and retailers' stockout policies on system total cost, net profit, service level, and average inventory level in a two-echelon inventory-distribution system, and developed an approximate inventory model to evaluate system performance. There are many papers addressing the interaction and coordination between inventory and marketing, financing, distribution, and production.

Motivated by the above observations, we will use the EOQ-based model with perishable items to analyze postponement in this section. The objective of this section is to investigate whether or not the postponement system can outperform the independent system with perishable items. We formulate two models to describe the supply chain, and give an algorithm to derive the optimal ordering strategies. We also investigate the effect of product deterioration on the total cost of the retailer and on inventory replenishment policies. Some numerical examples are provided to illustrate the theoretical results. We show that postponement strategy can give a lower total average cost under certain circumstances with perishable items. The results presented in this section provide insights for managers that guide them to find a proper tradeoff between postponement and non-postponement.

2.2.1 Notation and Assumptions

Consider a supply chain with n perishable products. These products are manufactured from the same type of raw materials and the end products only have slight differences. We assume that all the products and the raw materials decay at the same constant rate over time. The demand rates of the end-products are independent and constant. The unsatisfied demands (due to shortage) are completely backlogged. In an independent system, the end-products are ordered independently with different schedules so there are n EOQ decisions. However, if the customization process can be delayed after ordering, then the ordering decisions can be combined so that a single EOQ decision is made. This practice can be viewed as a form postponement strategy. Our objective is to apply the EOQ model with perishable products to examine the effects of form postponement on the total average cost.

2.2.1.1 Notation

- θ = deterioration rate of end-products and raw materials, $\theta \geq 0$,
- T_i = total cycle time for end-product i, $T_i > 0$,
- t_i = the time up to which the inventory of end-product i is positive in a cycle,
- $C(T_i, t_i)$ = total average cost per unit time for ordering and keeping end-product i,
- TC = total average cost per unit time for ordering and keeping end-products $1, 2, \cdots, n$ in an independent system,
- TC^* = the optimal total average cost per unit time for ordering and keeping end-products $1, 2, \cdots, n$ in an independent system,

- *TCP* = total average cost per unit time for ordering and keeping end-products $1, 2, \cdots, n$ in a postponement system,
- *TCP** = the optimal total average cost per unit time for ordering and keeping end-products $1, 2, \cdots, n$ in a postponement system.

2.2.1.2 Assumptions

1. The replenishment rate is infinite and the lead time is zero.
2. The end-product demand rates λ_i are constant and deterministic.
3. All the end-products are produced from the same type of raw materials and the factor of the raw material to end-product is 1:1.
4. Shortages are allowed and completely backlogged.
5. An extra customization process cost per end-product p is incurred if the customization process is delayed. The lead-time for customization is negligible.
6. The distribution of deterioration time of the items follows the exponential distribution with parameter θ, i.e., a constant rate of deterioration.
7. Deterioration of the materials and end-products is considered only after they have been received into inventory and there is no replacement of deteriorated inventory.

2.2.2 Model Formulation

Based on the above assumptions, the inventory level of an end-product at time t, $I(t)$, is governed by the following differential equation.

$$\frac{dI(t)}{dt} = \begin{cases} -\theta I(t) - \lambda, & 0 \le t \le t_0, \\ -\lambda, & t_0 \le t \le T. \end{cases} \tag{2.2.17}$$

with the boundary condition $I(t_0) = 0$, where t_0 is the time up to which the inventory level is positive in a cycle. The solution of (2.2.17) is

$$I(t) = \begin{cases} \lambda[e^{\theta(t_0 - t)} - 1]/\theta, & 0 \le t \le t_0, \\ -\lambda(t - t_0), & t_0 \le t \le T. \end{cases} \tag{2.2.18}$$

$I(t)$ follows the pattern depicted in Fig. 2.6.

Based on (2.2.18), we obtain the total average cost per unit time for ordering and keeping end-product as follows.

$$C(t_0, T|\, \theta) = \frac{k}{T} + c\lambda + \frac{\lambda(c\theta + h)(e^{\theta t_0} - \theta t_0 - 1)}{T\theta^2} + \frac{\lambda b(T - t_0)^2}{2T}. \tag{2.2.19}$$

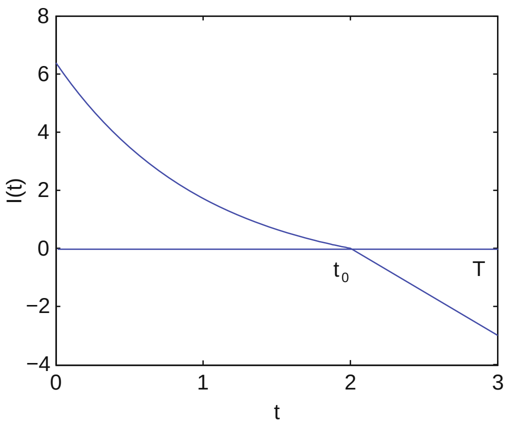

Fig. 2.6 Graphical representation of inventory level

The necessary conditions for the minimum value of $C(t_0, T|\,\theta)$ are

$$\frac{\partial C(t_0, T|\,\theta)}{\partial t_0} = \frac{\lambda(c\theta + h)(e^{\theta t_0} - 1)}{T\theta} + \lambda b\left(\frac{t_0}{T} - 1\right) = 0, \tag{2.2.20}$$

$$\frac{\partial C(t_0, T|\,\theta)}{\partial T} = -\frac{k}{T^2} - \frac{\lambda(c\theta + h)(e^{\theta t_0} - \theta t_0 - 1)}{T^2\theta^2} + \frac{\lambda b}{2}\left(1 - \frac{t_0^2}{T^2}\right) = 0 \tag{2.2.21}$$

After rearranging the terms in (2.2.20) and (2.2.21), we get

$$-\frac{k}{\lambda(c\theta + h)} - \frac{(e^{\theta t_0} - \theta t_0 - 1)}{\theta^2} + \frac{(c\theta + h)(e^{\theta t_0} - 1)^2}{2b\theta^2} + \frac{t_0(e^{\theta t_0} - 1)}{\theta} = 0, \tag{2.2.22}$$

$$T - t_0 - \frac{(c\theta + h)(e^{\theta t_0} - 1)}{b\theta} = 0. \tag{2.2.23}$$

Lemma 2.2.1 *If* $c\theta + h > 0$*, then the point* $(t_0^* > 0, T^* > 0)$ *that solves (2.2.22) and (2.2.23) simultaneously exists and is unique. The point* $(t_0^* > 0, T^* > 0)$ *is also the unique global optimum for the problem* $\min\{C(t_0, T|\theta){:}0 < t_0 < T < \infty\}$.

Proof Our lemma is a special case of Propositions 2 and 3 of Dye and Ouyang [36]. □

Thus t_0^* can be uniquely determined as a function of θ, say $t_0^* = t(\theta)$, and T^* can be uniquely determined as a function of θ, say $T^* = T(\theta)$. This also implies that $C(t_0^*, T^*|\,\theta)$ can be unique determined as a function of θ, say $C(t_0^*, T^*|\,\theta) = C(t(\theta), T(\theta)|\,\theta)$.

Theorem 2.2.2 $\tilde{C}(\theta) \stackrel{\text{def}}{=} C(t(\theta), T(\theta)|\,\theta) = \min_{0<t_0<T<\infty} C(t_0, T|\,\theta)$ *is an increasing and continuous function of* θ *in* $[0, +\infty)$*, and* $\lim\limits_{\theta\to 0} \tilde{C}(\theta) = c\lambda + \sqrt{\frac{2k\lambda hb}{b+h}}$.

Proof Recalling that the power series for e^x is $\sum_{n=0}^{\infty}\frac{(\theta t)^n}{n!}$, we have

$$\begin{aligned} C(t_0,T\,|\,\theta) &= \frac{k}{T} + c\lambda + \frac{\lambda(c\theta+h)\left(\sum_{n=0}^{\infty}\frac{(\theta t_0)^n}{n!} - \theta t_0 - 1\right)}{T\theta^2} + \frac{\lambda b(T-t_0)^2}{2T} \\ &= \frac{k}{T} + c\lambda + \frac{t_0^2\lambda(c\theta+h)\left(\sum_{n=2}^{\infty}\frac{(\theta t_0)^{n-2}}{n!}\right)}{T} + \frac{\lambda b(T-t_0)^2}{2T}. \end{aligned} \tag{2.2.24}$$

For $\theta \geq 0$, it is obvious that $C(t_0,T\,|\,\theta)$ is an increasing function of θ for each fixed value of $t_0 > 0$ and $T > 0$. If $\theta_1 < \theta_2$, we have

$$\begin{aligned} \tilde{C}(\theta_2) &= C(t(\theta_2),T(\theta_2)\,|\,\theta_2) \\ &> C(t(\theta_2),T(\theta_2)\,|\,\theta_1) \\ &\geq C(t(\theta_1),T(\theta_1)\,|\,\theta_1) \\ &= \tilde{C}(\theta_1). \end{aligned}$$

Thus, $\tilde{C}(\theta)$ is an increasing function of θ in $[0,+\infty)$.

$$\begin{aligned} f_1(t_0,T,\theta) &\overset{\text{def}}{=} -\frac{k}{\lambda(c\theta+h)} - \frac{(e^{\theta t_0} - \theta t_0 - 1)}{\theta^2} + \frac{(c\theta+h)(e^{\theta t_0}-1)^2}{2b\theta^2} \\ &\qquad + \frac{t_0(e^{\theta t_0}-1)}{\theta}, \\ f_2(t_0,T,\theta) &\overset{\text{def}}{=} T - t_0 - \frac{(c\theta+h)(e^{\theta t_0}-1)}{b\theta}. \end{aligned}$$

For $\theta > -\frac{h}{c}$, we have

$$\begin{aligned} \frac{\partial f_1}{\partial t_0} &= \frac{\lambda(c\theta+h)(e^{\theta t_0}-1)e^{\theta t_0}}{b\theta} + \lambda t_0 e^{\theta t_0}, & \frac{\partial f_1}{\partial T} &= 0, \\ \frac{\partial f_2}{\partial t_0} &= -1 - \frac{(c\theta+h)e^{\theta t_0}}{b}, & \frac{\partial f_2}{\partial T} &= 1. \end{aligned}$$

from where we deduce that

$$\begin{vmatrix} \frac{\partial f_1}{\partial t_0} & \frac{\partial f_1}{\partial T} \\ \frac{\partial f_2}{\partial t_0} & \frac{\partial f_2}{\partial T} \end{vmatrix} = \frac{\partial f_1}{\partial t_0} = \frac{\lambda(c\theta+h)(e^{\theta t_0}-1)e^{\theta t_0}}{b\theta} + \lambda t_0 e^{\theta t_0} > 0.$$

From the implicit function theorem, we know that $t(\theta)$ and $T(\theta)$ are continuous functions of θ in $[0,+\infty)$, respectively. Moreover, $C(t_0,T|\theta)$ is a continuously differentiable real function for $0 < t_0 < T$, and $\theta > -\frac{h}{c}$. Thus, $\tilde{C}(\theta)$ is also a continuous function of θ in $[0,+\infty)$.

Because $\tilde{C}(\theta)$ is continuous in $[0, +\infty)$, we have

$$\begin{aligned}\lim_{\theta\to 0}\tilde{C}(\theta) &= \tilde{C}(0)\\ &= \min_{0<t_0<T<\infty}\left\{\frac{k}{T}+c\lambda+\frac{\lambda h t_0^2}{2T}+\frac{\lambda b(T-t_0)^2}{2T}\right\}\\ &= c\lambda+\sqrt{\frac{2k\lambda hb}{b+h}}.\end{aligned}$$

□

Theorem 2.2.3 *$t(\theta)$ is a decreasing function of θ in $[0, +\infty)$, and $t(\theta) \le \sqrt{\frac{2kb}{\lambda h(b+h)}}$.*

Proof $t(\theta)$ is the unique solution of Eq. (2.2.20). After rearranging the terms in (2.2.20), we get

$$\begin{aligned}\frac{k}{\lambda(c\theta+h)} &= -\frac{(e^{\theta t_0}-\theta t_0-1)}{\theta^2}+\frac{t_0(e^{\theta t_0}-1)}{\theta}+\frac{(c\theta+h)(e^{\theta t_0}-1)^2}{2b\theta^2},\\ &= t_0^2\sum_{i=1}^{+\infty}\left(\frac{1}{i!}-\frac{1}{(i+1)!}\right)(\theta t_0)^{i-1}+\frac{(c\theta+h)(e^{\theta t_0}-1)^2}{2b\theta^2}.\end{aligned} \quad (2.2.25)$$

The left-hand side of (2.2.25) is a decreasing function of θ and the right side of (2.2.25) is an increasing function of θ for each fixed value of $t_0 > 0$. When θ increases, $t(\theta)$ has to decrease in order to satisfy equation (2.2.25). So $t(\theta)$ is a decreasing function of θ in $[0, +\infty)$ and $t(\theta) \le t(0) = \sqrt{\frac{2kb}{\lambda h(b+h)}}$. □

Because t_0^* and T^* cannot be determined in a closed form from (2.2.22) and (2.2.23), we have to determine them numerically using the following algorithm.

Algorithm 2.2.4 *Step 1: Obtain the value of t_0^* by solving the nonlinear Eq. (2.2.22) with the help of some mathematical software such as MatLab or Mathematica.*

Step 2: Compute T^* by using (2.2.23).

Step 3: The corresponding optimal cost per unit time $C(t_0^*, T^* \mid \theta)$ can be obtained by (2.2.19).

Remark 2.2.5 If $\theta\sqrt{\frac{2kb}{\lambda h(b+h)}}$ is small enough, we can give an approximate optimal solution of (2.2.19). We can approximate $e^{\theta t_0}$ by the first three terms in its power series. Then, we have $C(t_0, T|\theta) \approx \frac{k}{T} + c\lambda + \frac{\lambda(c\theta+h)(\sum_{n=0}^{2}\frac{(\theta t)^n}{n!}-\theta t_0-1)}{T\theta^2} + \frac{\lambda b(T-t_0)^2}{2T} = \frac{k}{T} + c\lambda + \frac{t_0^2\lambda(c\theta+h)}{2T} + \frac{\lambda b(T-t_0)^2}{2T}$. This is the classic EOQ model. By the EOQ formula, we can obtain the approximate optimal cost is $c\lambda + \sqrt{\frac{2k\lambda(h+c\theta)b}{b+h+c\theta}}$. From the approximate optimal cost, we can find that deterioration effectively adds an additional component to the holding cost, from h to $h + c\theta$.

2.2.3 The Postponement and Independent Systems

Now we discuss the postponement system and the independent system. In the independent system all the end-products are ordered independently (i.e., without postponement). The total average cost for ordering and keeping the n end-products is

$$
\begin{aligned}
TC(\theta) &= \sum_{i=1}^{n} C(t_i, T_i|\theta) \\
&= \sum_{i=1}^{n} \left\{ \frac{k}{T_i} + c\lambda_i + \frac{\lambda_i(c\theta + h)(e^{\theta t_i} - \theta t_i - 1)}{T_i\theta^2} + \frac{\lambda_i b(T_i - t_i)^2}{2T_i} \right\}. \quad (2.2.26)
\end{aligned}
$$

In the postponement system, all the raw materials are ordered (i.e., postponing the customization process) and the demand rate is $\hat{\lambda} = \lambda_1 + \lambda_2 + \cdots + \lambda_n$. The total average cost for ordering and keeping the n end-products is given by (excluding the customization cost)

$$
TCP(\hat{t}, \hat{T}|\theta) = \frac{k}{\hat{T}} + c\hat{\lambda} + \frac{\hat{\lambda}(c\theta + h)(e^{\theta\hat{t}} - \theta\hat{t} - 1)}{\hat{T}\theta^2} + \frac{\hat{\lambda}b(\hat{T} - \hat{t})^2}{2\hat{T}}. \quad (2.2.27)
$$

The difference in the optimal total average cost per unit time of the two systems is defined as $Z^* = TCP^*(\theta) - TC^*(\theta)$.

Theorem 2.2.6 *There exists a $\bar{\theta} > 0$ such that for any $0 \le \theta \le \bar{\theta}$, $TC^*(\theta) > TCP^*(\theta)$, i.e., the postponement system can give a lower total average cost than the independent system.*

Proof Because $\tilde{C}(\theta)$ is continuous on $[0, +\infty)$, and

$$
\begin{aligned}
\lim_{\theta \to 0} TCP^*(\theta) &= TCP^*(0) \\
&= c\hat{\lambda} + \sqrt{\frac{2kh\hat{\lambda}b}{b+h}} \quad (2.2.28)
\end{aligned}
$$

$$
\begin{aligned}
\lim_{\theta \to 0} TC^*(\theta) &= TC^*(0) \\
&= c\hat{\lambda} + \sum_{i=1}^{n} \sqrt{\frac{2kh\lambda_i b}{b+h}} \quad (2.2.29)
\end{aligned}
$$

In Section 2.1 we have proved that (2.2.28) − (2.2.29) $<$ 0. So there exists a $\bar{\theta} > 0$ such that for any $0 \le \theta \le \bar{\theta}$, $TC^*(\theta) > TCP^*(\theta)$.

Remark 2.2.7 If θ is small, we can obtain an approximate optimal cost as $c\lambda + \sqrt{\frac{2k\lambda(h+c\theta)b}{b+h+c\theta}}$. Then $Z^* \approx \left(\sqrt{\sum_{i=1}^{n}\lambda_i} - \sum_{i=1}^{n}\sqrt{\lambda_i}\right)\sqrt{\frac{2k(h+c\theta)b}{b+h+c\theta}}$. From this equation, we see that the postponement system can outperform the independent system, and the absolute value of Z^* becomes larger when θ becomes larger.

2.2.4 Numerical Examples

We give some numerical examples to illustrate how the deterioration rate impacts on the optimal total average cost and postponement. To illustrate the results, we consider the example in Padmanabhan and Vrat [90].

Example 2.2.8 In order to study how various deterioration rates affect the optimal cost of the EOQ model, deterioration sensitivity analysis is performed. The value of the deterioration rate is changed=(0, 0.02, 0.04, 0.06, 0.08, 0.10, 0.20, 0.30, 0.40, 0.50, 0.60). The demand rate is $\lambda = 600$, the common variable ordering cost c is 5, the common fixed ordering cost k is 250, the common unit holding cost h per unit time is 1.75, and the unit backorder cost b is 3 (all in appropriate units).

Applying the solution procedure in Section 2.2.2, we derive the results shown in Table 2.2 and Fig. 2.7, from which the following observations can be made.

1. $\tilde{c}(\theta)$ is an increasing, continuous and concave function of θ in $[0, +\infty)$.
2. $t(\theta)$ is a decreasing function of θ in $[0, +\infty)$.
3. θt^* is less sensitive to θ. The reason is that $t(\theta)$ is a decreasing function of θ.

Table 2.2 The impact of deterioration rate on inventory replenishment policies

θ	0	0.02	0.04	0.06	0.08	0.10	0.20	0.30	0.40	0.5	0.6
t^*	0.548	0.526	0.505	0.486	0.468	0.4527	0.385	0.336	0.299	0.269	0.244
θt^*	0	0.011	0.020	0.029	0.037	0.045	0.077	0.1008	0.119	0.134	0.147
$\tilde{C}(\theta)$	3576	3587	3597	3606	3615	3625	3661	3690	3713	3733	3750

Example 2.2.9 In order to study how various deterioration rates affect the difference of the postponement system and the independent system, we assume that there are eleven end-products. For the eleven products, we assume that $\lambda_1 = 550$, $\lambda_2 = 560$, $\lambda_3 = 570$, $\lambda_4 = 580$, $\lambda_5 = 590$, $\lambda_6 = 600$, $\lambda_7 = 610$, $\lambda_8 = 620$, $\lambda_9 = 630$, $\lambda_{10} = 640$, $\lambda_{11} = 650$. The related other data are the same as the data of Example 2.2.8. Applying the solution procedure in Section 2.2.2, we obtain the results of the sensitivity analysis with these parameters, which are shown in Table 2.3 and Fig. 2.8, from which the following observations can be made.

1. The postponement system yields savings in the total average cost.

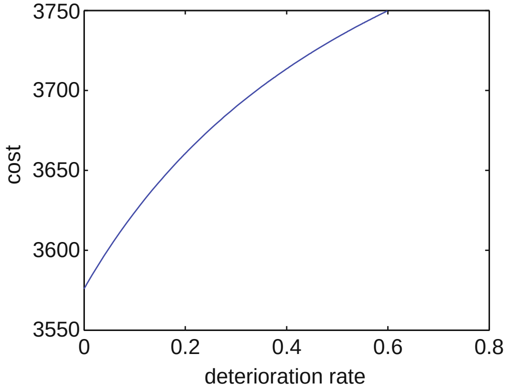

Fig. 2.7 The impact of deterioration rate on cost

Table 2.3 The impact of deterioration rate on the difference in cost

θ	0	0.02	0.04	0.06	0.08	0.10	0.20	0.30	0.40	0.5	0.6
z^*	–4422	–4506	–4585	–4659	–4730	–4796	–5083	–5309	–5493	–5646	–5775
$\left\lvert \frac{z^*}{TC^*} \right\rvert$	0.112	0.114	0.116	0.118	0.119	0.120	0.126	0.131	0.135	0.138	0.140

2. The absolute value of Z^* becomes larger when the deterioration rate becomes larger.
3. The absolute value of $\frac{Z^*}{TC^*}$ becomes larger when the deterioration rate becomes larger.

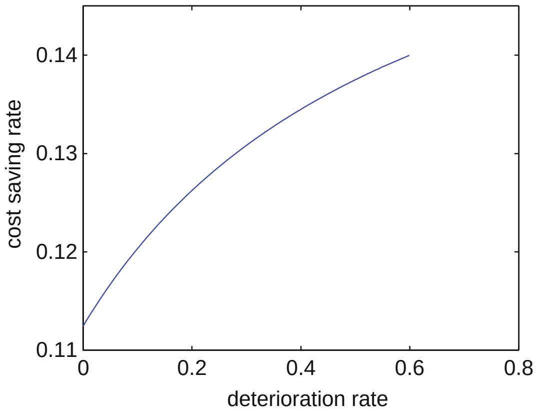

Fig. 2.8 The impact of deterioration rate on the difference in cost between the two systems

2.3 Concluding Remarks

This chapter seeks to evaluate the potential benefits of using pull postponement in ordering based on the EOQ model. The EOQ model is chosen for modelling not only because it is a foundation in inventory management, but also it is suitable for identifying the cost savings in an easy way. For imperishable products, the benefits are reductions in both total average cost and economic order quantity. There are savings even when there are planned backorders with different backorder costs. For perishable products with a constant deterioration rate θ, it has been shown that the postponement strategy outperforms the independent strategy when θ is small. Our numerical experiments show that the difference of the two strategy will become larger when θ becomes larger.

However, our model limits the use of the pull postponement strategy in ordering. We assume that the cost of customization is negligible and the ordering cost is fixed, regardless of order quantity. Our findings are still valid if the customization cost is less than the savings. Moreover, there are extra savings from a lower fixed ordering cost due to economies of scale and scope in joint ordering [131]. In addition, we assume that the deterioration rate of the raw materials is the same as that of the end-products. But the raw materials, such as IC, chips, are easy to transfer to other chassis by design change, the deterioration rate of the raw materials is often smaller than that of the end-products. So the postponement can yield more savings in the total average cost in practice.

Besides, the pull postponement strategy can be applied in a more generic supply chain, which consists of ordering, production, distribution, and so on. As an extension, the economic production quantity (EPQ) model can be used to demonstrate how a postponement strategy can be implemented when part of the production processes is standardized and common for a product category, while the remaining part is distinct for customization. Logistics postponement can then be considered to examine if localizing some customization processes can further reduce the total average cost and inventory level. Another research direction is to use stochastic models instead of deterministic models such as the EOQ and EPQ models since stochastic models are more flexible in dealing with realistic problems. An EPQ model is presented in Chapter 3, while two analyses, which are based on stochastic models, are discussed in Chapters 4 and 5.

Chapter 3
Analysis of Postponement Strategy by EPQ-based Models

In this chapter we develop EPQ-based models with and without stockout to examine the impact of postponement. We formulate the total average cost functions of the two scenarios for producing and keeping n end-products in a supply chain, in which their demands are known and deterministic. Using standard optimization techniques, we show that postponed customization of end-products results in a lower total average cost in certain circumstances. We also find that two key factors that influence postponement decisions are variance of the machine utilization rates and variance of the backorder costs.

This chapter is organized as follows. In Section 3.1 postponement strategy by an EPQ-based model without stockout are discussed. In Section 3.2 postponement strategy by an EPQ-based model with planned backorders are addressed. We conclude the chapter in Section 3.3.

3.1 Analysis of Postponement Strategy by an EPQ-based Model Without Stockout

3.1.1 Proposed Model and Assumptions

The major interest of this section are the effects of pull postponement on improving the total average cost per unit time and EPQ when stockouts are forbidden.[1] We develop a mathematical model to describe a supply chain with n end-products. The supply chain composes of a group of suppliers, a manufacturer and a group of customers. It is assumed that raw materials are shipped from the suppliers to the manufacturer to produce different kinds of end-products in-house. The end-products can be classified into different product categories. Our focus is on one of the product categories, in which there are n end-products for different customers from different market segments. The end-products only have slight differences, e.g., notebook computers bundled with different CPUs, RAMs or peripheral devices. From a man-

[1] The following discussion in this section is largely based on the ideas and results presented in Wan [125].

T.C.E. Cheng et al., *Postponement Strategies in Supply Chain Management*, International Series in Operations Research & Management Science 143,
DOI 10.1007/978-1-4419-5837-2_3,

ufacturing perspective, if they are produced independently with different production schedules, then there are n EPQ decisions. A schematic diagram of the independent system is shown in Fig. 3.1.

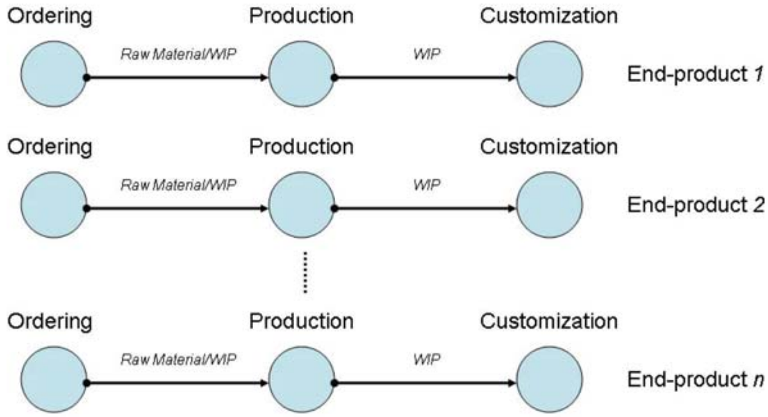

Fig. 3.1 A schematic diagram of the independent system

However, if the assembly process of CPUs, RAMs or peripheral devices can be delayed, then the core modules can be produced in aggregate such that there is only one EPQ decision. In other words, there is a point of product differentiation in the production stage. After the point of product differentiation, the assembly process can be carried out when orders are received. This practice can be viewed as a postponement strategy (see Huang and Lo [56] for an example of the Taiwan PC industry). A schematic diagram of the postponement system is shown in Fig. 3.2.

Our objective is to apply the EPQ model to examine whether this strategy is more cost-effective than the original one by comparing their total average cost per

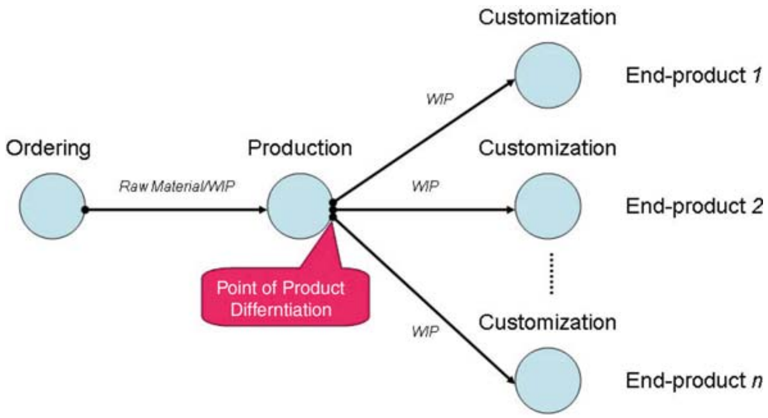

Fig. 3.2 A schematic diagram of the postponement system

unit time functions and EPQs. Our cost function consists of fixed production cost, variable production cost and inventory holding cost. Two cases are presented. In the first case, we assume customer demands are met continuously by the current production batch, while in the second case, we assume customer demands are only met after production is complete. Each of the two mathematical models begins with considering two end-products, and will be generalized to consider n end-products in later sections. In sum, we wish to investigate the following eight hypotheses in this section (Table 3.1).

Table 3.1 A summary of the eight hypotheses without stockout

No. of machines(Measure)	Economic Production Quantity	Optimal total average cost
2(a)	H1	H2
2(b)	H3	H4
n(a)	H5	H6
n(b)	H7	H8

(a) *Demands are met instantly by the current production batch*
(b) *Demands are met after production is complete*

H1: Postponement leads to a lower EPQ for a supply chain with 2 end-products, in which customer demands are met instantly by the current production batch.
H2: Postponement leads to a lower optimal total average cost per unit time for a supply chain with 2 end-products, in which customer demands are met instantly by the current production batch.
H3: Postponement leads to a lower EPQ for a supply chain with 2 end-products, in which customer demands are met after production is complete.
H4: Postponement leads to a lower optimal total average cost per unit time for a supply chain with 2 end-products, in which customer demands are met after production is complete.
H5: Postponement leads to a lower EPQ for a supply chain with n end-products, in which customer demands are met instantly by the current production batch.
H6: Postponement leads to a lower optimal total average cost per unit time for a supply chain with n end-products, in which customer demands are met instantly by the current production batch.
H7: Postponement leads to a lower EPQ for a supply chain with n end-products, in which customer demands are met after production is complete.
H8: Postponement leads to a lower optimal total average cost per unit time for a supply chain with n end-products, in which customer demands are met after production is complete.

We adopt Zipkin's EPQ model and his notation throughout this chapter (Zipkin [131]). Definitions of the notation and the general assumptions are presented below.

3.1.1.1 Notation

- i=end-product ($i = 1, 2, \cdots, n$),
- λ_i=demand rate for end-product i, $\lambda_i > 0$,

- μ_i = production rate for end-product i, $\mu_i > 0$,
- c = common variable production cost, $c > 0$,
- k = common fixed set-up cost, $k > 0$,
- h = common unit holding cost per unit time, $h > 0$,
- p = extra unit customization cost per unit time, $p \geq 0$,
- r_i = reorder point for end-product i, $r_i \geq 0$,
- q_i = production quantity for end-product i, $q_i > 0$,
- q_i^* = economic production quantity (EPQ) for end-product i, $q_i^* > 0$,
- $C(q_i)$ = total average cost per unit time for producing and keeping end-product i with production quantity q_i, $C(q_i) > 0$,
- $C(q_i^*)$ = total average cost per unit time for producing and keeping end-product i with EPQ q_i^*, $C(q_i^*) > 0$,
- TC = total average cost per unit time for producing and keeping end-products $1, \cdots, n$ with production quantities $q_1, q_2, \cdots, q_n$, respectively, $TC > 0$,
- TC^* = total average cost per unit time for producing and keeping end-products $1, \cdots, n$ with EPQ $q_1^*, q_2^*, \cdots, q_n^*$, respectively, $TC^* > 0$.
- IP = inventory position over time.

3.1.1.2 General Assumptions

(i) A cycle means the time between the production of two consecutive batches. It consists of two parts: active time and idle time (Zipkin [131]). When production starts, inventory accumulates until it is enough for the cycle. Then, production stops. Inventory starts to decline and finally drops to zero. Another cycle begins when inventory reaches zero.

(ii) Customer demand rate (λ_i) is constant and deterministic for end-product i.

(iii) Each production quantity (q_i) is of the same size for end-product i.

(iv) Inventory holding cost for raw material and work-in-process inventory is ignored (Zipkin [131]).

(v) Common variable production cost (c), fixed set-up cost (k), end-product holding cost (h) are the same for all end-products.

(vi) Extra customization process cost per end-product (p) is incurred if the customization process is delayed.

In Sections 3.1.2 and 3.1.3, the eight hypotheses are examined and discussed in detail.

3.1.2 2 Machines for 2 End-Products

In this section, we study whether or not a postponement system is better than a non-postponement system in terms of EPQ and optimal total average cost per unit time when there are 2 machines for 2 distinct end-products. Apart from the general assumptions presented in the above section, two further assumptions are adopted as follows.

Further Assumptions

(i) End-products 1 and 2 are produced separately by two machines.
(ii) For each end-product, the production rate is larger than the demand rate. That is, $\mu_1 > \lambda_1$ and $\mu_2 > \lambda_2$.

3.1.2.1 Demands Are Met Continuously

We first assume that the customer demands are met continuously by the current production batch. For example, if the customers are downstream stations of the manufacturer in a production line of a factory, the customer demands may be met continuously. Alternatively, if the product transportation time from the manufacturer to the customers is short and the transportation frequency is high, we can also assume that the demands are met continuously.

We first consider the case of independent processes, i.e., one machine for one end-product. By the well-known EPQ formula, we obtain the total average cost per unit time for producing and keeping end-product i as follows [131].

$$C(q_i) = c\lambda_i + \frac{k\lambda_i}{q_i} + \frac{h}{2}\left(1 - \frac{\lambda_i}{\mu_i}\right) q_i, \ i = 1,2. \tag{3.1.1}$$

The first two terms are the fixed (independent of production quantity) and variable (dependent of production quantity) components of the production cost per unit time, while the last term is the average inventory holding cost per unit time for end-product i. A graph of the inventory position over time is illustrated in Fig. 3.3.

If the production of end-products 1 and 2 is operated by two machines, in which the end-product characteristics are fully built, then the total average cost per unit time for producing and keeping the two end-products is:

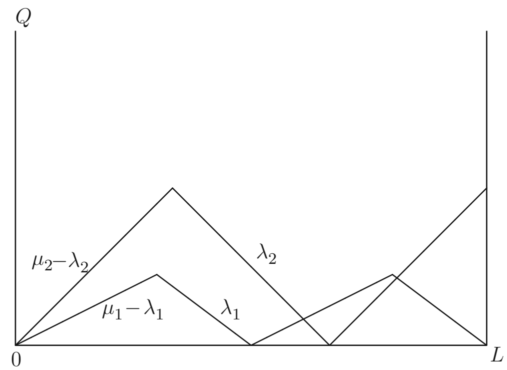

Fig. 3.3 IP for end-product 1 and 2 in the independent system when the demand is met continuously

$$\begin{aligned} TC &= C(q_1) + C(q_2) \\ &= c(\lambda_1 + \lambda_2) + k\left(\frac{\lambda_1}{q_1} + \frac{\lambda_2}{q_2}\right) + \frac{h}{2}\left[\left(1 - \frac{\lambda_1}{\mu_1}\right)q_1 + \left(1 - \frac{\lambda_2}{\mu_2}\right)q_2\right]. \end{aligned} \tag{3.1.2}$$

Minimizing TC in (3.1.2), we obtain the EPQ for end-product 1, the EPQ for end-product 2 and the optimal total average cost per unit time, respectively, as follows [131].

$$q_1^* = \sqrt{\frac{2k\lambda_1}{h\left(1 - \frac{\lambda_1}{\mu_1}\right)}},$$

$$q_2^* = \sqrt{\frac{2k\lambda_2}{h\left(1 - \frac{\lambda_2}{\mu_2}\right)}},$$

and

$$\begin{aligned} TC^* &= C(q_1^*) + C(q_2^*) \\ &= c(\lambda_1 + \lambda_2) + \sqrt{2kh\lambda_1\left(1 - \frac{\lambda_1}{\mu_1}\right)} + \sqrt{2kh\lambda_2\left(1 - \frac{\lambda_2}{\mu_2}\right)}. \end{aligned} \tag{3.1.3}$$

Now consider the situation where the core productions of the two end-products are combined, but slight customization is carried out later. We regard this strategy as a postponement strategy. One practical example is notebook computer production, in which common modules (without CPUs) are produced and CPUs are installed at the time when customer orders are received. In our model, we assume the combined core production is completed by a single machine whose production rate is $\mu_1 + \mu_2$, the variable production cost is c and the fixed set-up cost is k. Let the extra customization cost per end-product per unit time be p. The total average cost per unit time for producing (excluding the customization process) and keeping the two end-products becomes

$$\begin{aligned} TC &= C(q_1 + q_2) \\ &= c(\lambda_1 + \lambda_2) + k\left(\frac{\lambda_1 + \lambda_2}{q_1 + q_2}\right) + \frac{h}{2}\left(1 - \frac{\lambda_1 + \lambda_2}{\mu_1 + \mu_2}\right)(q_1 + q_2). \end{aligned} \tag{3.1.4}$$

A graphical illustration is given in Fig. 3.4.

Minimizing TC in (3.1.4), we obtain the EPQ and the optimal total average cost per unit time, respectively, as follows

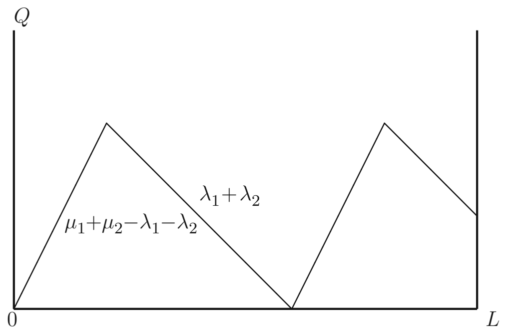

Fig. 3.4 IP for end-product 1 and 2 in the postponement system when the demand is met continuously

$$(q_1+q_2)^* = \sqrt{\frac{2k(\lambda_1+\lambda_2)}{h\left(1-\dfrac{\lambda_1+\lambda_2}{\mu_1+\mu_2}\right)}},$$

and

$$\begin{aligned} TC^* &= C[(q_1+q_2)^*] \\ &= c(\lambda_1+\lambda_2) + \sqrt{2kh(\lambda_1+\lambda_2)\left(1-\frac{\lambda_1+\lambda_2}{\mu_1+\mu_2}\right)}. \end{aligned} \tag{3.1.5}$$

Analysis of EPQs for the Two Systems

We let

$$\begin{aligned} G &= \frac{(\lambda_1+\lambda_2)}{1-\dfrac{\lambda_1+\lambda_2}{\mu_1+\mu_2}} \\ &= \frac{(\lambda_1+\lambda_2)(\mu_1+\mu_2)}{(\mu_1-\lambda_1)+(\mu_2-\lambda_2)}. \end{aligned}$$

Further, we let

$$x_i = \frac{\lambda_i}{\mu_i-\lambda_i} \text{ (for } i=1,2).$$

So

$$\lambda_i = \frac{\mu_i x_i}{1+x_i} \text{ (for } i=1,2).$$

Substituting $\lambda_i = \dfrac{\mu_i x_i}{1+x_i}$ into G, we have

$$G = \frac{\left(\dfrac{\mu_1 x_1}{1+x_1} + \dfrac{\mu_2 x_2}{1+x_2}\right)(\mu_1+\mu_2)}{\mu_1 - \dfrac{\mu_1 x_1}{1+x_1} + \mu_2 - \dfrac{\mu_2 x_2}{1+x_2}}$$

$$= \mu_1 x_1 + \mu_2 x_2 - \frac{\mu_1\mu_2(x_2-x_1)^2}{\mu_1(1+x_2)+\mu_2(1+x_1)}$$

$$\le \mu_1 x_1 + \mu_2 x_2$$

$$= \frac{\lambda_1\mu_1}{\mu_1-\lambda_1} + \frac{\lambda_2\mu_2}{\mu_2-\lambda_2}$$

$$= \frac{\lambda_1}{1-\dfrac{\lambda_1}{\mu_1}} + \frac{\lambda_2}{1-\dfrac{\lambda_2}{\mu_2}}.$$

It is easy to obtain

$$\sqrt{\frac{(\lambda_1+\lambda_2)}{1-\dfrac{\lambda_1+\lambda_2}{\mu_1+\mu_2}}} \le \sqrt{\frac{\lambda_1}{1-\dfrac{\lambda_1}{\mu_1}}} + \sqrt{\frac{\lambda_2}{1-\dfrac{\lambda_2}{\mu_2}}}$$

$$\Rightarrow \quad \sqrt{\frac{2k(\lambda_1+\lambda_2)}{h\left(1-\dfrac{\lambda_1+\lambda_2}{\mu_1+\mu_2}\right)}} \le \sqrt{\frac{2k\lambda_1}{h\left(1-\dfrac{\lambda_1}{\mu_1}\right)}} + \sqrt{\frac{2k\lambda_2}{h\left(1-\dfrac{\lambda_2}{\mu_2}\right)}}$$

$$\Rightarrow \quad (q_1+q_2)^* \le q_1^* + q_2^*.$$

This result indicates that the EPQ of the postponement system is lower than the EPQ of the independent system. Thus, *H1* is supported.

Analysis of *TCs for the Two Systems**

The difference in the optimal total average cost per unit time of the two systems is defined as Z^*, given by (3.1.5)−(3.1.3), as follows.

$$Z^* = \sqrt{2kh(\lambda_1+\lambda_2)\left(1-\frac{\lambda_1+\lambda_2}{\mu_1+\mu_2}\right)} - \sqrt{2kh\lambda_1\left(1-\frac{\lambda_1}{\mu_1}\right)}$$

$$-\sqrt{2kh\lambda_2\left(1-\frac{\lambda_2}{\mu_2}\right)}$$

$$= \frac{2kh\left[(\lambda_1+\lambda_2)\left(1-\frac{\lambda_1+\lambda_2}{\mu_1+\mu_2}\right)-\lambda_1\left(1-\frac{\lambda_1}{\mu_1}\right)-\lambda_2\left(1-\frac{\lambda_2}{\mu_2}\right)\right]}{A}$$

$$-\frac{4kh\sqrt{\lambda_1\lambda_2\left(1-\frac{\lambda_1}{\mu_1}\right)\left(1-\frac{\lambda_2}{\mu_2}\right)}}{A}$$

$$= \frac{2kh\left[\frac{(\lambda_1\mu_2-\lambda_2\mu_1)^2}{\mu_1\mu_2(\mu_1+\mu_2)}-2\sqrt{\lambda_1\lambda_2\left(1-\frac{\lambda_1}{\mu_1}\right)\left(1-\frac{\lambda_2}{\mu_2}\right)}\right]}{A},$$

where $A=\sqrt{2kh(\lambda_1+\lambda_2)\left(1-\frac{\lambda_1+\lambda_2}{\mu_1+\mu_2}\right)}+\sqrt{2kh\lambda_1\left(1-\frac{\lambda_1}{\mu_1}\right)}+\sqrt{2kh\lambda_2\left(1-\frac{\lambda_2}{\mu_2}\right)}$.

It should be noted that Z^* can be positive, zero or negative depending on the values of $k, h, \lambda_1, \lambda_2, \mu_1$ and μ_2. For example, consider the following two instances, $(k, h, \lambda_1, \lambda_2, \mu_1, \mu_2)_j$, j=1,2, of Z^*: $(1, 2, 100, 150, 200, 250)_1$ and $(1, 2, 100, 150, 1{,}000, 160)_2$. The first instance leads to a negative Z^*, while the second instance yields a positive Z^*. It shows that the optimal total average cost per unit time of the postponement system may not outperform the optimal total average cost per unit time of the independent system. One observation is that when their utilization rates are close, that is, $\frac{\lambda_1}{\mu_1} \approx \frac{\lambda_2}{\mu_2}$, then Z^* is negative. The major reason for a positive Z^* is that the postponement system overproduces the end-product that has a lower demand rate. Thus, *H2* is not supported.

3.1.2.2 Demands Are Met After Production Is Complete

Now we assume that the product demands are met only after a whole production batch is finished. This scenario is more appropriate for describing the inventory level of the end-products that need to be moved to another warehouse in batches or to be further processed in batches, or for which instant consumption is not possible (Zipkin [131]).

This time production starts when the inventory level drops to $\frac{\lambda q}{\mu}$. In other words, there is some overlapping between cycles. Production stops when the inventory level reaches q. The demand during production is being fulfilled by the inventory of the previous batch (Zipkin [131]). A graphical illustration is given in Fig. 3.5. It is different from the previous model described in Section 3.1.2 because the average

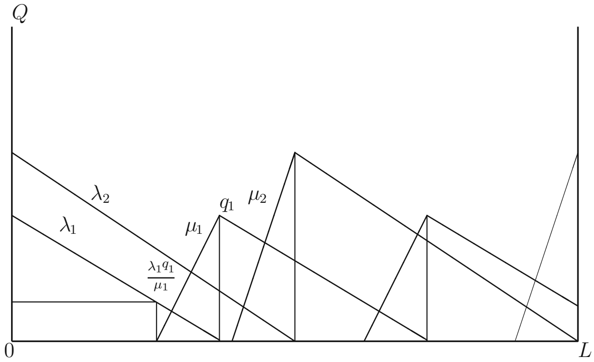

Fig. 3.5 IP for end-product 1 and 2 in the independent system when the demand is met after production is finished

inventory is higher as no demand is fulfilled by the current production batch. In general, this model operates like a model whose safety stock is $\frac{\lambda q}{\mu}$.

By revising the EPQ formula, we obtain a new total average cost per unit time for producing and keeping end-product i as follows [131].

$$C(q_i) = c\lambda_i + \frac{k\lambda_i}{q_i} + \frac{h}{2}\left(1 + \frac{\lambda_i}{\mu_i}\right) q_i, \quad i = 1, 2. \tag{3.1.6}$$

Again, if two machines are used for producing end-products 1 and 2 separately, then the total average cost per unit time for producing and keeping these two end-products is:

$$\begin{aligned} TC &= C(q_1) + C(q_2) \\ &= c(\lambda_1 + \lambda_2) + k\left(\frac{\lambda_1}{q_1} + \frac{\lambda_2}{q_2}\right) \\ &\quad + \frac{h}{2}\left[\left(1 + \frac{\lambda_1}{\mu_1}\right) q_1 + \left(1 + \frac{\lambda_2}{\mu_2}\right) q_2\right]. \end{aligned} \tag{3.1.7}$$

Minimizing TC in (3.1.7), we obtain the EPQ for end-product 1, the EPQ for end-product 2 and the optimal total average cost per unit time, respectively, as follows [131].

$$q_1^* = \sqrt{\frac{2k\lambda_1}{h\left(1 + \frac{\lambda_1}{\mu_1}\right)}},$$

$$q_2^* = \sqrt{\frac{2k\lambda_2}{h\left(1+\frac{\lambda_2}{\mu_2}\right)}},$$

and

$$\begin{aligned} TC^* &= C(q_1^*) + C(q_2^*) \\ &= c(\lambda_1+\lambda_2) + \sqrt{2kh\lambda_1\left(1+\frac{\lambda_1}{\mu_1}\right)} \\ &\quad + \sqrt{2kh\lambda_2\left(1+\frac{\lambda_2}{\mu_2}\right)}. \end{aligned} \tag{3.1.8}$$

Again, if one machine is used to produce the two end-products in aggregate with production rate $\mu_1 + \mu_2$, the total average cost per unit time equals

$$\begin{aligned} TC &= C(q_1+q_2) \\ &= c(\lambda_1+\lambda_2) + k\left(\frac{\lambda_1+\lambda_2}{q_1+q_2}\right) \\ &\quad + \frac{h}{2}\left(1+\frac{\lambda_1+\lambda_2}{\mu_1+\mu_2}\right)(q_1+q_2). \end{aligned} \tag{3.1.9}$$

A graphical illustration is given in Fig. 3.6.

Minimizing *TC* in (3.1.9), we obtain the EPQ and the optimal total average cost per unit time, respectively, as follows.

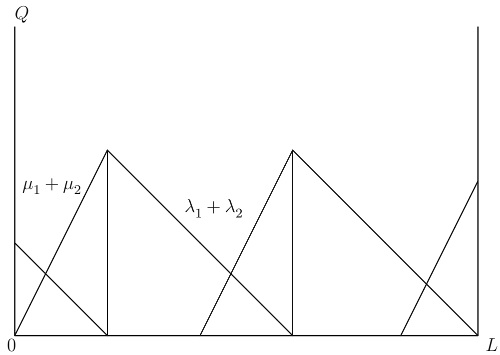

Fig. 3.6 IP for end-product 1 and 2 in the postponement system when the demand is met after production is finished

$$(q_1+q_2)^* = \sqrt{\frac{2k(\lambda_1+\lambda_2)}{h\left(1+\dfrac{\lambda_1+\lambda_2}{\mu_1+\mu_2}\right)}},$$

and

$$TC^* = c(\lambda_1+\lambda_2) + \sqrt{2kh(\lambda_1+\lambda_2)\left(1+\frac{\lambda_1+\lambda_2}{\mu_1+\mu_2}\right)}. \qquad (3.1.10)$$

Analysis of EPQs for the Two Systems

We let

$$\begin{aligned}
H &= (q_1+q_2)^* - q_1^* - q_2^* \\
&= \sqrt{\frac{2k(\lambda_1+\lambda_2)}{h\left(1+\dfrac{\lambda_1+\lambda_2}{\mu_1+\mu_2}\right)}} - \sqrt{\frac{2k\lambda_1}{h\left(1+\dfrac{\lambda_1}{\mu_1}\right)}} - \sqrt{\frac{2k\lambda_2}{h\left(1+\dfrac{\lambda_2}{\mu_2}\right)}} \\
&= \frac{\dfrac{\lambda_1+\lambda_2}{h(1+\dfrac{\lambda_1+\lambda_2}{\mu_1+\mu_2})} - \dfrac{2k\lambda_1}{h\left(1+\dfrac{\lambda_1}{\mu_1}\right)} - \dfrac{2k\lambda_2}{h\left(1+\dfrac{\lambda_2}{\mu_2}\right)} - \dfrac{4k}{h}\sqrt{\dfrac{\lambda_1\lambda_2}{\left(1+\dfrac{\lambda_1}{\mu_1}\right)\left(1+\dfrac{\lambda_2}{\mu_2}\right)}}}{B},
\end{aligned}$$

where $B = \sqrt{\dfrac{2k(\lambda_1+\lambda_2)}{h\left(1+\dfrac{\lambda_1+\lambda_2}{\mu_1+\mu_2}\right)}} + \sqrt{\dfrac{2k\lambda_1}{h\left(1+\dfrac{\lambda_1}{\mu_1}\right)}} + \sqrt{\dfrac{2k\lambda_2}{h\left(1+\dfrac{\lambda_2}{\mu_2}\right)}}.$

Further, we let

$$y_i = \frac{\lambda_i}{\mu_i+\lambda_i} \quad (\text{for } i = 1,2),$$

and

$$\lambda_i = \frac{\mu_i x_i}{1-x_i} \quad (\text{for } i = 1,2).$$

Substituting the above equations into H, we obtain

$$H = \frac{2k}{h}\left[\frac{\dfrac{\mu_1\mu_2(y_1+y_2)^2}{\mu_1(1-y_2)+\mu_2(1-y_1)} - 2\sqrt{\mu_1\mu_2 y_1 y_2}}{B}\right].$$

By numerical trials, it is shown that H can be positive, zero or negative. For example, if $k=1$, $h=2$, $\lambda_1=1$, $\lambda_2=100$, $\mu_1=150$, $\mu_2=120$, then $H>0$.

On the other hand, if $k = 1$, $h = 2$, $\lambda_1 = 50$, $\lambda_2 = 100$, $\mu_1 = 150$, $\mu_2 = 120$, then $H < 0$. This result indicates that the EPQ for the postponement system is not lower than the independent system, especially when the utilization rate $\frac{\lambda_i}{\mu_i}$ is very small for one of the production systems. For instance, if $\frac{\mu_1}{\lambda_1} \to \infty$, then $y_1 \to 0$. Thus, $H \to \frac{2k}{h}\left[\frac{\frac{\mu_1\mu_2 y_2^2}{\mu_1(1-y_2)+\mu_2}}{B}\right] > 0$. Based on our findings, *H3* is not supported.

Analysis of TC^*s for the Two Systems

The difference in the optimal total average cost per unit time of the two systems, given by (3.1.10)−(3.1.8), is as follows.

$$
\begin{aligned}
Z^* &= \sqrt{2kh(\lambda_1+\lambda_2)\left(1+\frac{\lambda_1+\lambda_2}{\mu_1+\mu_2}\right)} - \sqrt{2kh\lambda_1\left(1+\frac{\lambda_1}{\mu_1}\right)} \\
&\quad - \sqrt{2kh\lambda_2\left(1+\frac{\lambda_2}{\mu_2}\right)} \\
&= \frac{2kh\left[(\lambda_1+\lambda_2)\left(1+\frac{\lambda_1+\lambda_2}{\mu_1+\mu_2}\right) - \lambda_1\left(1+\frac{\lambda_1}{\mu_1}\right) - \lambda_2\left(1+\frac{\lambda_2}{\mu_2}\right)\right]}{C} \\
&\quad - \frac{4kh\sqrt{\lambda_1\lambda_2\left(1+\frac{\lambda_1}{\mu_1}\right)\left(1+\frac{\lambda_2}{\mu_2}\right)}}{C} \\
&= -\left[\frac{2kh\left[\frac{(\lambda_1\mu_2-\lambda_2\mu_1)^2}{\mu_1\mu_2(\mu_1+\mu_2)} + 2\sqrt{\lambda_1\lambda_2\left(1+\frac{\lambda_1}{\mu_1}\right)\left(1+\frac{\lambda_2}{\mu_2}\right)}\right]}{C}\right] \\
&< 0 \quad (h, k, \mu_1, \mu_2, \lambda_1, \lambda_2 > 0),
\end{aligned}
$$

where $C = \sqrt{2kh(\lambda_1+\lambda_2)\left(1+\frac{\lambda_1+\lambda_2}{\mu_1+\mu_2}\right)} + \sqrt{2kh\lambda_1\left(1+\frac{\lambda_1}{\mu_1}\right)} + \sqrt{2kh\lambda_2\left(1+\frac{\lambda_2}{\mu_2}\right)}$.

The result shows that the optimal total average cost per unit time for postponement is lower. Thus, *H4* is supported.

3.1.3 n Machines for n End-Products

This section extends our mathematical model to the production of n end-products. Further assumptions are presented below.

Further Assumptions

(i) End-products are produced separately by n machines.
(ii) For all end-products, the production rate is larger than the demand rate. That is, $\mu_i > \lambda_i$ for all i.

3.1.3.1 Demands Are Met Continuously

For the independent system, we derive the total average cost per unit time for producing and keeping n end-products by summing (3.1.1) over all $i, i = 1, 2, \cdots, n$.

$$TC = \sum_{i=1}^{n} C(q_i) = c\sum_{i=1}^{n}\lambda_i + k\sum_{i=1}^{n}\frac{\lambda_i}{q_i} + \frac{h}{2}\left[\sum_{i=1}^{n}\left(1-\frac{\lambda_i}{\mu_i}\right)q_i\right]$$

$$= c\sum_{i=1}^{n}\lambda_i + k\sum_{i=1}^{n}\frac{\lambda_i}{q_i} + \frac{h}{2}\left(\sum_{i=1}^{n}q_i - \sum_{i=1}^{n}\frac{\lambda_i q_i}{\mu_i}\right), \quad (3.1.11)$$

where $i = 1, 2, \cdots, n$.

Minimizing TC in (3.1.11), we obtain the EPQs for the n end-products and the optimal total average cost per unit time, respectively, as follows.

$$q_i^* = \sqrt{\frac{2k\lambda_i}{h\left(1-\frac{\lambda_i}{\mu_i}\right)}},$$

and

$$TC^* = \sum_{i=1}^{n} C(q_i^*)$$

$$= c\sum_{i=1}^{n}\lambda_i + \sum_{i=1}^{n}\sqrt{2kh\lambda_i\left(1-\frac{\lambda_i}{\mu_i}\right)}. \quad (3.1.12)$$

For the postponement system, if the aggregate production rate is $\sum_{i=1}^{n}\mu_i$, then the total average cost per unit time is equal to

$$TC = C\left(\sum_{i=1}^{n} q_i\right)$$

$$= c\sum_{i=1}^{n}\lambda_i + k\left(\frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{n}q_i}\right) + \frac{h}{2}\left(1-\frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{n}\mu_i}\right)\left(\sum_{i=1}^{n}q_i\right)$$

$$= c\sum_{i=1}^{n}\lambda_i + k\left(\frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{n}q_i}\right)$$

$$+\frac{h}{2}\left(\sum_{i=1}^{n}q_i - \frac{\sum_{i=1}^{n}\lambda_i\sum_{i=1}^{n}q_i}{\sum_{i=1}^{n}\mu_i}\right). \tag{3.1.13}$$

Minimizing TC in (3.1.13), we obtain the EPQ and the optimal total average cost per unit time, respectively, as follows.

$$\left(\sum_{i=1}^{n}q_i\right)^* = \sqrt{\frac{2k\sum_{i=1}^{n}\lambda_i}{h\left(1-\frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{n}\mu_i}\right)}},$$

and

$$TC^* = C\left[\left(\sum_{i=1}^{n}q_i\right)^*\right]$$

$$= c\sum_{i=1}^{n}\lambda_i + \sqrt{2kh\left(\sum_{i=1}^{n}\lambda_i\right)\left(1-\frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{n}\mu_i}\right)}. \tag{3.1.14}$$

Analysis of EPQs for the Two Systems

In this section, we wish to compare if $(q_1+q_2+\cdots+q_n)^* \le q_1^* + q_2^* + \cdots + q_n^*$ still holds true, provided that $i = 2$ is true (proved in Section 3.1.2.1). By mathematical induction, we assume $\sqrt{\frac{2k\sum_{i=1}^{n}\lambda_i}{h\left(1-\frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{n}\mu_i}\right)}} \le \sum_{i=1}^{n}\sqrt{\frac{2k\lambda_i}{h\left(1-\frac{\lambda_i}{\mu_i}\right)}}$ is true. Then,

$$\sqrt{\frac{2k\sum_{i=1}^{n}\lambda_i\sum_{i=1}^{n}\mu_i}{h\sum_{i=1}^{n}(\mu_i-\lambda_i)}} \le \sum_{i=1}^{n}\sqrt{\frac{2k\lambda_i\mu_i}{h(\mu_i-\lambda_i)}}.$$

Thus,

$$\sqrt{\frac{2k(\sum_{i=1}^{n}\lambda_i+\lambda_{n+1})(\sum_{i=1}^{n}\mu_i+\mu_{n+1})}{h[\sum_{i=1}^{n}(\mu_i-\lambda_i)+(\mu_{n+1}-\lambda_{n+1})]}}$$

$$\leq \sqrt{\frac{2k\sum_{i=1}^{n}\lambda_i\sum_{i=1}^{n}\mu_i}{h\sum_{i=1}^{n}(\mu_i-\lambda_i)}}+\sqrt{\frac{2k\lambda_{n+1}\mu_{n+1}}{h(\mu_{n+1}-\lambda_{n+1})}}$$

$$\leq \sum_{i=1}^{n}\sqrt{\frac{2k\lambda_i\mu_i}{h(\mu_i-\lambda_i)}}+\sqrt{\frac{2k\lambda_{n+1}\mu_{n+1}}{h(\mu_{n+1}-\lambda_{n+1})}}$$

$$=\sum_{i=1}^{n+1}\sqrt{\frac{2k\lambda_i}{h\left(1-\frac{\lambda_i}{\mu_i}\right)}}.$$

By mathematical induction, the inequality holds for all integers $i \geq 1$. The result proves that the postponement system leads to a lower EPQ. Thus, *H5* is supported. In the next section, we compare the optimal total average cost per unit time of the two systems.

Analysis of TC^*s for the Two Systems

The difference in the optimal total average cost per unit time of the two systems, given by (3.1.14)–(3.1.12), is equal to

$$Z^*=\sqrt{2kh\sum_{i=1}^{n}\lambda_i\left(1-\frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{n}\mu_i}\right)}-\sum_{i=1}^{n}\sqrt{2kh\lambda_i\left(1-\frac{\lambda_i}{\mu_i}\right)}$$

$$=\frac{2kh[\sum_{i=1}^{n}\frac{\lambda_i^2}{\mu_i}-\frac{(\sum_{i=1}^{n}\lambda_i)^2}{\sum_{i=1}^{n}\mu_i}]}{\sqrt{2kh\left(\sum_{i=1}^{n}\lambda_i\right)\left(1-\frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{n}\mu_i}\right)}+\sum_{i=1}^{n}\sqrt{2kh\lambda_i\left(1-\frac{\lambda_i}{\mu_i}\right)}}$$

$$+\frac{-4kh\sqrt{\sum_{j=1}^{n}\sum_{i=1,i\neq j}^{n-1}\lambda_j\lambda_i\left(1-\frac{\lambda_j}{\mu_j}\right)\left(1-\frac{\lambda_i}{\mu_i}\right)}}{\sqrt{2kh\left(\sum_{i=1}^{n}\lambda_i\right)\left(1-\frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{n}\mu_i}\right)}+\sum_{i=1}^{n}\sqrt{2kh\lambda_i\left(1-\frac{\lambda_i}{\mu_i}\right)}}.$$

We let the first term be Z_1^* and the second term be Z_2^*. Also, we let $a_i = \dfrac{\lambda_i}{\sqrt{\mu_i}}$, $b_i = \sqrt{\mu_i}$. By the Cauchy-Schwarz inequality, for the two sets of real numbers $a_1, a_2, \cdots, a_n$ and $b_1, b_2, \cdots, b_n$,

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leq \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right)$$

$$\left(\sum_{i=1}^{n} \lambda_i\right)^2 \leq \left(\sum_{i=1}^{n} \frac{\lambda_i^2}{\mu_i}\right)\left(\sum_{i=1}^{n} \mu_i\right)$$

$$\frac{\left(\sum_{i=1}^{n} \lambda_i\right)^2}{\sum_{i=1}^{n} \mu_i} \leq \sum_{i=1}^{n} \frac{\lambda_i^2}{\mu_i}.$$

Therefore, $Z_1^* \geq 0$ and $Z_2^* < 0$. Z^* can be positive, zero or negative, given that $k, h > 0$. It implies that the optimal total average cost per unit time for the postponement system does not give a lower optimal total average cost per unit time. This conclusion is consistent with the proof in Section 3.1.2.1 when $i = 2$. Hence, *H6* is not supported.

3.1.3.2 Demands Are Met After Production Is Complete

For the independent system, we derive the total average cost per unit time for producing and keeping n end-products by summing (3.1.6) over all $i, i = 1, 2, \cdots, n$.

$$\begin{aligned} TC &= \sum_{i=1}^{n} C(q_i) \\ &= c\sum_{i=1}^{n} \lambda_i + k\sum_{i=1}^{n} \frac{\lambda_i}{q_i} + \frac{h}{2}\left[\sum_{i=1}^{n}\left(1 + \frac{\lambda_i}{\mu_i}\right) q_i\right], \end{aligned} \tag{3.1.15}$$

where $n = 1, 2, \cdots, n$.

Minimizing TC in (3.1.15), we obtain the EPQs for the n end-products and the optimal total average cost per unit time, respectively, as follows.

$$q_i^* = \sqrt{\frac{2k\lambda_i}{h\left(1 + \dfrac{\lambda_i}{\mu_i}\right)}},$$

and

$$TC^* = \sum_{i=1}^{n} C(q_i^*)$$
$$= c\sum_{i=1}^{n}\lambda_i + \sum_{i=1}^{n}\sqrt{2kh\lambda_i\left(1+\frac{\lambda_i}{\mu_i}\right)}. \quad (3.1.16)$$

If postponement system is implemented so that the aggregate production rate is $\sum_{i=1}^{n}\mu_i$, then the total average cost per unit time becomes

$$TC = C\left(\sum_{i=1}^{n} q_i\right)$$
$$= c\sum_{i=1}^{n}\lambda_i + k\left(\frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{n}q_i}\right) + \frac{h}{2}\left(1+\frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{n}\mu_i}\right)\left(\sum_{i=1}^{n}q_i\right)$$
$$= c\sum_{i=1}^{n}\lambda_i + k\left(\frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{n}q_i}\right)$$
$$+\frac{h}{2}\left(\sum_{i=1}^{n}q_i + \frac{\sum_{i=1}^{n}\lambda_i\sum_{i=1}^{n}q_i}{\sum_{i=1}^{n}\mu_i}\right). \quad (3.1.17)$$

Minimizing TC in (3.1.17), we get the EPQ and the optimal total average cost per unit time, respectively, as follows.

$$\left(\sum_{i=1}^{n}q_i\right)^* = \sqrt{\frac{2k\sum_{i=1}^{n}\lambda_i}{h\left(1+\frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{n}\mu_i}\right)}},$$

and

$$TC^* = c\sum_{i=1}^{n}\lambda_i + \sqrt{2kh\left(\sum_{i=1}^{n}\lambda_i\right)\left(1+\frac{\sum_{i=1}^{n}\lambda_i}{\sum_{i=1}^{n}\mu_i}\right)}. \quad (3.1.18)$$

Analysis of EPQs for the Two Systems

In this section, we wish to compare if $(q_1+q_2+\cdots+q_n)^* \le q_1^* + q_2^* + \cdots + q_n^*$. In Section 3.1.2.2, we showed that it is not true for $n = 2$. This time, if $n = 5$, k=1,$h = 2, \lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 1, \mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = 100$, then $(q_1 + q_2 + \cdots + q_n)^* < q_1^* + q_2^* + \cdots + q_n^*$. On the other hand, if $n = 5$, k=1,$h =$

$2, \lambda_1 = 999, \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = 1, \mu_1 = 1,000, \mu_2 = \mu_3 = \mu_4 = \mu_5 = 100$, then $(q_1 + q_2 + \cdots + q_n)^* > q_1^* + q_2^* + \cdots + q_n^*$. This result proves that the postponement system does not result in a lower EPQ. Therefore, *H7* is not supported. In the next section, we compare the optimal total average cost per unit time of the two systems.

Analysis of TC^*s for the Two Systems

The difference in the optimal total average cost per unit time of the two systems, given by (3.1.18)−(3.1.16), is

$$Z^* = \sqrt{2kh \sum_{i=1}^{n} \lambda_i \left(1 + \frac{\sum_{i=1}^{n} \lambda_i}{\sum_{i=1}^{n} \mu_i}\right)} - \sum_{i=1}^{n} \sqrt{2kh\lambda_i \left(1 + \frac{\lambda_i}{\mu_i}\right)}.$$

It has been proved that $Z^* < 0$ for $i = 2$ in Section 3.1.2.2. We wish to prove if $Z^* < 0$ for all integers $i > 2$ by mathematical induction. First we assume $Z^* < 0$ for $i = n$. That is,

$$\sqrt{2kh \left(\sum_{i=1}^{n} \lambda_i\right) \left(1 + \frac{\sum_{i=1}^{n} \lambda_i}{\sum_{i=1}^{n} \mu_i}\right)} < \sum_{i=1}^{n} \sqrt{2kh\lambda_i \left(1 + \frac{\lambda_i}{\mu_i}\right)}.$$

Then,

$$\sqrt{2kh \left(\sum_{i=1}^{n} \lambda_i + \lambda_{n+1}\right) \left(1 + \frac{\sum_{i=1}^{n} \lambda_i + \lambda_{n+1}}{\sum_{i=1}^{n} \mu_i + \mu_{n+1}}\right)}$$

$$< \sqrt{2kh \left(\sum_{i=1}^{n} \lambda_i\right) \left(1 + \frac{\sum_{i=1}^{n} \lambda_i}{\sum_{i=1}^{n} \mu_i}\right)} + \sqrt{2kh\lambda_{n+1} \left(1 + \frac{\lambda_{n+1}}{\mu_{n+1}}\right)}$$

$$< \sum_{i=1}^{n} \sqrt{2kh\lambda_i \left(1 + \frac{\lambda_i}{\mu_i}\right)} + \sqrt{2kh\lambda_{n+1} \left(1 + \frac{\lambda_{n+1}}{\mu_{n+1}}\right)}$$

$$= \sum_{i=1}^{n+1} \sqrt{2kh\lambda_i \left(1 + \frac{\lambda_i}{\mu_i}\right)}.$$

By mathematical induction, $Z^* < 0$ for all integers $i \geq 1$. It implies postponement system results in a lower optimal total average cost per unit time. Hence, *H8* is supported.

Analysis of the Extra Customization Cost p

In this section, we consider the extra customization cost for implementing a postponement strategy. Recall that in Sections 3.1.2 and 3.1.3, we assumed that a customization process is required for a postponement strategy to differentiate the end-products when customer demands are known. The customization process incurs an extra cost for the production of each end-product. The average customization cost is $(\sum_{i=1}^{n}\lambda_i)p$, which is independent of q_i, for $i = 1, 2, \cdots, n$. The difference in the optimal total average cost per unit time of the two systems, in which the customization process is included, is $Z^* + (\sum_{i=1}^{n}\lambda_i)p$. Postponement is more cost-effective if $Z^* + (\sum_{i=1}^{n}\lambda_i)p < 0$, i.e., $(\sum_{i=1}^{n}\lambda_i)p < -Z^*$. It is not possible for the case when demands are met instantly by the current production batch (Sections 3.1.2.1 and 3.1.3.1) because $Z^* > 0$. Therefore, it is certainly not worth implementing the postponement strategy. In the case when demands are met after production is complete (Sections 3.1.2.2 and 3.1.3.2), we can determine whether or not postponement strategy should be adopted by checking if this inequality $Z^* + (\sum_{i=1}^{n}\lambda_i)p < 0$ holds.

3.2 Analysis of Postponement Strategy by an EPQ-based Model with Planned Backorders[2]

In the above section, we assumed that there is no stockout. However, stockout plays an important role both in theoretical analysis and actual practice. Consider the same EPQ system, but relax the requirement that all demands be met from stock on hand. We further assume that all demands are ultimately filled, though perhaps after a delay. That is, demands not filled immediately are backordered. It is natural to consider EPQ-based models with backorders to analyze postponement. There are some papers addressing the EPQ model with planned backorders when the end-product demand is met continuously by the current production batch [24, 99]. In these studies, the EPQ and the optimal total average cost per unit time for producing and keeping one endproduct were given.

Motivated by the above observations, we develop the EPQ model with backorders when the demand is met after production is finished in this section. We give the cost function and the optimal strategy of an EPQ-based model with planned backorders when the demand is met after production is finished. We derive the optimal total average costs per unit time of a postponement system and a non-postponement system under four different circumstances, respectively. By comparing the optimal total average costs of the two systems, we evaluate the impact of postponement on the manufacturer. Our results show that the postponement strategy can yield a lower total average cost under certain circumstances. We also find that the key factors in

[2] The following discussion in this section is largely based on the ideas and results presented in Li et al. [72].

postponement decisions are the variance of the machine utilization rates and the variance of the backorder costs.

3.2.1 Proposed Model and Assumptions

Consider a supply chain with a manufacturer and *n* customers. The manufacturer produces *n* different products in response to the demands of the *n* customers. These products are manufactured from the same type of raw material and the end products have only slight differences. These products are independent without any supply-demand links between them. The customers' demand rates and the manufacturer's production rates are deterministic and constant. The manufacturer can produce the *n* products independently on *n* machines under different production schedules such that there are *n* EPQ decisions. It is viewed as a non-postponement system. However, if the customization process can be delayed, the manufacturer can first produce a generic product. Then the production of the generic product can be carried out under the same production schedule such that there is only one EPQ decision. It is viewed as a form postponement system. Our objective is to apply the EPQ-based model with backorders to examine whether postponement outperforms non-postponement. There are two scenarios to describe the model. In the first scenario, we assume that the end-product demands are met continuously by the current production batch. In the second scenario, we assume that the end-product demands are only met after production is finished. In addition, we consider two cases of backorder costs. In the general case, we assume different backorder costs for different end-products. In the special case, we assume the backorder costs are the same for all the end-products. In sum, there are four cases to be discussed, and we wish to investigate the following four hypotheses in this section (Table 3.2).

Table 3.2 A summary of the four hypotheses with stockout

No. of machines(Measure)	Different backorder cost	Same backorder cost
n(a)	H9	H10
n(b)	H11	H12

(a) *Demands are met instantly by the current production batch*
(b) *Demands are met after production is complete*

H9: Postponement leads to a lower optimal total average cost per unit time for the manufacturer when the demands are met continuously and the planned backorder costs are not all equal.

H10: Postponement leads to a lower optimal total average cost per unit time for the manufacturer when the demands are met continuously and the planned backorder costs are the same for all the end-products.

H11: Postponement leads to a lower optimal total average cost per unit time for the manufacturer when the demands are met after production is finished and the planned backorder costs are not all equal.

H12: Postponement leads to a lower optimal total average cost per unit time for the manufacturer when the demands are met after production is finished and the planned backorder costs are the same for all the end-products.

Definitions of the notation of this section are introduced below.

- b_i = unit backorder cost per unit time for end-product $i, b_i \geq 0$,
- v_i = planned backorder quantity for end-product $i, v_i \leq 0$,
- L_i = total cycle time for end-product $i, L_i > 0$,
- L_i' = backorder lead-time for end-product $i, L_i' > 0$,
- $C(q_i, v_i)$ = total average cost per unit time for producing and keeping end-product i with production quantity q_i and planned backorder quantity v_i, $C(q_i, v_i) > 0$,
- $C\left(q_i^*, v_i^*\right)$ = the optimal total average cost per unit time for producing and keeping end-product i,
- TCP = total average cost per unit time for producing and keeping end-products $1, 2, \cdots, n$ in the postponement system with production quantity $q_1+q_2+\cdots+q_n$, $TCP > 0$,
- TCP^* = the optimal total average cost per unit time for producing and keeping end-products $1, 2, \cdots, n$ in the postponement system.

In addition, the following assumptions are made:

(i) A production cycle means the time between the production of two consecutive batches. The end-product demand rate λ_i and the production rate μ_i are deterministic and constant. To avoid unrealistic and trivial cases, we assume $\mu_i > \lambda_i, i = 1, 2, \cdots, n$. When production starts, inventory accumulates until it is enough for the cycle. Then, production stops and inventory starts to decline. When inventory drops below zero, the product is backordered (Figs. 3.7, 3.8, 3.9 and 3.10).

(ii) Demands not filled immediately are backordered and all the demands are ultimately filled. The manufacturer always uses any inventory on hand to fill the demands. Backorders accumulate only when the manufacturer runs out of stock entirely, which means that all the products will be backordered synchronously in the postponement system (Figs. 3.7, 3.8, 3.9 and 3.10).

(iii) The inventory holding cost for raw materials is ignored. In the non-postponement system, we only consider the holding cost for the end-products. In the postponement system, we only consider the holding cost for the generic products. Because the generic product and all the end-products have only slight differences, we assume that the holding cost for the generic product and all the end-products are the same.

(iv) The manufacturer incurs a common setup cost for setting up a production run and an item-specific setup cost for each product. Because all the end-products have only slight differences, the item-specific setup cost is usually much less than the common setup cost, we can assume that all the item-specific setup costs are the same. For simplicity of analysis, we further assume that all the item-specific setup cost are zero and the fixed setup cost is the only common setup cost. The manufacturer incurs common fixed setup cost, k, in each production cycle when production starts in the postponement system or in the non-postponement system.

(v) Because all the end-products have only slight differences, we assume that c and p are the same for all the end-products, respectively. Moreover, an extra customization process cost is incurred only if the customization process is delayed. In practice, the time for customization is very short. For example, an apparel manufacturer can postpone its color dyeing process at the very end of the production. The dyeing process can be finished quickly after the orders are received. So the lead-time of customization can be assumed to be negligible for simplicity of analysis.

In the following section, we discuss the four hypotheses in detail by examining whether or not the postponement system is more cost-effective than the non-postponement system.

3.2.2 Demands Are Met Continuously

3.2.2.1 Different Backorder Costs

First, we consider the general case in which the backorder costs are not all equal. A graph of the inventory position of end-product i over time in the non-postponement system is illustrated in Fig. 3.7. The horizontal axis t denotes time. The vertical axis IP denotes the inventory position of end-product i over time. Each cycle consists of an active period when production occurs and an idle period following production. In the active period, inventory increases with a slope of $\mu_i - \lambda_i$. In the idle period, inventory decreases with a slope of $-\lambda_i$. When production starts, the manufacturer incurs a fixed setup cost. When the inventory position is positive, there is inventory on hand and the manufacturer incurs holding cost. When the inventory is negative, the product is backordered and the manufacturer incurs backorder cost. The production quantity q_i and the planned backorder quantity v_i for end-product i in each cycle are our decision variables, which also determine the production cycle time L_i and the backorder lead-time L_i'. The objective of the EPQ model is to find the optimal q_i and v_i to minimize the average cost per unit time for producing and keeping end-product i.

The total average cost per unit time for producing and keeping end-product i is as follows

$$C\left(q_i, v_i\right) = c\lambda_i + \frac{k\lambda_i}{q_i} + \frac{h\left(\rho_i q_i + v_i\right)^2}{2\rho_i q_i} + \frac{b_i v_i^2}{2\rho_i q_i}, \tag{3.2.19}$$

where $\rho_i = 1 - \dfrac{\lambda_i}{\mu_i}$.

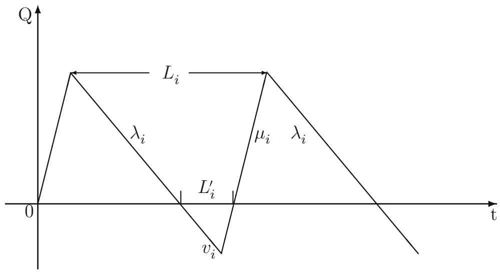

Fig. 3.7 IP for product i in the non-postponement system when the demand is met continuously

The first two terms are the variable production cost and the fixed production cost per unit time, the third term is the average inventory holding cost per unit time for end product *i*, while the last term is the average backorder quantity for end-product *i*.

Minimizing Eq. (3.7), we obtain the EPQ and optimal backorder quantity for end-product *i*, respectively, as follows (Cardenas-Barron, [24], Ronald et al. [99])

$$q_i^* = \sqrt{\frac{2k\lambda_i}{h\rho_i}}\sqrt{\frac{b_i + h}{b_i}},$$

$$v_i^* = -\sqrt{\frac{2k\rho_i\lambda_i h}{b_i\left(b_i + h\right)}}.$$

The optimal average cost per unit time for producing and keeping end-product *i* is

$$C\left(q_i^*, v_i^*\right) = c\lambda_i + \sqrt{\frac{2kh\lambda_i b_i\rho_i}{b_i + h}}. \tag{3.2.20}$$

In the non-postponement system, the production of the end-products is processed independently by n different machines with production rate μ_i, on which the end-products are customized. The optimal total average cost for producing and keeping the *n* end-products is the sum of all the costs of products *i* and is given by

$$\begin{aligned} TC^* &= \sum_{i=1}^{n} C\left(v_i^*, q_i^*\right) \\ &= c\sum_{i=1}^{n}\lambda_i + \sum_{i=1}^{n}\sqrt{\frac{2kh\lambda_i b_i\rho_i}{b_i + h}}. \end{aligned} \tag{3.2.21}$$

In the postponement system, the customization process is delayed. According to assumptions, the production of the generic product can be viewed as being processed by a single machine whose production rate is $\mu = \mu_1 + \mu_2 + \cdots + \mu_n$.

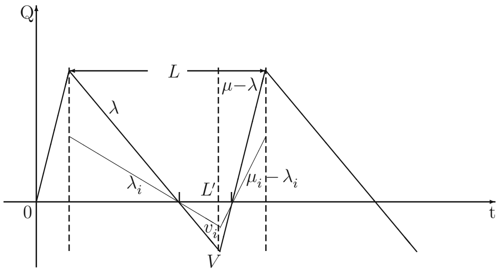

Fig. 3.8 IP for n products in the postponement system when the demand is met continuously

The unit variable production cost is c, the fixed setup cost is k, the unit holding cost per unit time is h, and the extra unit customization cost is p. A graph of the inventory position of the generic product over time in the postponement system is illustrated in Fig. 3.8. In this figure the thick lines denote the inventory position of the total generic product and the thin lines denote the inventory position of the generic product used to produce product i. According to assumptions, the production cycle of the total generic product and that of the generic product used to produce product i are the same and so is the backorder lead-time. There is only one EPQ problem in this system. The decision variables are the production quantity Q and the planned backorder quantity V of the generic product in one production cycle, which also determine the cycle time L and the backorder lead-time L'. The objective of the EPQ model is to find the optimal Q^* and V^* to minimize the total average cost per unit time in the postponement system. For simplicity, we do not consider the extra unit customization cost p, which will be discussed later.

Because we always use any inventory on hand to fill demands and because backorders accumulate only when we run out of stock entirely, the backorder lead-time for end-product i is the same ($L'_1 = L'_2 = \cdots = L'_n = L'$) and the core production rate for product i becomes μ'_i. From Fig. 3.8, we can observe that they yield to the following equations:

$$\mu = \mu_1 + \mu_2 + \cdots + \mu_n = \mu'_1 + \mu'_2 + \cdots + \mu'_n,$$

$$L' = \frac{-V}{\lambda} + \frac{-V}{\mu - \lambda} = \frac{-V}{\lambda\rho},$$

$$L' = L'_i = \frac{-v_i}{\lambda_i} + \frac{-v_i}{\mu'_i - \lambda_i} = \frac{-v_i}{\lambda_i \rho'_i},$$

where $\rho'_i = 1 - \dfrac{\lambda_i}{\mu'_i}$.

Since

$$V = v_1 + v_2 + \cdots + v_n,$$

we have

$$
\begin{aligned}
L'\lambda\rho &= -V \\
&= -v_1 - v_2 - \cdots - v_n \\
&= L'\lambda_1\rho_1' + L'\lambda_2\rho_2' + \cdots + L'\lambda_n\rho_n', \\
\lambda\rho &= \lambda_1\rho_1' + \lambda_2\rho_2' + \cdots + \lambda_n\rho_n', \\
\frac{\lambda^2}{\mu} &= \sum_{i=1}^{n} \frac{\lambda_i^2}{\mu_i'}.
\end{aligned}
$$

We let $a_i = \dfrac{\lambda_i}{\sqrt{\mu_i'}}, b_i = \sqrt{\mu_i'}$. By the Cauchy-Schwarz inequality

$$\left(\sum_{i=1}^{n} a_i b_i\right)^2 \leq \left(\sum_{i=1}^{n} a_i^2\right)\left(\sum_{i=1}^{n} b_i^2\right),$$

we have

$$\left(\sum_{i=1}^{n} \lambda_i\right)^2 \leq \left(\sum_{i=1}^{n} \frac{\lambda_i^2}{\mu_i'}\right)\left(\sum_{i=1}^{n} \mu_i'\right);$$

therefore,

$$\sum_{i=1}^{n} \frac{\lambda_i^2}{\mu_i'} - \frac{\lambda^2}{\mu} \geq 0$$

and

$$\sum_{i=1}^{n} \frac{\lambda_i^2}{\mu_i'} - \frac{\lambda^2}{\mu} = 0 \iff \frac{\lambda_1}{\mu_1'} = \frac{\lambda_2}{\mu_2'} = \cdots \frac{\lambda_n}{\mu_n'}.$$

So we have

$$\rho = \rho_1' = \rho_2' = \cdots = \rho_n'.$$

Thus, in the postponement system the total average cost per unit time for producing (excluding the customization process) and keeping the n end-products becomes

$$TCP = c\lambda + \frac{k\lambda}{Q} + \frac{h\,(\rho Q + V)^2}{2\rho Q} + \sum_{i=1}^{n} b_i\left(\frac{-v_i L'}{2L}\right)$$

$$= c\lambda + \frac{k\lambda}{Q} + \frac{h\left(\rho Q + V\right)^2}{2\rho Q} + \sum_{i=1}^{n} \frac{b_i L' \lambda_i \rho_i' L'}{2\dfrac{Q}{\lambda}}$$

$$= c\lambda + \frac{k\lambda}{Q} + \frac{h\left(\rho Q + V\right)^2}{2\rho Q} + \sum_{i=1}^{n} \frac{b_i \lambda_i \rho_i' \left(L' \rho \lambda\right)^2}{2\rho\lambda\rho Q}$$

$$= c\lambda + \frac{k\lambda}{Q} + \frac{h\left(\rho Q + V\right)^2}{2\rho Q} + \left(\sum_{i=1}^{n} \frac{b_i \lambda_i}{\lambda}\right) \frac{V^2}{2\rho Q}$$

$$= c\lambda + \frac{k\lambda}{Q} + \frac{h\left(\rho Q + V\right)^2}{2\rho Q} + \frac{\hat{B} V^2}{2\rho Q} \qquad (3.2.22)$$

$$= C\left(Q, V\right),$$

where $\hat{B} = \sum_{i=1}^{n} \dfrac{b_i \lambda_i}{\lambda}$.

Minimizing Eq. (3.2.22), we obtain the EPQ and optimal backorder quantity respectively, as follows.

$$Q^* = \sqrt{\frac{2k\lambda}{h\rho}} \sqrt{\frac{\hat{B} + h}{\hat{B}}},$$

$$V^* = -\sqrt{\frac{2k\rho\lambda h}{\hat{B}\left(\hat{B} + h\right)}},$$

and

$$TCP^* = C\left(Q^*, V^*\right) = c\lambda + \sqrt{\frac{2kh\lambda\hat{B}\rho}{\hat{B} + h}}. \qquad (3.2.23)$$

The difference in the optimal total average cost per unit time of the two systems is defined as Z^*, given by (3.2.23) − (3.2.21), as follows:

$$Z^* = TCP^* - TP^*$$

$$= \sqrt{\frac{2kh\lambda\hat{B}\rho}{\hat{B} + h}} - \sum_{i=1}^{n} \sqrt{\frac{2kh\lambda_i b_i \rho_i}{b_i + h}}$$

$$= \frac{2kh\left[\sum_{i=1}^{n}\lambda_i\left(\frac{\rho\hat{B}}{\hat{B}+h}-\frac{\rho_i b_i}{b_i+h}\right)-2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sqrt{\frac{\lambda_i\lambda_j b_i b_j\rho_i\rho_j}{(b_i+h)\left(b_j+h\right)}}\right]}{\sqrt{\frac{2kh\lambda\hat{B}\rho}{\hat{B}+h}}+\sum_{i=1}^{n}\sqrt{\frac{2kh\lambda_i b_i\rho_i}{b_i+h}}}. \tag{3.2.24}$$

It should be noted that the term $\sum_{i=1}^{n}\lambda_i\left(\frac{\rho\hat{B}}{\hat{B}+h}-\frac{\rho_i b_i}{b_i+h}\right)$ and Z^* can be positive. For example, when $i=2, h=2, k=50, c=20, b_1=50, \lambda_1=990, \mu_1=1{,}000, b_2=500, \lambda_2=1, \mu_2=1{,}000$, then $Z^*=206.5>0$, $\frac{Z^*}{TC^*}=1.03\%>0$. Therefore, Z^* can be positive, zero or negative (Table 3.3). It implies that the postponement system does not always give a lower optimal total average cost per unit time. Thus, *H9* is not supported.

3.2.2.2 Same Backorder Cost

Now we consider a special case in which the backorder cost b_i is the same for all the end-products. Letting $b_1=b_2=\cdots=b_n=b$ in formula (3.2.21), (3.2.23) and (3.2.24), we obtain the following results.

The optimal total average cost for producing and keeping the n end-products in the non-postponement system is given by

$$\begin{aligned} TC^* &= \sum_{i=1}^{n} C\left(v_i^*, q_i^*\right) \\ &= c\sum_{i=1}^{n}\lambda_i+\sum_{i=1}^{n}\sqrt{\frac{2kh\lambda_i b\rho_i}{b+h}}. \end{aligned} \tag{3.2.25}$$

The optimal total average cost per unit time in the postponement system is as follows:

$$\begin{aligned} TCP^* &= C\left(Q^*, V^*\right) \\ &= c\lambda+\sqrt{\frac{2khb\lambda\rho}{b+h}}. \end{aligned} \tag{3.2.26}$$

The difference in the optimal total average cost per unit time of the two systems is defined as Z^*, given by (3.2.26) − (3.2.25), as follows:

$$Z^* = c\lambda+\sqrt{\frac{2khb\lambda\rho}{b+h}}-c\sum_{i=1}^{n}\lambda_i-\sum_{i=1}^{n}\sqrt{\frac{2kh\lambda_i b\rho_i}{b+h}}$$

$$= \frac{\sqrt{\frac{2khb}{b+h}\left(\sum_{i=1}^{n}\frac{\lambda_i^2}{\mu_i}-\frac{\lambda^2}{\mu}-2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sqrt{\lambda_i\lambda_j\rho_i\rho_j}\right)}}{\sqrt{\rho\lambda}+\sum_{i=1}^{n}\sqrt{\rho_i\lambda_i}}. \quad (3.2.27)$$

By the Cauchy-Schwarz inequality, we have

$$\sum_{i=1}^{n}\frac{\lambda_i^2}{\mu_i}-\frac{\lambda^2}{\mu}\geq 0$$

and

$$\sum_{i=1}^{n}\frac{\lambda_i^2}{\mu_i}-\frac{\lambda^2}{\mu}=0 \iff \frac{\lambda_1}{\mu_1}=\frac{\lambda_2}{\mu_2}=\cdots\frac{\lambda_n}{\mu_n}.$$

Z^* can be positive. For example, when $i = 2, h = 2, k = 50, c = 20, b = 100, \lambda_1 = 50, \mu_1 = 1{,}000, \lambda_2 = 950, \mu_2 = 1{,}000, Z^* = 120.1 > 0, \frac{Z^*}{TC^*} = 0.59\%$. Thus, Z^* can be positive, zero or negative (Table 3.4). *H10* is not supported.

In the first scenario, *H9* and *H10* are not supported. The postponement system does not always give a lower optimal total average cost per unit time when customer demands are continuously met by the current production batch. But we can observe that if $b_1 = b_2 = \cdots = b_n = b$ and $\frac{\lambda_1}{\mu_1} = \frac{\lambda_2}{\mu_2} = \cdots = \frac{\lambda_n}{\mu_n}$, then $Z^* < 0$. It implies that the variance in backorder costs $b_1, b_2, \cdots, b_n$ and the variance in machine utilization rates $\frac{\lambda_1}{\mu_1}, \frac{\lambda_2}{\mu_2}, \cdots, \frac{\lambda_n}{\mu_n}$ are key factors in a postponement decision. When there are a large number of end-products, we can group them into different product families based on machine utilization rates and backorder costs. Those products whose machine utilization rates and backorder costs are equal or close can share a single lot size and gain a lower total average cost.

3.2.3 Demands Are Met After Production Is Complete

3.2.3.1 Different Backorder Costs

First, we consider the general case in which the backorder costs are not all equal. A graph of the inventory position of end-product i over time in the non-postponement system is illustrated in Fig. 3.9. In one production cycle, there are two lines. One increases with a slope of μ_i, which denotes the inventory position over time when product i is being produced. The other line decreases with a slope of λ_i, which denotes the inventory position over time when product i is being consumed. The production quantity q_i and the planned backorder quantity v_i for end-product i in each cycle are our decision variables, which also determine the production cycle

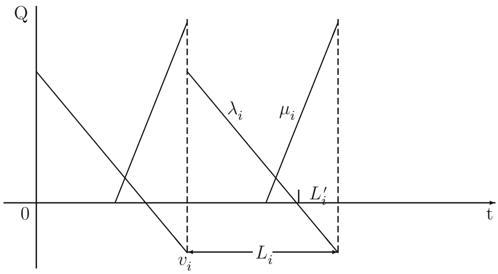

Fig. 3.9 IP for product i in the non-postponement system when the demand is met after production is finished

time L_i and the backorder lead-time L_i'. The objective is to find the optimal q_i^* and v_i^* to minimize the average cost per unit time for producing and keeping end-product i.

The total average cost per unit time for producing and keeping end-product i is as follows

$$C(q_i, v_i) = c\lambda_i + \frac{k\lambda_i}{q_i} + \frac{h}{2}\left(\frac{q_i\lambda_i}{\mu_i} + \frac{(q_i+v_i)^2}{q_i}\right) + \frac{b_i v_i^2}{2q_i}.$$

The cost $C(q_i, v_i)$ is a function of two variables. $C(q_i, v_i)$ is continuously differentiable and strictly convex on its domain. To minimize this cost, we equate its partial derivatives to zero. We obtain the EPQ and the optimal backorder quantity, respectively, as follows.

$$q_i^* = \sqrt{\frac{2k\lambda_i}{h\left(\frac{b_i}{b_i+h} + \frac{\lambda_i}{\mu_i}\right)}}$$

$$v_i^* = -\frac{b_i}{b_i+h} q_i^*.$$

The optimal total average cost per unit time for producing and keeping n end-products in the non-postponement system is given by

$$\begin{aligned} TC^* &= \sum_{i=1}^{n} C\left(q_i^*, v_i^*\right) \\ &= c\sum_{i=1}^{n}\lambda_i + \sum_{i=1}^{n}\sqrt{2kh\lambda_i\left(\frac{b_i}{b_i+h} + \frac{\lambda_i}{\mu_i}\right)}. \end{aligned} \tag{3.2.28}$$

Similarly, in the postponement system we assume that the core production is carried out by a single machine whose production rate is $\mu = \mu_1 + \mu_2 + \cdots + \mu_n$, the variable production cost is c, the fixed setup cost is k, the unit holding cost per unit time is h, and the extra unit customization cost is p. A graph of the inventory position of the generic product over time in the postponement system is illustrated in Fig. 3.10. The production quantity Q and the planned backorder quantity V for generic product in each cycle are our decision variables, which also determine the production cycle time L and the backorder lead-time L'. The objective is to find the optimal Q^* and V^*.

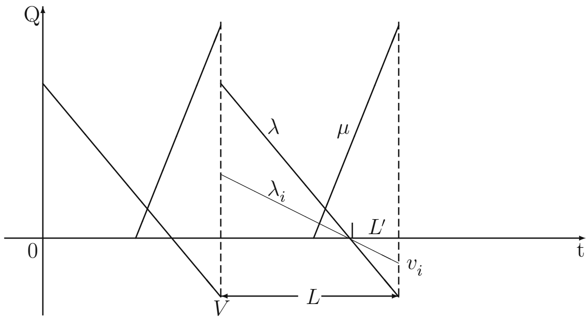

Fig. 3.10 IP for n products in the postponement system when the demand is met after production is finished

From Fig. 3.10, we can observe that

$$L = \frac{Q}{\lambda} = \frac{q_i}{\lambda_i},$$
$$L' = -\frac{V}{\lambda},$$
$$L'_i = -\frac{v_i}{\lambda_i},$$
$$L' = L'_1 = L'_2 = \cdots = L'_n.$$

Thus, the total average cost for producing and keeping these n end-products in the postponement system is

$$TCP = c\lambda + \frac{k\lambda}{Q} + \frac{h}{2}\left(\frac{Q\lambda}{\mu} + \frac{(Q+V)^2}{Q}\right) + \sum_{i=1}^{n} b_i \left(\frac{-v_i L'}{2L}\right)$$

$$= c\lambda + \frac{k\lambda}{Q} + \frac{h}{2}\left(\frac{Q\lambda}{\mu} + \frac{(Q+V)^2}{Q}\right) + \sum_{i=1}^{n} b_i \left(\frac{-L'\lambda_i v}{2Q}\right)$$

$$= c\lambda + \frac{k\lambda}{Q} + \frac{h}{2}\left(\frac{Q\lambda}{\mu} + \frac{(Q+V)^2}{Q}\right) + \sum_{i=1}^{n} \frac{b_i \lambda_i V^2}{2\lambda Q}$$

$$= c\lambda + \frac{k\lambda}{Q} + \frac{h}{2}\left(\frac{Q\lambda}{\mu} + \frac{(Q+V)^2}{Q}\right) + \frac{\hat{B}V^2}{2Q} \tag{3.2.29}$$

$$= C\,(Q, V),$$

where $\hat{B} = \sum_{i=1}^{n} \frac{b_i \lambda_i}{\lambda}$.

Minimizing (3.2.29), we obtain the EPQ and the optimal total average cost per unit, respectively, as follows.

$$Q^* = \sqrt{\frac{2k\lambda}{h\left(\frac{\hat{B}}{\hat{B}+h} + \frac{\lambda}{\mu}\right)}},$$

$$V^* = -\frac{\hat{B}}{\hat{B}+h} Q^*,$$

and

$$TCP^* = C\left(Q^*, V^*\right)$$

$$= c\lambda + \sqrt{2kh\lambda\left(\frac{\hat{B}}{\hat{B}+h} + \frac{\lambda}{\mu}\right)}. \tag{3.2.30}$$

The difference in the optimal total average cost per unit time of the two systems is defined as Z^*, given by (3.2.30) − (3.2.28), as follows

$$Z^* = TCP^* - TC^*$$

$$= \sqrt{2kh\lambda\left(\frac{\hat{B}}{\hat{B}+h} + \frac{\lambda}{\mu}\right)} - \sum_{i=1}^{n}\sqrt{2kh\lambda_i\left(\frac{b_i}{b_i+h} + \frac{\lambda_i}{\mu_i}\right)}$$

$$= \frac{2kh}{D}\Bigg[\sum_{i=1}^{n}\frac{\lambda_i h\left(\hat{B}-b_i\right)}{\left(\hat{B}+h\right)(b_i+h)} + \left(\frac{\lambda^2}{\mu} - \sum_{i=1}^{n}\frac{\lambda_i^2}{\mu_i}\right)$$

$$-2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sqrt{\lambda_i\lambda_j\left(\frac{\lambda_i}{\mu_i} + \frac{b_i}{b_i+h}\right)\left(\frac{\lambda_j}{\mu_j} + \frac{b_j}{b_j+h}\right)}\Bigg]. \tag{3.2.31}$$

where $D = \sqrt{2kh\lambda\left(\dfrac{\hat{B}}{\hat{B}+h}+\dfrac{\lambda}{\mu}\right)} + \sum_{i=1}^{n}\sqrt{2kh\lambda_i\left(\dfrac{b_i}{b_i+h}+\dfrac{\lambda_i}{\mu_i}\right)}$.

The term $\sum_{i=1}^{n} \dfrac{\lambda_i h\left(\hat{B}-b_i\right)}{\left(\hat{B}+h\right)(b_i+h)}$ can be positive, zero or negative. For example, when $i = 2, h = 1, k = 1, c = 2, b_1 = 1, \lambda_1 = 1{,}000, \mu_1 = 1{,}200, b_2 = 1{,}000, \lambda_2 = 10, \mu_2 = 12, Z^* = 1.7496 > 0, \dfrac{Z^*}{TC^*} = 0.0084\% > 0$. Thus, Z^* can be positive, zero or negative (Table 3.5). *H11* is not supported.

3.2.3.2 Same Backorder Cost

If $b_1 = b_2 = \cdots = b_n = b$, according to Eq. (3.2.31), the difference in the optimal total average cost per unit time of the two systems is given by

$$Z^* = TCP^* - TC \tag{3.2.32}$$

$$= -\frac{\sqrt{2kh}\left(\sum_{i=1}^{n}\frac{\lambda_i^2}{\mu_i} - \frac{\lambda^2}{\mu} + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}\sqrt{\lambda_i\lambda_j\left(\frac{\lambda_i}{\lambda_i}+\frac{b}{b+h}\right)\left(\frac{\lambda_j}{\lambda_j}+\frac{b}{b+h}\right)}\right)}{\sqrt{\lambda\left(\frac{b}{b+h}+\frac{\lambda}{\mu}\right)} + \sum_{i=1}^{n}\sqrt{\lambda_i\left(\frac{b}{b+h}+\frac{\lambda_i}{\mu_i}\right)}} < 0.$$

We have shown that the postponement system always gives a lower optimal total average cost per unit time than the non-postponement system. Thus, *H12* is supported.

In the second scenario *H11* is not supported, but *H12* is supported. It implies that the variance in backorder costs $b_1, b_2, \cdots, b_n$ is a key factor in a postponement decision. When there are a large number of end-products, we can group them into different product families based on backorder costs. Those products whose backorder costs are equal or close can share a single lot size and gain a lower total average cost.

Now we consider the extra customization cost in the second scenario. It is obvious that the average customization cost per unit time is $\left(\sum_{i=1}^{n}\lambda_i\right)p$. The difference in the optimal total average cost per unit time of the two systems is $Z^* + \left(\sum_{i=1}^{n}\lambda_i\right)p$. Postponement is more cost-effective if $Z^* + \left(\sum_{i=1}^{n}\lambda_i\right)p < 0$.

3.2.3.3 Numerical Examples

We give numerical examples to illustrate how postponement and the key factors impact on the optimal total average cost of the two scenarios we have presented in this section. We assume that the manufacturer produces five end-products. They can be produced in non-postponement system or in postponement system. The difference in the optimal total average cost per unit time between the two systems is denoted as Z^*. $Z^* < 0$ means that the postponement system outperforms the non-postponement system. Z^*/TC^* denotes the relative difference between the two systems. For the five products, the unit common variable production cost c is 20, the common fixed setup cost k is 50, and the common unit holding cost h per unit time is 2 (all in appropriate units).

For the first scenario in which the demands are met continuously, we first assume that $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_0 = 250$, $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_0 = 500$ and that b_i is variable. The values of their various parameters and the results are shown in Table 3.3, from which the following observations can be made:

Table 3.3 Impact of backorder costs when the demands are met continuously

	c	k	h	λ_0	μ_0	b_1	b_2	b_3	b_4	b_5	Z^*	$\frac{Z^*}{TC^*}$ (%)
1	20	50	2	250	500	100	100	100	100	100	–432.7	–1.68
2	20	50	2	250	500	80	90	100	110	120	–421.5	–1.68
3	20	50	2	250	500	60	80	100	120	140	–432.0	–1.68
4	20	50	2	250	500	40	70	100	130	160	–430.7	–1.67
5	20	50	2	250	500	20	60	100	140	180	–427.0	–1.66

- The postponement system yields savings in the total average cost.
- The absolute value of Z^* and $\frac{Z^*}{TC^*}$ becomes smaller when the variance of the backorder costs becomes larger. This means that the smaller the variance of the backorder costs, the more cost-effective the postponement system is when the demands are met continuously.

For the first scenario, we then assume that $b_1 = b_2 = b_3 = b_4 = b_5 = b = 100$, $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_0 = 500$ and that λ_i is variable. The values of their various parameters and the results are shown in Table 3.4, from which the following observations can be made:

- The postponement system yields savings in the total average cost.
- The absolute value of Z^* and $\frac{Z^*}{TC^*}$ becomes smaller when the variance in the machine utilization rate becomes larger. This means that the smaller the variance of the machine utilization rates, the more cost-effective the postponement system is when the demands are met continuously.

For the second scenario in which the demands are met after the production is finished, we first assume that $\lambda_1 = \lambda_2 = \lambda_3 = \lambda_4 = \lambda_5 = \lambda_0 = 250$, $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_0 = 500$ and that b_i is variable. The values of their various parameters and the results are shown in Table 3.5, from which the following observations can be made:

- The postponement system yields savings in the total average cost.
- The absolute value of Z^* and $\frac{Z^*}{TC^*}$ becomes smaller when the variance in the backorder costs becomes larger. This means that the smaller the variance of the

Table 3.4 Impact of machine utilization rates when demands are met continuously

	c	k	h	b	λ_1	λ_2	λ_3	λ_4	λ_5	μ_0	Z^*	$\frac{Z^*}{TC^*}$(%)
1	20	50	2	100	250	250	250	250	250	500	–432.7	–1.68
2	20	50	2	100	190	220	250	280	310	500	–421.3	–1.63
3	20	50	2	100	130	190	250	310	370	500	–385.1	–1.50
4	20	50	2	100	70	160	250	340	430	500	–315.9	–1.23
5	20	50	2	100	10	130	250	370	490	500	–168.8	–0.66

Table 3.5 Impact of backorder costs when the demands are met after production is finished

	c	k	h	λ_0	μ_0	b_1	b_2	b_3	b_4	b_5	Z^*	$\frac{Z^*}{TC^*}$(%)
1	20	50	2	250	500	100	100	100	100	100	–752.0	–2.85
2	20	50	2	250	500	80	90	100	110	120	–751.7	–2.85
3	20	50	2	250	500	60	80	100	120	140	–751.1	–2.85
4	20	50	2	250	500	40	70	100	130	160	–749.7	–2.84
5	20	50	2	250	500	20	60	100	140	180	–745.4	–2.83

backorder costs, the more cost-effective the postponement system is when the demands are met after production is finished.

For the second scenario, we then assume that $b_1 = b_2 = b_3 = b_4 = b_5 = b = 100$, $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5 = \mu_0 = 500$, and that λ_i is variable. The values of their various parameters and the results are shown in Table 3.6, from which the following observations can be made:

- The postponement system yields savings in the total average cost.
- The absolute value of Z^* and $\frac{Z^*}{TC^*}$ becomes smaller when the variance in the machine utilization rate becomes larger.

Table 3.6 Impact of machine utilization rates when the demands are met after production is finished

	c	k	h	b	λ_1	λ_2	λ_3	λ_4	λ_5	μ_0	Z^*	$\frac{Z^*}{TC^*}$(%)
1	20	50	2	100	250	250	250	250	250	500	–752.0	–2.85
2	20	50	2	100	190	220	250	280	310	500	–749.8	–2.84
3	20	50	2	100	130	190	250	310	370	500	–742.5	–2.82
4	20	50	2	100	70	160	250	340	430	500	–726.8	–2.76
5	20	50	2	100	10	130	250	370	490	500	–683.1	–2.60

From the the numerical examples in this section (Tables 3.3, 3.4, 3.5 and 3.6) and the numerical examples in Sections 3.2.2 and 3.2.3, we can derive the following results.

1. Under some circumstances, the postponement system yields savings in the total average cost. But when the variance of the backorder costs is very large, or the variance of the machine utilization rates is very large in the case that the demands

Table 3.7 A summary of the findings of the eight hypotheses without stockout

No. of machines (Measure)	Economic Production Quantity	Optimal total average cost
2(a)	H1 is supported	H2 is not supported
2(b)	H3 is not supported	H4 is supported
n(a)	H5 is supported	H5 is not supported
n(b)	H7 is not supported	H8 is supported

(a) *Demands are met instantly by the current production batch*
(b) *Demands are met after production is complete*

are met continuously, it is possible that the postponement system does not outperform the non-postponement system. So the manufacturer must be careful to find a proper trade-off between postponement and non-postponement in such cases.

2. The smaller the variance of the backorder costs and the variance of the machine utilization rates, the more cost-effective the postponement system is in the two scenarios.
3. The cost saving in the second scenario is more than that in the first scenario. This means that it is more appropriate to apply postponement when the demands are met after production is finished.

3.3 Concluding Remarks

This chapter examined the impact of postponement based on some EPQ-based models without and with stockout. We formulated the total average cost functions of the two scenarios in a supply chain, in which their demands are known and deterministic.

For the first scenario in which there is no stockout, we examined the impacts of postponement on EPQ and the optimal total average cost per unit time for a supply chain that offers n end-products with slight customization. We analyzed eight hypotheses in this section and a summary table is shown below (Table 3.7).

Our findings reveal that postponement results in lower EPQ when demands are met instantly by the current production batch. However, saving in the optimal total average cost per unit time is not guaranteed when demands are met instantly by the current production batch. On the other hand, postponement does not give a lower EPQ but a lower optimal total average cost per unit time when demands are met after production is complete. Generally speaking, the postponement system may outperform the independent system as the aggregate production rate ($\mu_1 + \mu_2$ in Section 3.1.2 or $\sum_{i=1}^{n} \mu_i$ in Section 3.1.3) is only one of the feasible choices for the postponement system. It does not always lead to an optimal solution for all μ_i. We use these two production rates because they are more equitable for comparison purposes. If we let the aggregate production rate be another decision variable, we can always find a lower average inventory holding cost per unit time for a postponement system so that saving from the optimal total average cost per unit time is enough to cover the extra customization cost.

For the second scenario in which there is backorder, four hypotheses were considered. In the first and the second hypotheses the demands are met instantly by the current production batch. In the third and the fourth hypotheses the demands are met after production is completed. We analyzed the four hypotheses in this section and a summary table is shown below (Table 3.8).

Table 3.8 A summary of the findings of the four hypotheses with stockout

No. of machines(Measure)	Different backorder cost	Same backorder cost
n(a)	H9 is not supported	H10 is not supported
n(b)	H11 is not supported	H12 is supported

(a) *Demands are met instantly by the current production batch*
(b) *Demands are met after production is complete*

The postponement system may not always outperform the non-postponement system. *H9* and *H10* were not supported. But when all the backorder costs are equal and all the machine utilization rates are equal, the postponement system can outperform the non-postponement system. The key factors in a postponement decision are the variance in backorder costs and the variance in machine utilization rates. *H11* was not supported but *H12* was supported. The key factors in a postponement decision are the variance in backorder costs. If the backorder costs are equal, the postponement system can outperform the non-postponement system. Our analysis and numerical example imply that the end-products can be classified into different product families based on machine utilization rates and backorder costs. The production of those products whose machine utilization rates and backorder costs are equal or close can be handled by postponement and a lower total average cost can be gained.

In general, this chapter has demonstrated how postponement can achieve a lower cost in a supply chain. Although a number of simplifying assumptions are made in our model, our analysis should still be valid for more general systems. One research direction is to include the customization cost in the comparison so that the cost difference between a postponement system and a non-postponement system is more evident. An analysis that includes a customization cost will be discussed in Chapter 5.

Chapter 4
Evaluation of a Postponement System with an (*r*, *q*) Policy

In this chapter we study the cost impact of the pull postponement strategy by comparing the total average cost function with the optimal or an approximately optimal total average cost of an (r,q) policy. This is a stochastic model of a single end-product supply chain that consists of a supplier, a manufacturer and a number of customers. We develop two distinct models to represent the inventory system of the manufacturer. We employ Markov chain analysis to determine the exact average inventory level and the exact average accumulated backorder per period at the steady state so that the total average cost can be evaluated analytically. Also, we design an algorithm to find a near optimal total average cost per period. Our results show that the postponement system is more cost effective when the lead-time is zero, while the (r,q) inventory system is more effective when the lead-time is greater than zero.

This chapter is organized as follows. The proposed models of the postponement system and the non-postponement system are described in Sections 4.1 and 4.2. Then the algorithm for finding a near optimal total average cost of an (r,q) policy is discussed in Section 4.3. In Section 4.4, the total average cost for the postponement system is derived. In Sections 4.5 and 4.6, comparison results are generated by both optimization and simulation techniques. Some concluding remarks are given in Section 4.7.[1]

4.1 The Proposed Models and Assumptions

In this section, an (r,q) inventory policy is formulated to compare the cost difference between a postponement system and a non-postponement system, in which they supply a common end-product. We develop a supply chain model which involves a supplier, a manufacturer and a group of customers. We assume the model is discrete in time period. As the manufacturer, we treat a postponement system as a zero stock policy or the so-called make-to-order system. It has been argued whether or not a zero stock policy of a single product is a postponement system, as there is no point

[1] The following discussion in this chapter is largely based on the ideas and results presented in Wan [125].

T.C.E. Cheng et al., *Postponement Strategies in Supply Chain Management*, International Series in Operations Research & Management Science 143, DOI 10.1007/978-1-4419-5837-2_4,

of product differentiation. However, it is justified to refer it to as a postponement system because the system can be divided into both forecast driven and demand driven processes, though the demand driven process is only for producing a single product. Therefore, it is more appropriate to consider this system as a special case of postponement application. In this system, when demand from customer at period t, $D(t)$, arrives, we order $D(t)$ from the supplier at time t. Customer will receive the order at period $t + L$, where L is the order lead-time. A schematic diagram is shown in Fig. 4.1. On the other hand, in a non-postponement system, we keep a base stock inventory as r. When demand at period t, $D(t)$, arrives, we check the inventory level $IL(t-1)$ ($IL(t-1)$ is also the beginning inventory level at period t). Note that $IL(t) = IL(t-1) + A(t) - D(t)$. If $IL(t) \geq 0$, then $D(t)$ is fulfilled at period t; otherwise $B(t) = -IL(t)$ is backordered. A schematic diagram is shown in Fig. 4.2.

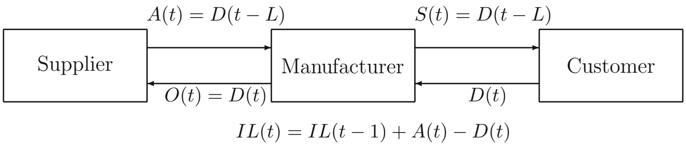

Fig. 4.1 A schematic diagram of the postponement system

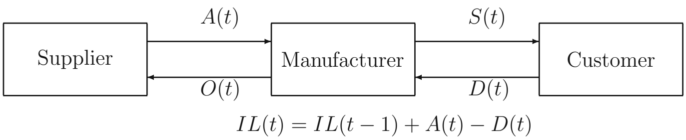

Fig. 4.2 A schematic diagram of the non-postponement system

Definitions of the notation of this chapter are introduced below.

- k = Fixed order cost, $k > 0$;
- c = Variable cost per unit, $c > 0$;
- h = Inventory holding cost per unit per period, $h > 0$;
- b = Backorder cost per unit per period, $b > 0$;
- r = Reorder point, $r \geq 0$;
- q = Order quantity;
- $I(t)$ = On-hand inventory at the end of period t, $I(t) \geq 0$;
- $B(t)$ = Accumulated backorder at the end of period t, $B(t) \geq 0$;
- $O(t)$ = Order quantity placed at the end of period t, $O(t) \geq 0$;
- $A(t)$ = Order arrived in the beginning of period t, $A(t) \geq 0$;
- $IL(t)$ = Inventory level at the end of period t, $IL(t) = I(t) - B(t)$;
- $IP(t)$ = Inventory position at the end of period t before placing order, $IP(t) = I(t) - B(t)+$outstanding orders at period t;
- $S(t)$ = Order shipped out at the end of period t, $S(t) \geq 0$;

- $D(t)$ = Demand in the beginning of period t, $D(t) \geq 0$;
- L = Order lead-time for the end-product, $L \geq 0$;
- $\overline{I}$ = Average inventory per period, $\overline{I} \geq 0$;
- $\overline{B}$ = Average backorder per period, $\overline{B} \geq 0$;
- $\overline{OF}$ = Average order frequency per period, $\overline{OF} \geq 0$;
- $\overline{D}$ = Average demand per period, $\overline{D} \geq 0$;
- $C(r,q)$ = Total average cost per period when r and q are given,$C(r,q) > 0$;
- $C(r^*,q^*)$ = Optimal total average cost per period when optimal r^* and q^* are given, $C(r^*,q^*) > 0$;
- R = Range of inventory level $IL(t)$, $R \geq 0$;
- N = Total number of states in a Markov chain, $N \geq 0$;
- S_τ = State of a Markov chain, $S = 1,2,\cdots,N$, $\tau = 0,1,\cdots,L-1$;
- τ = An indicator to record the number of periods for next order arrival. The number of periods remain for next order arrival is $L-\tau$, where $\tau = 0,1,\cdots,L-1$.
- π = Row vector of the steady state probabilities π_i, $i = 1,2,\cdots,N$;
- p_{ij} = Transition probability that a system is in state j from state i after 1 period, $i,j \in N$;
- $\mathbf{P}$ = Transition matrix of a Markov chain;
- $Pr\{.\}$ = Probability of an event.

The assumptions of this model are presented below.

(i) There is only one end-product supplied.
(ii) Order lead-time L is constant.
(iii) The supplier has unlimited capacity.
(iv) Order $A(t)$ arrives in the beginning of period t with no delay, $A(0) = 0$.
(v) Customer demand at period t, $D(t)$, is not known until the beginning of period t, $D(0) = 0$.
(vi) Customer demand at period t, $D(t)$, is handled in the beginning of period t.
(vii) Backorders are fulfilled immediately when there is enough inventory at the beginning of period t.

4.2 System Dynamics for a Non-postponement System

We employ an (r,q) policy to manage the inventory. We update the inventory level by the following system dynamics.

$$IL(t) = IL(t-1) + A(t) - D(t),\ t = 1,2,\cdots \tag{4.2.1}$$

$$B(t) = -\min\{IL(t),0\},\ t = 0,1,\cdots \tag{4.2.2}$$

$$I(t) = \max\{IL(t),0\},\ t = 0,1,\cdots \tag{4.2.3}$$

$$IP(t) = IL(t) + \sum_{i=t+1-L}^{t-1} O(i),\ t = 0,1,\cdots$$

and

$$O(i) = 0 \text{ for all } i < 0. \tag{4.2.4}$$

If lead-time $L = 0$, then the inventory is replenished immediately by a quantity q whenever $IP(t) \leq r$. That is, $IL(t) = IL(t-1) + q - D(t)$ if $IP(t) \leq r$. Otherwise, $IL(t) = IL(t-1) - D(t)$. For $L \geq 1$, The reorder decision $O(t) = A(t+L) = q$ is made when $IP(t) \leq r$. Otherwise, $O(t) = A(t+L) = 0$.

Our objective is to find an optimal (r^*, q^*) policy that will result in an optimal total average cost per period $C(r^*, q^*)$. Then we compare it with the total average cost per period of the postponement system.

The objective function for the non-postponement system is as follows.

$$\min C(r, q) = k\overline{OF} + c\overline{D} + h\overline{I} + b\overline{B},$$

where

- **(i)** $\overline{OF}$ is the average order frequency,
- **(ii)** $\overline{D}$ is the average demand,
- **(iii)** $\overline{I}$ is the average on-hand inventory,
- **(iv)** $\overline{B}$ is the average backorder per period.

Assume that customer demand at period t follows a uniform distribution $U(0, a)$, where a is the maximum value of a demand per period. It implies that $Pr\{D(t) = i\} = \frac{1}{a+1}$ for $i = 0, 1, \cdots, a$. $Pr\{D(t) = i\}$ is independent of period t. The average demand $\overline{D}$ is $\frac{a}{2}$ and the average order frequency $\overline{OF}$ is $\frac{\overline{D}}{q} = \frac{a}{2q}$.

$IL(t)$, by definition, depends on $IL(t-1)$, order arrival $A(t)$ and customer demand $D(t)$. If the order arrival $A(t)$ is known, then $IL(t)$ is not affected by the inventory level before period $t-1$. It satisfies the Markovian property that the conditional probability of any future event is dependent on present events $A(t)$ and $D(t)$ and independent of all past events. That is, $Pr\{IL(t) = j|IL(t-1) = i\} = Pr\{IL(t) = j|IL(0) = i_0, IL(1) = i_1, \cdots, IL(t-1) = i_{t-1}\}$ (Hillier and Lieberman [53]). Thus, we can develop a Markov chain to determine the steady state probability distribution of $IL(t)$ so that the average inventory $\overline{I}$ and average backorder $\overline{B}$ can be evaluated for further cost analysis and comparison.

4.3 The Algorithm for Finding a Near Optimal Total Average Cost of an (*r*, *q*) Policy

4.3.1 The Markov Chain Development

In this section, we refer the inventory level to as a Markov process, in which each possible value is defined as a state S_τ and the change in one period is called a transition. $Pr\{IL(t) = j|IL(t-1) = i\}$ is the transition probability. As suggested by Render et al. [98], there are four further assumptions for Markov chain development. With another two assumptions tailored for this model, the six assumptions are stated as follows.

(i) The number of states N is finite and constant for each transition.
(ii) All possible states are included in the Markov chain. It is also known as collectively exhaustive.
(iii) The system can be in only one of the states. It follows the mutually exclusive property.
(iv) All transition probabilities in the transition matrix $\mathbf{P} = (p_{ij})$ are unchanged in each transition.
(v) We restrict that there is no more than one single order outstanding during the order lead-time L (Hadley and Whitin [51]). The reorder decision is simply to place an order q, when $IL(t) \leq r$ and there is no order outstanding. It makes the Markov process less complicated.
(vi) It is assumed that the order quantity is sufficient to cover the demands during lead-time so that customers do not have to wait for more than L periods for having the end-products. That is, $a \leq q$ for $L = 0$ and $La \leq q$ for $L \geq 1$.

Variables that move the system from a present state to the next state are order arrival and demand. Without an order arrival, i.e., $A(t) = 0$, the next state will be one of the possible states that corresponds to the range $IL(t-1)$ and $IL(t-1) - a$, giving that $IL(t-1) - a \leq IL(t) \leq IL(t-1)$. The transition probability is $\frac{1}{a+1}$ for all possible states and 0, otherwise. The system becomes more complicated if order arrival is taken into account. The order decision depends on the reorder point r, the order quantity q and the lead-time L. It is important to note that $IL(t) \leq r + q$ as no multiple orders are allowed in the system. However, the lowest point that $IL(t)$ can reach is dependent on lead-time L and a. In turn, the lowest point also determines the total number of possible states N of the Markov chain. In what follows, we wish to find the total number of possible states N required for describing the inventory level $IL(t)$ in which r, q, a and L are decision variables.

4.3.1.1 The Total Number of Possible States N

States mean all possible outcomes of a Markov chain (Feller [42]). It is noted that some of the states are dependent of lead-time L while some are not. To see this point, let a Markov chain be in state S at period t, where state S corresponds to $IL(t) = S$. If $S \leq r$, that means state S can either move to one of the next states $S, S-1, \cdots, S-a$ with corresponding $IL(t+1) = S, S-1, \cdots, S-a$, provided that an order will not arrive in the next period; or move to one of the next states $S+q, S+q-1, \cdots, S+q-a$ with corresponding $IL(t+1) = S+q, S+q-1, \cdots, S+q-a$, provided that an order will arrive in the next period. However, if $S > r$, then state S is independent of L. Therefore, only for those states $S \leq r$, we have to record the number of periods remain for the next order arrival as it affects the choice of next state. We use $IL(t) = S_\tau$ to define $IL(t) = S$ and there are $L - \tau$ periods remain for next order arrival. For instance, k_0 means $IL(t) = k$ and the next order will arrive L periods later, while k_2 means $IL(t) = k$ and the next order will arrive $L - 2$ periods later. One important point to note is that N is increased by lead-time L.

It is worthy to see how τ affects the number of states required N by listing all possible outcomes of the Markov chain for a given set of r, q, a and L $(L > 0)$ as follows.

When there is no outstanding order, the possible states are:
$r+q, r+q-1, r+q-2, \cdots, r+1$.
When $\tau = 0$, the possible states (S_0) are: $r, r-1, r-2, \cdots, r-a+1$.
When $\tau = 1$, (S_1): $r, r-1, r-2, \cdots, r-a+1, r-a, r-a-1, \cdots, r-2a+1$.
When $\tau = 2$, (S_2): $r, r-1, r-2, \cdots, r-a+1, r-a, r-a-1, \cdots,$
$r-2a+1, r-2a, r-2a-1, \cdots, r-3a+1$.
...
When $\tau = L-1$, (S_{L-1}):
$r, r-1, r-2, \cdots, r-a+1, r-a, r-a-1, \cdots, r-2a+1, r-2a, r-2a-1, \cdots, r-3a+1, \cdots, r-(L-1)a, r-(L-1)a-1, \cdots, r-La+1$.

Table 4.1 Number of states required

τ	Number of states required
No outstanding order	q
0	a
1	$2a$
2	$3a$
...	...
$L-1$	La

Table 4.1 summarizes our findings. According to the table, the total number of states required is equal to

$$N = q + \frac{L(L+1)a}{2}. \tag{4.3.5}$$

It implies that we need to develop an $N \times N$ transition matrix $\mathbf{P}$ to account for all transition probabilities of the inventory level.

Although the current numbering system is highly sophisticated and it involves negative numbers, it is an essential process for us to understand all elements and states that constitute the transition matrix $\mathbf{P}$. After completing this process, it is more convenient to convert them into numerical order for further calculations and analyses. However, a table should be kept for cross-referencing purposes.

In the following, we give two examples to demonstrate how our system works. In the first one, we assume the lead-time is zero while the lead-time is non-zero in the second example.

Example 4.3.1 If $q = 3, a = 1, r = 0$ and $L = 0$, then $N = 3$ according to (4.3.5). Since $L = 0$, we omit the subscript τ. The required states are explained as follows.

State 1: If $IL(t)$ is in State 1 ($IL(t) = 3$), then it can either keep in State 1 ($IL(t+1) = 3$) or move to State 2 ($IL(t+1) = 2$) in the next period with $p_{11} = p_{12} = \frac{1}{2}$ and $p_{13} = 0$.

State 2: If $IL(t)$ is in State 2 ($IL(t) = 2$), then it can either keep in State 2 ($IL(t+1) = 2$) or move to State 3 ($IL(t+1) = 1$) in the next period with $p_{22} = p_{23} = \frac{1}{2}$ and $p_{21} = 0$.

State 3: If $IL(t)$ is in State 3 ($IL(t) = 1$), then it can either keep in State 3 ($IL(t+1) = 1$) or move to State 1 ($IL(t+1) = 3$) in the next period with $p_{33} = p_{31} = \frac{1}{2}$ and $p_{32} = 0$. It can move to State 1 ($IL(t+1) = 3$) because inventory is replenished simultaneously when demand is 1 in State 3. When it is in State 1, then the process repeats so that there is no other possible states missing.

The transition matrix $\mathbf{P}$ is

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{bmatrix} = \begin{bmatrix} 0.5 & 0.5 & 0 \\ 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \end{bmatrix}.$$

Example 4.3.2 If $q = 3, a = 1, r = 0$ and $L = 2$, then $N = 6$. States are explained as follows.

State 1: If $IL(t)$ is in State 1 ($IL(t) = 3$), then it can either keep in State 1 ($IL(t+1) = 3$) or move to State 2 ($IL(t+1) = 2$) in the next period with $p_{11} = p_{12} = \frac{1}{2}$ and $p_{13} = p_{14} = p_{15} = p_{16} = 0$.

State 2: If $IL(t)$ is in State 2 ($IL(t) = 2$), then it can either keep in State 2 ($IL(t+1) = 2$) or move to State 3 ($IL(t+1) = 1$) in the next period with $p_{22} = p_{23} = \frac{1}{2}$ and $p_{21} = p_{24} = p_{25} = p_{26} = 0$.

State 3: If $IL(t)$ is in State 3 ($IL(t) = 1$), then it can either keep in State 3 ($IL(t+1) = 1$) or move to State 4 ($IL(t+1) = 0_0$) in the next period with $p_{33} = p_{34} = \frac{1}{2}$ and $p_{31} = p_{32} = p_{35} = p_{36} = 0$. A subscript is used as a counter to indicate the number of periods from an order placement when there is no inventory in the system. If the subscript is equal to 0, it means the order is just placed. If the subscript is equal to 1, it means the order is placed 1 period before.

State 4: If $IL(t)$ is in State 4 ($IL(t) = 0_0$), then it can either move to State 5 ($IL(t+1) = 0_1$) or State 6 ($IL(t+1) = -1_1$) in the next period with $p_{45} = p_{46} = \frac{1}{2}$ and $p_{41} = p_{42} = p_{43} = p_{44} = 0$.

State 5: If $IL(t)$ is in State 5 ($IL(t) = 0_1$), then it can either move to State 1 if demand is 0, or State 2 if demand is 1 in the next period with $p_{51} = p_{52} = \frac{1}{2}$ and $p_{53} = p_{54} = p_{55} = p_{56} = 0$.

State 6: If $IL(t)$ is in State 6 ($IL(t) = -1_1$), then it can either move to State 2 if demand is 0, or State 3 if demand is 1 in the next period with $p_{62} = p_{63} = \frac{1}{2}$ and $p_{61} = p_{64} = p_{65} = p_{66} = 0$. This is the last state to consider as the system starts to repeat.

The transition matrix **P** is

$$\mathbf{P} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} & p_{15} & p_{16} \\ p_{21} & p_{22} & p_{23} & p_{24} & p_{25} & p_{26} \\ p_{31} & p_{32} & p_{33} & p_{34} & p_{35} & p_{36} \\ p_{41} & p_{42} & p_{43} & p_{44} & p_{45} & p_{46} \\ p_{51} & p_{52} & p_{53} & p_{54} & p_{55} & p_{56} \\ p_{61} & p_{62} & p_{63} & p_{64} & p_{65} & p_{66} \end{bmatrix}$$

$$= \begin{bmatrix} 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 & 0 & 0 & 0 \\ 0 & 0.5 & 0.5 & 0 & 0 & 0 \end{bmatrix}.$$

It is interesting to note that the transition matrix (**P**) is independent of the reorder point r. In fact, only the corresponding inventory levels depend on r. To see this point, we can revisit Example 4.3.2 for a general r. The six states are presented below.

State 1: If $IL(t)$ is in State 1 ($IL(t) = r + 3$), then it can either keep in State 1 ($IL(t) = r + 3$) or move to State 2 ($IL(t) = r + 2$) in the next period with $p_{11} = p_{12} = \frac{1}{2}$ and $p_{13} = p_{14} = p_{15} = p_{16} = 0$.

State 2: If $IL(t)$ is in State 2 ($IL(t) = r + 2$), then it can either keep in State 2 ($IL(t) = r + 2$) or move to State 3 ($IL(t) = r + 1$) in the next period with $p_{22} = p_{23} = \frac{1}{2}$ and $p_{21} = p_{24} = p_{25} = p_{26} = 0$.

State 3: If $IL(t)$ is in State 3 ($IL(t) = r + 1$), then it can either keep in State 3 ($IL(t) = r + 1$) or move to State 4 ($IL(t) = r_0$) in the next period with $p_{33} = p_{34} = \frac{1}{2}$ and $p_{31} = p_{32} = p_{35} = p_{36} = 0$.

State 4: If $IL(t)$ is in State 4 ($IL(t) = r_0$), then it can either move to State 5 ($IL(t) = r_1$) or State 6 ($IL(t) = (r - 1)_1$) in the next period with $p_{45} = p_{46} = \frac{1}{2}$ and $p_{41} = p_{42} = p_{43} = p_{44} = 0$.

State 5: If $IL(t)$ is in State 5 ($IL(t) = r_1$), then it can either move to State 1 if demand is 0, or State 2 if demand is 1 in the next period with $p_{51} = p_{52} = \frac{1}{2}$ and $p_{53} = p_{54} = p_{55} = p_{56} = 0$.

State 6: If $IL(t)$ is in State 6 ($IL(t) = (r-1)_1$), then it can either move to State 2 if demand is 0, or State 3 if demand is 1 in the next period with $p_{62} = p_{63} = \frac{1}{2}$ and $p_{61} = p_{64} = p_{65} = p_{66} = 0$. This is the last state to consider as the system starts to repeat.

It is obvious that the transition matrix **P** is as the same as that in the second example but the transition probabilities refer to a different inventory levels. We have the following theorem.

Theorem 4.3.3 *The transition matrix* (**P**) *and its number of states N are independent of the reorder point r.*

4.3.1.2 The Steady State Probabilities when $L = 0$

Let $\mathbf{P}^{(t)}$ be the transition matrix that passes through t transitions. In what follows, we wish to show that the transition matrix $\mathbf{P}^{(t)}$ reaches a steady state. Recall that there are totally four variables in describing the inventory system, namely r, q, a and L. However, r has been shown to be arbitrary in calculating N in Section 5.3.2. Without loss of generality, we can let $r = 0$ in the following analysis. In view of the fact that the system becomes more complex when $L > 0$, first we assume $L = 0$ so that $N = q$ by (4.3.5) and **P** is an $q \times q$ transition matrix.

A Markov chain has a limit, $\lim_{t\to\infty} \mathbf{P}^{(t)}$, and this limit is independent of the initial state if and only if it is an irreducible ergodic Markov chain (Feller [42]). Jenson and Bard [58], and Hillier and Lieberman [53] classified an irreducible chain as a chain that there is only one class that all states in the chain communicate. Specifically, any two states communicate if they are accessible from each other (Jenson and Bard [58]). An ergodic Markov chain is defined as a finite-state Markov chain that consists of recurrent states only and they are aperiodic (Hillier and Lieberman [53]). If all states can communicate with one another, then it is possible that a state can be revisited again after finite transitions. They are referred to as recurrent states. These states are aperiodic if they can return to themselves after a random number of transitions greater than 1 (Jenson and Bard [58]).

In our system, the inventory level moves from $r + q$ to any state in the Markov chain until an order arrives. Upon replenishment, there is a chance that the inventory level is back to $r + q$ from any state and starts to move to other states. Ross [100] stated that if a state is recurrent and it communicates with another state, say state j, then state j is also recurrent. $r + q$ is one recurrent state. It supports our view that the chain is irreducible and recurrent. Moreover, since the demand is stochastic (uniformly distributed), so is the inventory level. Therefore, the problem can be viewed as an irreducible ergodic Markov chain process so that a limit $\lim_{t\to\infty} \mathbf{P}^{(t)}$ exists. According to Jensen and Bard [58], **P** satisfies the following two conditions.

$$\pi(\mathbf{P} - \mathbf{I}) = \mathbf{0}, \tag{4.3.6}$$

$$\sum_{i=1}^{q} \pi_i = 1, \tag{4.3.7}$$

where π is the vector of the steady state probabilities and $\mathbf{I}$ is the $q \times q$ identity matrix.

Example 4.3.4 If $a = 3, q = 5$ and $\pi = [\pi_1\ \pi_2\ \pi_3\ \pi_4\ \pi_5]$, then $\pi(\mathbf{P} - \mathbf{I}) = \mathbf{0}$ can be written as

$$[\pi_1\ \pi_2\ \pi_3\ \pi_4\ \pi_5]\begin{bmatrix} -0.75 & 0.25 & 0.25 & 0.25 & 0 \\ 0 & -0.75 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0 & -0.75 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0 & -0.75 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0 & -0.75 \end{bmatrix} = 0,$$

$$[\pi_1\ \pi_2\ \pi_3\ \pi_4\ \pi_5]\begin{bmatrix} -3 & 1 & 1 & 1 & 0 \\ 0 & -3 & 1 & 1 & 1 \\ 1 & 0 & -3 & 1 & 1 \\ 1 & 1 & 0 & -3 & 1 \\ 1 & 1 & 1 & 0 & -3 \end{bmatrix} = 0.$$

Also, it can be expressed as five linear equations as follows.

$$\begin{aligned} -3\pi_1 + \pi_5 + \pi_4 + \pi_3 &= 0, \\ -3\pi_2 + \pi_1 + \pi_5 + \pi_4 &= 0, \\ -3\pi_3 + \pi_2 + \pi_1 + \pi_5 &= 0, \\ -3\pi_4 + \pi_3 + \pi_2 + \pi_1 &= 0, \\ -3\pi_5 + \pi_4 + \pi_3 + \pi_2 &= 0. \end{aligned}$$

The $6th$ equation is $\sum_{i=1}^{5} \pi_i = 1$. There are totally six equations and five unknowns. By eliminating one equation from the first five equations and by Gaussian Elimination, the solution is $\pi_1 = \pi_2 = \pi_3 = \pi_4 = \pi_5 = \frac{1}{5}$. In general, the two conditions can be expressed as

$$-a\pi_j + \sum_{i=1}^{a} (\pi_{j-i|j>i}, \pi_{j-i+q|j\leq i}) = 0, \tag{4.3.8}$$

$$\sum_{i=1}^{q} \pi_i = 1, \text{ for } j = 1, 2, \cdots, q. \tag{4.3.9}$$

From (4.3.8), $a\pi_j = \sum_{i=1}^{a} (\pi_{j-i|j>i}, \pi_{j-i+q|j\leq i})$. There are a terms in both sides. Let $\pi_j = \frac{1}{q}$ for all $j = 1, 2, \cdots, q$. It satisfies both Equations (4.3.8) and (4.3.9). This solution is necessarily unique [58, 42] and independent of a because no matter how many choices a present state can move to the next state, the probability that the present state is one of the q states is equal, provided that replenishment lead-time L is 0. In fact, this Markov chain is also known as a doubly stochastic matrix as the sum of any row and sum of any column is equal to 1. One property of a doubly stochastic matrix is that its steady state probability for any π_j is equal to the reciprocal of the number of states, $\frac{1}{N}$ (Feller [42]). It supports our finding: for $L = 0$, the steady state probability for any $\pi_j = \frac{1}{q}$, which is independent of a.

In fact, this result can be examined numerically. Recall from our example when $a = 3$ and $q = 5$. The initial transition matrix $\mathbf{P}^{(0)}$ is expressed below.

$$\mathbf{P}^{(0)} = \begin{bmatrix} 0.25 & 0.25 & 0.25 & 0.25 & 0 \\ 0 & 0.25 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0 & 0.25 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0 & 0.25 & 0.25 \\ 0.25 & 0.25 & 0.25 & 0 & 0.25 \end{bmatrix}$$
$$= 0.25 \begin{bmatrix} 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \end{bmatrix}.$$

$$\mathbf{P}^{(2)} = 0.25^3 \begin{bmatrix} 13 & 13 & 12 & 13 & 13 \\ 13 & 13 & 13 & 12 & 13 \\ 13 & 13 & 13 & 13 & 12 \\ 12 & 13 & 13 & 13 & 13 \\ 13 & 12 & 13 & 13 & 13 \end{bmatrix}.$$

$$\mathbf{P}^{(20)} = 0.2 \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 \end{bmatrix}.$$

Define $p_{ij}^{(t)} = \mathbf{P}^{(t)}$. As $t \to \infty$, $p_{ij}^{(t)} = 0.2$ for all i and j. It implies the Markov chain is stabilized. When the system is in steady state, regardless of the initial state, the probability that the system is in one of the possible states is $\frac{1}{q}$. It is equivalent to our previous conclusion. We have the following theorem (Feller [42]).

Theorem 4.3.5 *For $L = 0$, the steady state probability for any $\pi_j = \frac{1}{q}$, $j = 1, 2, \cdots, q$.*

4.3.1.3 The Steady State Probabilities when $L \geq 1$

The core of this section is to analyze if the system reaches a steady state. That is, $\pi(\mathbf{P} - \mathbf{I}) = \mathbf{0}$. In other words, it is to prove that 1 is an eigenvalue of $\mathbf{P}$.

Definition 4.3.6 A nonnegative matrix is stochastic if all its row sums or column sums are equal to 1 (Bronson[16]).

Proposition 4.3.7 *A stochastic $n \times n$ matrix $\mathbf{A}$ has an eigenvalue equal to 1.*

Proof Without loss of generality, we let the sum of any row of $\mathbf{A}$ be 1. Then the sum of any row of $(\mathbf{A} - \mathbf{I})$ is zero. Further, we let a set of column vectors, denoted as $\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_n$, to represent the corresponding columns of **A-I**. $\mathbf{A}_1 + \mathbf{A}_2 + \cdots + \mathbf{A}_n = \mathbf{0}$. It means that they are linearly dependent. So det(**A-I**)=0. It proves that $\mathbf{A}$ has an eigenvalue equal to 1.

Since $\mathbf{P}$ is a stochastic square matrix, it has an eigenvalue equal to 1. It implies $\pi\mathbf{P} = \pi$ and thus $\pi(\mathbf{P} - \mathbf{I}) = \mathbf{0}$, for some non-zero row vectors π. We can take a $\pi \neq 0$ such that the two conditions (Equations (4.3.6) and (4.3.7)) from Jensen and Bard [58] are still applicable in the case when $L > 0$ after normalization. That is,

$$\pi(\mathbf{P} - \mathbf{I}) = \mathbf{0}, \tag{4.3.10}$$

$$\sum_{i=1}^{N} \pi_i = 1, \tag{4.3.11}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix.

In order to combine these two conditions, we introduce an $(N+1) \times N$ augmented matrix, **P'**, whose elements are presented below [58].

$$\mathbf{P'} = \left[\begin{array}{cccccc} 1 & p_{12} & p_{13} & p_{14} & \cdots & p_{1N} \\ 1 & p_{22}-1 & p_{23} & p_{24} & \cdots & p_{2N} \\ 1 & p_{32} & p_{33}-1 & p_{34} & \cdots & p_{3N} \\ 1 & p_{42} & p_{43} & p_{44}-1 & \cdots & p_{4N} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \\ 1 & p_{N2} & p_{N3} & p_{N4} & \cdots & p_{NN}-1 \\ \hline 1 & 0 & 0 & 0 & \cdots & 0 \end{array}\right].$$

The advantage of using **P'** is that (4.3.10) and (4.3.11) can be solved directly by performing Gaussian Elimination (column operations). We begin our analysis by letting $L = 1$ and $a = 1$. Again, $r = 0$ without any loss of generality. By assumption, $q \geq 1$. Let $q = 5$, a 7×6 matrix **P'**, is expressed below.

$$\mathbf{P'} = \left[\begin{array}{cccccc} 1 & 0.5 & 0 & 0 & 0 & 0 \\ 1 & -0.5 & 0.5 & 0 & 0 & 0 \\ 1 & 0 & -0.5 & 0.5 & 0 & 0 \\ 1 & 0 & 0 & -0.5 & 0.5 & 0 \\ 1 & 0 & 0 & 0 & -0.5 & 0.5 \\ 1 & 0.5 & 0 & 0 & 0 & -1 \\ \hline 1 & 0 & 0 & 0 & 0 & 0 \end{array}\right].$$

After performing the Gaussian Elimination,

$$\mathbf{P'} = \left[\begin{array}{cccccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 \\ 1 & 1 & 3 & 0 & 0 & 0 \\ 1 & 1 & 1 & 4 & 0 & 0 \\ 1 & 1 & 1 & 1 & 5 & 0 \\ 1 & 0 & 0 & 0 & 0 & 10 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right].$$

The solution is $\pi_1 = \frac{1}{10}, \pi_2 = \frac{2}{10}, \pi_3 = \frac{2}{10}, \pi_4 = \frac{2}{10}, \pi_5 = \frac{2}{10}, \pi_6 = \frac{1}{10}$. This means $Pr\{IL(\infty) = 5\} = Pr\{IL(\infty) = 0\} = \frac{1}{10}, Pr\{IL(\infty) = 4\} = Pr\{IL(\infty) = 3\} = Pr\{IL(\infty) = 2\} = Pr\{IL(\infty) = 1\} = \frac{2}{10}$ when the system is in steady state.

When $q = 6$, **P'** becomes

$$\mathbf{P'} = \left[\begin{array}{ccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 3 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 4 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 5 & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & 6 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 12 \\ \hline 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{array}\right].$$

The solution is $\pi_1 = \pi_7 = \frac{1}{12}, \pi_2 = \pi_3 = \pi_4 = \pi_5 = \pi_6 = \frac{2}{12}$. This means $Pr\{IL(\infty) = 6\} = Pr\{IL(\infty) = 0\} = \frac{1}{12}, Pr\{IL(\infty) = 5\} = Pr\{IL(\infty) = 4\} = Pr\{IL(\infty) = 3\} = Pr\{IL(\infty) = 2\} = Pr\{IL(\infty) = 1\} = \frac{2}{12}$ when the system is in steady state.

In general, it is easy to see that

$$\mathbf{P'} = \underbrace{\left[\begin{array}{cccccccc} 1 & 0.5 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & -0.5 & 0.5 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & -0.5 & 0.5 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & -0.5 & 0.5 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & -0.5 & \cdots & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 0 & 0 & 0 & 0 & \cdots & -0.5 & 0.5 \\ 1 & 0.5 & 0 & 0 & 0 & \cdots & 0 & -1 \\ \hline 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \end{array}\right]}_{q+1 \text{ columns}}.$$

Inductively, using Gaussian Elimination (column operations), we obtain

$$\mathbf{P'} = \left[\begin{array}{cccccccc} 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 3 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 1 & 4 & 0 & \cdots & 0 & 0 \\ 1 & 1 & 1 & 1 & 5 & \cdots & 0 & 0 \\ 1 & 1 & 1 & 1 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & 1 & 1 & 1 & 1 & \cdots & q & 0 \\ 1 & 0 & 0 & 0 & 0 & \cdots & 0 & 2q \\ \hline 1 & 1 & 1 & 1 & 1 & \cdots & 1 & 1 \end{array}\right].$$

By using backward substitution, the solution is $\pi_1 = \pi_{q+1} = \frac{1}{2q}, \pi_i = \frac{1}{q}$ for $i = 2, 3, \cdots, q$. This means $Pr\{IL(\infty) = q\} = Pr\{IL(\infty) = 0\} = \frac{1}{2q}$ and $Pr\{IL(\infty) = i\} = \frac{1}{q}$ for $i = 1, 2, 3, \cdots, q-1$ when the system is in steady state.

However, starting from $L = 1$ and $a = 2$, there is no such pattern for analysis. Therefore, linear equations are developed and solved directly. In fact, we generated another eleven data sets and solve them in order to find a general pattern of the steady state probabilities. Interestingly, there are some general patterns found and they are useful for expanding our model in a more generic form. Two tables are presented below. The first one (Table 4.2) is a summary table of the required states N and their corresponding inventory levels of the 13 data sets. The second one (Table 4.3) is a summary table of the probabilities of the thirteen data sets.

Table 4.2 The 13 data sets and their required states

Set	1	2	3	4	5	6	7	8	9	10	11	12	13
L	1	1	1	1	1	2	2	2	2	2	3	3	3
a	1	1	2	2	3	1	1	2	2	3	1	1	2
q	5	6	5	6	6	5	6	6	10	10	6	7	7
π_1	5	6	5	6	6	5	6	6	10	10	6	7	7
π_2	4	5	4	5	5	4	5	5	9	9	5	6	6
π_3	3	4	3	4	4	3	4	4	8	8	4	5	5
π_4	2	3	2	3	3	2	3	3	7	7	3	4	4
π_5	1	2	1	2	2	1	2	2	6	6	2	3	3
π_6	0	1	0	1	1	0_0	1	1	5	5	1	2	2
π_7	–	0	–1	0	0	0_1	0_0	0_0	4	4	0_0	1	1
π_8	–	–	–	–1	–1	-1_1	0_1	0_1	3	3	0_1	0_0	0_0
π_9	–	–	–	–	–2	–	-1_1	-1_0	2	2	0_2	0_1	0_1
π_{10}	–	–	–	–	–	–	–	-1_1	1	1	-1_1	0_2	0_2
π_{11}	–	–	–	–	–	–	–	-2_1	0_0	0_0	-1_2	-1_1	-1_0
π_{12}	–	–	–	–	–	–	–	-3_1	0_1	0_1	-2_2	-1_2	-1_1
π_{13}	–	–	–	–	–	–	–	–	-1_0	-1_0	–	-2_2	-1_2
π_{14}	–	–	–	–	–	–	–	–	-1_1	-1_1	–	–	-2_1
π_{15}	–	–	–	–	–	–	–	–	-2_1	-2_0	–	–	-2_2
π_{16}	–	–	–	–	–	–	–	–	-3_1	-2_1	–	–	-3_1
π_{17}	–	–	–	–	–	–	–	–	–	-3_1	–	–	-3_2
π_{18}	–	–	–	–	–	–	–	–	–	-4_1	–	–	-4_2
π_{19}	–	–	–	–	–	–	–	–	–	-5_1	–	–	-5_2

Table 4.3 The probability distributions (in terms of N) of the 13 data sets

Set	1	2	3	4	5	6	7	8	9	10	11	12	13
L	1	1	1	1	1	2	2	2	2	2	3	3	3
a	1	1	2	2	3	1	1	2	2	3	1	1	2
q	5	6	5	6	6	5	6	6	10	10	6	7	7
π_1	$\frac{1}{10}$	$\frac{1}{12}$	$\frac{1}{15}$	$\frac{1}{18}$	$\frac{1}{24}$	$\frac{1}{20}$	$\frac{1}{24}$	$\frac{1}{54}$	$\frac{1}{90}$	$\frac{1}{160}$	$\frac{1}{48}$	$\frac{1}{56}$	$\frac{1}{189}$
π_2	$\frac{2}{10}$	$\frac{2}{12}$	$\frac{2}{15}$	$\frac{2}{18}$	$\frac{2}{24}$	$\frac{3}{20}$	$\frac{3}{24}$	$\frac{3}{54}$	$\frac{3}{90}$	$\frac{3}{160}$	$\frac{4}{48}$	$\frac{4}{56}$	$\frac{4}{189}$
π_3	$\frac{2}{10}$	$\frac{2}{12}$	$\frac{3}{15}$	$\frac{3}{18}$	$\frac{3}{24}$	$\frac{4}{20}$	$\frac{4}{24}$	$\frac{6}{54}$	$\frac{6}{90}$	$\frac{6}{160}$	$\frac{7}{48}$	$\frac{7}{56}$	$\frac{10}{189}$
π_4	$\frac{2}{10}$	$\frac{2}{12}$	$\frac{3}{15}$	$\frac{3}{18}$	$\frac{4}{24}$	$\frac{4}{20}$	$\frac{4}{24}$	$\frac{8}{54}$	$\frac{8}{90}$	$\frac{10}{160}$	$\frac{8}{48}$	$\frac{8}{56}$	$\frac{17}{189}$
π_5	$\frac{2}{10}$	$\frac{2}{12}$	$\frac{3}{15}$	$\frac{3}{18}$	$\frac{4}{24}$	$\frac{4}{20}$	$\frac{4}{24}$	$\frac{9}{54}$	$\frac{9}{90}$	$\frac{13}{160}$	$\frac{8}{48}$	$\frac{8}{56}$	$\frac{23}{189}$
π_6	$\frac{1}{10}$	$\frac{2}{12}$	$\frac{2}{15}$	$\frac{3}{18}$	$\frac{4}{24}$	$\frac{2}{20}$	$\frac{4}{24}$	$\frac{9}{54}$	$\frac{9}{90}$	$\frac{15}{160}$	$\frac{8}{48}$	$\frac{8}{56}$	$\frac{26}{189}$
π_7	–	$\frac{1}{12}$	$\frac{1}{15}$	$\frac{2}{18}$	$\frac{3}{24}$	$\frac{1}{20}$	$\frac{2}{24}$	$\frac{6}{54}$	$\frac{9}{90}$	$\frac{16}{160}$	$\frac{4}{48}$	$\frac{8}{56}$	$\frac{27}{189}$
π_8	–	–	–	$\frac{1}{18}$	$\frac{2}{24}$	$\frac{1}{20}$	$\frac{1}{24}$	$\frac{2}{54}$	$\frac{9}{90}$	$\frac{16}{160}$	$\frac{2}{48}$	$\frac{4}{56}$	$\frac{18}{189}$
π_9	–	–	–	–	$\frac{1}{24}$	–	$\frac{1}{24}$	$\frac{3}{54}$	$\frac{9}{90}$	$\frac{16}{160}$	$\frac{1}{48}$	$\frac{2}{56}$	$\frac{6}{189}$
π_{10}	–	–	–	–	–	–	–	$\frac{3}{54}$	$\frac{9}{90}$	$\frac{16}{160}$	$\frac{2}{48}$	$\frac{1}{56}$	$\frac{2}{189}$
π_{11}	–	–	–	–	–	–	–	$\frac{3}{54}$	$\frac{6}{90}$	$\frac{12}{160}$	$\frac{2}{48}$	$\frac{2}{56}$	$\frac{9}{189}$
π_{12}	–	–	–	–	–	–	–	$\frac{1}{54}$	$\frac{2}{90}$	$\frac{3}{160}$	$\frac{1}{48}$	$\frac{2}{56}$	$\frac{9}{189}$
π_{13}	–	–	–	–	–	–	–	–	$\frac{3}{90}$	$\frac{8}{160}$	–	$\frac{1}{56}$	$\frac{5}{189}$
π_{14}	–	–	–	–	–	–	–	–	$\frac{3}{90}$	$\frac{5}{160}$	–	–	$\frac{9}{189}$
π_{15}	–	–	–	–	–	–	–	–	$\frac{3}{90}$	$\frac{4}{160}$	–	–	$\frac{8}{189}$
π_{16}	–	–	–	–	–	–	–	–	$\frac{1}{90}$	$\frac{6}{160}$	–	–	$\frac{3}{189}$
π_{17}	–	–	–	–	–	–	–	–	–	$\frac{6}{160}$	–	–	$\frac{7}{189}$
π_{18}	–	–	–	–	–	–	–	–	–	$\frac{3}{160}$	–	–	$\frac{4}{189}$
π_{19}	–	–	–	–	–	–	–	–	–	$\frac{1}{160}$	–	–	$\frac{1}{189}$

Recall the number of states N is larger than or equal to the range of inventory level R, where $R = q + La$. If Table 4.3 is presented in terms of R instead of N, then the probability distribution becomes symmetric. A revised table is shown below (Table 4.4).

Based on Table 4.4, several important observations are obtained. We could not prove them, but it is interesting to state them as conjectures for future research.

(i) In general, the probability distribution graph is trapezium-like with lower base R and upper base $(q - La)$. It is symmetric (see Fig. 4.3 for details).

(ii) The total area of the trapezium is equal to 1. Area of A=Area of C=$\frac{La}{2q}$ and Area of B=$\frac{q-La}{q}$.

(iii) $La = 0$ only if $L = 0$. In this case, Area of A=Area of C=0. The probability distribution becomes a horizontal line whose height is equal to $\frac{1}{q}$. It is consistent

Table 4.4 The probability distributions (in terms of R) of the 13 data sets

Set	1	2	3	4	5	6	7	8	9	10	11	12	13
L	1	1	1	1	1	2	2	2	2	2	3	3	3
a	1	1	2	2	3	1	1	2	2	3	1	1	2
q	5	6	5	6	6	5	6	6	10	10	6	7	7
$Pr\{IL(\infty) = q\}$	$\frac{1}{10}$	$\frac{1}{12}$	$\frac{1}{15}$	$\frac{1}{18}$	$\frac{1}{24}$	$\frac{1}{20}$	$\frac{1}{24}$	$\frac{1}{54}$	$\frac{1}{90}$	$\frac{1}{160}$	$\frac{1}{48}$	$\frac{1}{56}$	$\frac{1}{189}$
$Pr\{IL(\infty) = q-1\}$	$\frac{2}{10}$	$\frac{2}{12}$	$\frac{2}{15}$	$\frac{2}{18}$	$\frac{2}{24}$	$\frac{3}{20}$	$\frac{3}{24}$	$\frac{3}{54}$	$\frac{3}{90}$	$\frac{3}{160}$	$\frac{4}{48}$	$\frac{4}{56}$	$\frac{4}{189}$
$Pr\{IL(\infty) = q-2\}$	$\frac{2}{10}$	$\frac{2}{12}$	$\frac{3}{15}$	$\frac{3}{18}$	$\frac{3}{24}$	$\frac{4}{20}$	$\frac{4}{24}$	$\frac{6}{54}$	$\frac{6}{90}$	$\frac{6}{160}$	$\frac{7}{48}$	$\frac{7}{56}$	$\frac{10}{189}$
$Pr\{IL(\infty) = q-3\}$	$\frac{2}{10}$	$\frac{2}{12}$	$\frac{3}{15}$	$\frac{3}{18}$	$\frac{4}{24}$	$\frac{4}{20}$	$\frac{4}{24}$	$\frac{8}{54}$	$\frac{8}{90}$	$\frac{10}{160}$	$\frac{8}{48}$	$\frac{8}{56}$	$\frac{17}{189}$
$Pr\{IL(\infty) = q-4\}$	$\frac{2}{10}$	$\frac{2}{12}$	$\frac{3}{15}$	$\frac{3}{18}$	$\frac{4}{24}$	$\frac{4}{20}$	$\frac{4}{24}$	$\frac{9}{54}$	$\frac{9}{90}$	$\frac{13}{160}$	$\frac{8}{48}$	$\frac{8}{56}$	$\frac{23}{189}$
$Pr\{IL(\infty) = q-5\}$	$\frac{1}{10}$	$\frac{2}{12}$	$\frac{2}{15}$	$\frac{3}{18}$	$\frac{4}{24}$	$\frac{3}{20}$	$\frac{4}{24}$	$\frac{9}{54}$	$\frac{9}{90}$	$\frac{15}{160}$	$\frac{8}{48}$	$\frac{8}{56}$	$\frac{26}{189}$
$Pr\{IL(\infty) = q-6\}$	–	$\frac{1}{12}$	$\frac{1}{15}$	$\frac{2}{18}$	$\frac{3}{24}$	$\frac{1}{20}$	$\frac{3}{24}$	$\frac{8}{54}$	$\frac{9}{90}$	$\frac{16}{160}$	$\frac{7}{48}$	$\frac{8}{56}$	$\frac{27}{189}$
$Pr\{IL(\infty) = q-7\}$	–	–	–	$\frac{1}{18}$	$\frac{2}{24}$	–	$\frac{1}{24}$	$\frac{6}{54}$	$\frac{9}{90}$	$\frac{16}{160}$	$\frac{4}{48}$	$\frac{7}{56}$	$\frac{26}{189}$
$Pr\{IL(\infty) = q-8\}$	–	–	–	–	$\frac{1}{24}$	–	–	$\frac{3}{54}$	$\frac{9}{90}$	$\frac{16}{160}$	$\frac{1}{48}$	$\frac{4}{56}$	$\frac{23}{189}$
$Pr\{IL(\infty) = q-9\}$	–	–	–	–	–	–	–	$\frac{1}{54}$	$\frac{9}{90}$	$\frac{16}{160}$	–	$\frac{1}{56}$	$\frac{17}{189}$
$Pr\{IL(\infty) = q-10\}$	–	–	–	–	–	–	–	–	$\frac{8}{90}$	$\frac{15}{160}$	–	–	$\frac{10}{189}$
$Pr\{IL(\infty) = q-11\}$	–	–	–	–	–	–	–	–	$\frac{6}{90}$	$\frac{13}{160}$	–	–	$\frac{4}{189}$
$Pr\{IL(\infty) = q-12\}$	–	–	–	–	–	–	–	–	$\frac{3}{90}$	$\frac{10}{160}$	–	–	$\frac{1}{189}$
$Pr\{IL(\infty) = q-13\}$	–	–	–	–	–	–	–	–	$\frac{1}{90}$	$\frac{6}{160}$	–	–	–
$Pr\{IL(\infty) = q-14\}$	–	–	–	–	–	–	–	–	–	$\frac{3}{160}$	–	–	–
$Pr\{IL(\infty) = q-15\}$	–	–	–	–	–	–	–	–	–	$\frac{1}{160}$	–	–	–

with our previous findings. On the other hand, if $La = q$, then Area of B=0 and the probability distribution graph is triangular in shape.

(iv) When $q >> La$, then both Area of A and Area of C tend to zero and the probability distribution becomes a horizontal line again whose height is equal to $\frac{1}{q}$. This finding is identical to the case when $La = 0$.

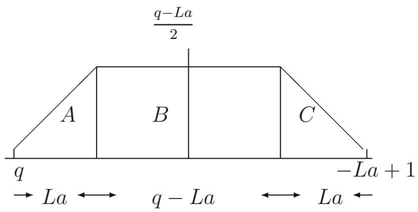

Fig. 4.3 A plot of probability distribution diagram

Although the shape of the probability distribution is estimated, it is more valuable if the steady state probability for any $IL(\infty)$ is derived exactly. A 2-step algorithm is introduced for this purpose. However, it is only a conjecture and only valid for L=0,1 and 2, and all a, q and r. The algorithm is fully explained in the following.

The Two-step Algorithm (Conjecture)

Step 1: Calculate $q(a+1)^L$. It is the denominator of the steady state probabilities for L=0,1 and 2.

Step 2: Derive the probability distribution.
If $L = 0$, then the probability distribution $Pr\{IL(\infty) = i\}$ is $\frac{1}{q}$ for $i = r+q, r+q-1, \cdots, r+1$ (Theorem 4.3.5).

If $L = 1$, then the probability distribution $Pr\{IL(\infty) = i\}$ is

$$\underbrace{\frac{1}{q(a+1)}, \frac{2}{q(a+1)}, \frac{3}{q(a+1)}, \cdots, \frac{a}{q(a+1)},}_{\text{first } a \text{ terms}}$$

for $i = r+q, r+q-1, \cdots, r+q-a+1$,

$$\underbrace{\frac{a+1}{q(a+1)}, \frac{a+1}{q(a+1)}, \cdots, \frac{a+1}{q(a+1)},}_{\text{next } q-a \text{ terms}}$$

for $i = r+q-a, r+q-a-1, \cdots, r+1$,

$$\underbrace{\frac{a}{q(a+1)}, \frac{a-1}{q(a+1)}, \cdots, \frac{1}{q(a+1)},}_{\text{last } a \text{ terms}} \text{ for } i = r, r-1, \cdots, r-a+1.$$

There are $R = q + a$ terms in total.
If $L = 2$, then the probability distribution $Pr\{IL(\infty) = i\}$ is

$$\underbrace{\frac{1}{q(a+1)^2}, \frac{1+2}{q(a+1)^2}, \frac{1+2+3}{q(a+1)^2}, \cdots, \frac{1+2+\cdots+a}{q(a+1)^2},}_{\text{first } a \text{ terms}}$$

for $i = r+q, r+q-1, \cdots, r+q-a+1$,

$$\underbrace{\frac{(a+1)^2-(1+2+\cdots+a)}{q(a+1)^2}, \cdots, \frac{(a+1)^2-1}{q(a+1)^2},}_{\text{next } a \text{ terms}}$$

for $i = r+q-a, r+q-a-1, \cdots, r+q-2a+1$,

$$\underbrace{\frac{(a+1)^2}{q(a+1)^2}, \frac{(a+1)^2}{q(a+1)^2}, \cdots, \frac{(a+1)^2}{q(a+1)^2},}_{\text{next } q-2a \text{ terms}}$$

for $i = r+q-2a, r+q-2a-1, \cdots, r+1$,

$$\underbrace{\frac{(a+1)^2-1}{q(a+1)^2}, \frac{(a+1)^2-(1+2)}{q(a+1)^2}, \cdots, \frac{(a+1)^2-(1+2+\cdots+a)}{q(a+1)^2}}_{\text{next } a \text{ terms}},$$

for $i = r, r-1, \cdots, r-a+1$,

$$\underbrace{\frac{1+2+\cdots+a}{q(a+1)^2}, \frac{1+2+\cdots+(a-1)}{q(a+1)^2}, \cdots, \frac{1}{q(a+1)^2}}_{\text{last } a \text{ terms}},$$

for $i = r, r-1, \cdots, r-2a+1$.

There are $R = q + 2a$ terms in total.

4.3.1.4 A Numerical Example

In this section, an example is used to demonstrate how the algorithm works to compute the probability distribution of each possible $IL(\infty)$ value. The accuracy of the algorithm is tested by solving the required linear equations. In this example, we let $L = 2, r = 0, a = 2$ and $q = 10$ so that $N = 16$ and $R = 14$. The required states are

$$(10, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0_0, 0_1, -1_0, -1_1, -2_1, -3_1).$$

The possible inventory levels are

$$(10, 9, 8, 7, 6, 5, 4, 3, 2, 1, 0, -1, -2, -3).$$

According to the algorithm,

Step 1: The denominator is $q(a+1)^L = (10)(2+1)^2 = 90$.

Step 2: The probability distribution solution is listed as follows.

The first 2 terms,
$Pr\{IL(\infty) = 10\} = \frac{1}{90}$ and $Pr\{IL(\infty) = 9\} = \frac{1+2}{90} = \frac{3}{90}$.

The next 2 terms,
$Pr\{IL(\infty) = 8\} = \frac{9-(1+2)}{90} = \frac{6}{90}$ and $Pr\{IL(\infty) = 7\} = \frac{9-1}{90} = \frac{8}{90}$.

The next 6 terms,
$Pr\{IL(\infty) = 6\} = Pr\{IL(\infty) = 5\} = Pr\{IL(\infty) = 4\} = Pr\{IL(\infty) = 3\} = Pr\{IL(\infty) = 2\} = Pr\{IL(\infty) = 1\} = \frac{9}{90}$.

The next 2 terms,
$Pr\{IL(\infty) = 0\} = \frac{9-1}{90} = \frac{8}{90}$ and $Pr\{IL(\infty) = -1\} = \frac{9-(1+2)}{90} = \frac{6}{90}$.

The last 2 terms,
$Pr\{IL(\infty) = -2\} = \frac{1+2}{90} = \frac{3}{90}$ and $Pr\{IL(\infty) = -3\} = \frac{1}{90}$.

Besides, the probability distribution can be solved by performing column operation of the following 17×16 matrix, denoted as **P'**.

$$\mathbf{P}' =$$

$$\left[\begin{array}{cccccccccccccccc}
1 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & \frac{-2}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & \frac{-2}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & \frac{-2}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & \frac{-2}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & \frac{-2}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & \frac{-2}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-2}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-2}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-2}{3} & \frac{1}{3} & 0 & \frac{1}{3} & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\
1 & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
1 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
1 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\
1 & 0 & 0 & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\
\hline
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{array}\right].$$

The probability of each state is summarized as follows.

N	10	9	8	7	6	5	4	3	2	1
Probability	$\frac{1}{90}$	$\frac{3}{90}$	$\frac{6}{90}$	$\frac{8}{90}$	$\frac{9}{90}$	$\frac{9}{90}$	$\frac{9}{90}$	$\frac{9}{90}$	$\frac{9}{90}$	$\frac{9}{90}$

N	0_0	0_1	-1_0	-1_1	-2_1	-3_1
Probability	$\frac{6}{90}$	$\frac{2}{90}$	$\frac{3}{90}$	$\frac{3}{90}$	$\frac{3}{90}$	$\frac{1}{90}$

By summing those states that are dependent of lead-time, say, $Pr\{IL(\infty) = 0\} = Pr\{IL(\infty) = 0_0\} + Pr\{IL(\infty) = 0_1\} = \frac{8}{90}$. It can be verified easily that the solution is identical to our previous conjecture.

4.3.2 The Algorithm for Finding a Near Optimal Total Average Cost

4.3.2.1 Pros and Cons of Using the Algorithm

In practice, a supply chain manager always performs various what-if analysis in order to plan for the future, such as inventory. Our proposed algorithm aims at helping managers to control their inventory levels with different (r, q) policies when facing different customer demands and order lead-times. Our algorithm is easy to use and it greatly reduces the complexity in solving a large number of linear equations, as N is compounded by L and a. For instance, if $L = 2$ and $a = 10$, then q should be at least 20 in accordance with the constraint $La \leq q$. The minimum N will be 50. It implies that there are 50 unknowns in total. It requires a significant amount of computation time to solve them. With our algorithm, the whole probability distribution is computed in only two steps. Besides, there is no need to develop a matrix before we can get the whole probability distribution for the required inventory level.

The major drawback of the algorithm is its accuracy because it is based on our findings rather than a mathematical proof. For validation, the algorithm should be tested for a larger number of samples so that it can support a wider range of L, a, r and q in high confidence. Another limitation is that it has an initial constraint $La \leq q$ for using the algorithm and the algorithm can be used only for integer inputs.

4.3.2.2 The Algorithm for Optimization of the Cost Function

The proposed algorithm is a useful tool for us to find the average backorder and average inventory for further analysis based on cost. Recall our objective function for the non-postponement system is $C(r,q) = k\overline{OF} + c\overline{D} + h\overline{I} + b\overline{B}$. When $L = 0$, the four variables, $\overline{OF}, \overline{D}, \overline{I}$ and $\overline{B}$ at steady state, can be expressed in terms of a, r and q as follows.

$$
\begin{aligned}
\overline{OF} &= \frac{a}{2q}.\\
\overline{D} &= \frac{a}{2}.\\
\overline{I} &= \sum_{i=0}^{r+q} iPr\{IL(\infty) = i\}\\
&= \sum_{i=r+1}^{r+q} iPr\{IL(\infty) = i\}\\
&= \frac{1}{q}\sum_{i=1}^{q}(r+i)\\
&= r + \frac{1+q}{2}.\\
\overline{B} &= 0.\\
C(r,q) &= \frac{ka}{2q} + \frac{ca}{2} + h\left(r + \frac{1+q}{2}\right).
\end{aligned}
$$

By differentiating $C(r,q)$ with respect to q and r,

$$\frac{\partial C(r,q)}{\partial r} = h > 0.$$

$$\frac{\partial C(r,q)}{\partial q} = -\frac{ka}{2q^2} + \frac{h}{2}.$$

Since $\frac{\partial C(r,q)}{\partial r}$ is always positive, $r^* = 0$. We have the following theorem.

Theorem 4.3.8 *When $L = 0$, the optimal reorder point r^* is zero for all a and q.*

Further, we let $\frac{\partial C(r,q)}{\partial q} = 0$,

$$q^* = \sqrt{\frac{ka}{h}} = \sqrt{\frac{2k\overline{D}}{h}}.$$

The optimal order quantity is equivalent to the well-known EOQ. Thus, the minimum average cost is

$$C(r^*, q^*) = \sqrt{hka} + \frac{1}{2}(ca + h). \tag{4.3.12}$$

For $L \geq 1$, the four variables, $\overline{OF}, \overline{D}, \overline{I}$ and $\overline{B}$, are expressed in terms of a, r and q as follows.

$$\overline{OF} = \frac{a}{2q}.$$
$$\overline{D} = \frac{a}{2}.$$
$$\overline{I} = \sum_{i=0}^{r+q} iPr\{IL(\infty) = i\}.$$
$$\overline{B} = \sum_{i=r-La+1}^{0} -iPr\{IL(\infty) = i\}.$$

As stated before, the probability distribution of $IL(\infty)$ is changing with different combinations of r, q and a, it is not easy to derive the minimum $C(r^*, q^*)$ directly. Therefore an iterative procedure is proposed for any given set of a and L. The heuristic is explained below.

A Heuristic for Searching for $C(r^*, q^*)$

Start the computation of q

- **(i)** Set $r = 0$ and $q = La$.
- **(ii)** Generate the probability distribution.
- **(iii)** Calculate $C(0, q)$.
- **(iv)** If $C(0, q-1) \geq C(0, q)$ and $C(0, q+1) \geq C(0, q)$, then $q = q^*$. Stop and goto the computation of r.

(v) $q = q + 1$. Goto **(ii)**.

Start the computation of r

(i) Set $r = 1$.
(ii) Calculate $C(r, q^*)$.
(iii) If $C(r-1, q^*) \geq C(r, q^*)$ and $C(r+1, q^*) \geq C(r, q^*)$, then $r = r^*$ and Stop. Otherwise $r = r + 1$ and Goto **(ii)**.

The iterative procedure starts with $r = 0$ because r does not affect the values of the probability distribution. If we increase r, the whole distribution will be shifted to the right and it implies that there are more safety stocks. When r is increased to a value larger than $La - 1$, unnecessary safety stocks are stacked. Thus, the upper bound of r is $La - 1$. On the other hand, if we keep increasing the value of q, again unwanted stocks are created. Therefore, there should be an optimal (r^*, q^*) that can result in a low average cost. However, our algorithm does not guarantee the solution is the global optimal. This is consistent with the finding from Browne and Zipkin [20] that there is no perfect method for solving an optimal (r, q) policy. In fact, it is not uncommon to search for an appropriate reorder point based on a fixed order quantity (see Matheus and Gelders [77], and Ng et al. [84]). The next section will present a numerical example for this algorithm.

4.3.2.3 A Numerical Example

In this example, we assume the four cost parameters are $k = 10, h = 1, b = 5$ and $c = 3$. Further we let $L = 1$ and $a = 10$ so that the initial q is 10. With initial $r = 0$ (the upper bound of r is 9), the iterative procedure for the searching of r^*, q^* and $C(r^*, q^*)$ is shown in Table 4.5. When r is restricted to zero, C decreases as q increases until $q = 17$. Then we fix $q = 17$ and relax r. When $r = 2$ and $q = 17$, a near optimal C is reached. That is, $C(2, 17) = 26.64$. The policy is: if the inventory level drops to 2, then place an order of quantity equal to 17. Twelve iterations are required to reach the desired solution.

4.4 System Dynamics for a Postponement System

As recalled, we treat a postponement system as a zero stock policy or a so-called make-to-order system. From the manufacturer's point of view, when demand from customer at period t, $D(t)$, arrives, we order $D(t)$ from the supplier at time t. The customer will receive the order at period $t + L$, where L is the order lead-time. The system dynamics for the postponement system are as follows.

$$IL(t) = A(t) - D(t) - B(t-1). \tag{4.4.13}$$

$$B(t) = -IL(t). \tag{4.4.14}$$

Table 4.5 Iterative procedure for $L = 1, a = 10$

Iteration	k	h	b	c	L	a	r	q	$\overline{OF}$	$\overline{D}$	$\overline{I}$	$\overline{B}$	C
1	10	1	5	3	1	10	0	10	0.5	5	2	1.5	29.5
2	10	1	5	3	1	10	0	11	0.4545	5	2.364	1.364	28.73
3	10	1	5	3	1	10	0	12	0.4167	5	2.75	1.25	28.17
4	10	1	5	3	1	10	0	13	0.3846	5	3.154	1.154	27.77
5	10	1	5	3	1	10	0	14	0.3571	5	3.571	1.071	27.50
6	10	1	5	3	1	10	0	15	0.3333	5	4	1	27.33
7	10	1	5	3	1	10	0	16	0.3125	5	4.438	0.9375	27.25
8	10	1	5	3	1	10	0	17*	0.2941	5	4.882	0.8824	27.24
9	10	1	5	3	1	10	0	18	0.2778	5	5.333	0.8333	27.28
10	10	1	5	3	1	10	1	17	0.2941	5	5.642	0.6417	26.79
11	10	1	5	3	1	10	2**	17	0.2941	5	6.449	0.4492	26.64***
12	10	1	5	3	1	10	3	17	0.2941	5	7.299	0.2995	26.74

Note: *, **, and *** are the optimal q, r, and total cost, respectively.

$$I(t) = 0. \tag{4.4.15}$$

The reorder decision $O(t) = A(t+L) = D(t)$ is made when $D(t) > 0$. Otherwise, $O(t) = A(t+L) = 0$.

The cost function for the postponement system is

$$C = k\overline{OF} + c\overline{D} + h\overline{I} + b\overline{B}. \tag{4.4.16}$$

Since the system is a make-to-order system, there is no on-hand inventory and all orders are backlogged. $\overline{I} = 0$ and $\overline{B} = L\overline{D}$ because there are L orders outstanding on average. The average order frequency $\overline{OF}$ is $\frac{a}{a+1}$ while the average demand $\overline{D}$ is $\frac{a}{2}$. Substitute them into (4.4.16),

$$C = \frac{ka}{a+1} + \frac{(c+Lb)a}{2}. \tag{4.4.17}$$

It is interesting to evaluate whether or not this postponement (make-to-order) system outperforms the non-postponement (r, q) system based on average cost. For $L = 0$, it can be analyzed by subtracting equation (4.4.17) from (4.3.12) directly. In fact, we compare the optimal (r^*, q^*) system with the postponement system. However, there is no exact equation for computing an optimal $C(r^*, q^*)$ when $L \geq 1$. We employ our heuristic and simulation to show the difference between the two systems. They will be fully discussed in the sequel.

4.5 Average Cost Comparison of the Two Systems When $L = 0$

In this section, we let the difference between (4.4.17) and (4.3.12) be Z. Z is expressed as follows.

$$Z = \sqrt{hka} + \frac{1}{2}(ca + h) - \frac{ka}{a+1} - \frac{ca}{2},$$
$$= \sqrt{hka} + \frac{h}{2} - \frac{ka}{a+1}.$$

The result indicates that the postponement system results in a lower total average cost in a wide variety of cases except when the fixed ordering cost k is relatively larger than the inventory holding cost h and the maximum demand per period a.

4.6 Average Cost Comparison of the Two Systems When $L \geq 1$

The simulation was run with the following truncated parameters.

L : 1, 2
a : 3 (low demand variability), 10 (high demand variability)
k : 0.1 (low fixed ordering cost), 15 (high fixed ordering cost)
c : 3 (low variable cost), 15 (high variable cost)
h : 1
b : 0.1 (low backorder cost), 15 (high backorder cost)

The values of k, h and b are based on a ratio analysis in which h is the base value. In our analysis, there were totally 32 cases. In each case, a near optimal $C(r^*, q^*)$ was found by employing our heuristic. Then this near optimal average cost was used to compare with the average cost of a postponement system. The percentage difference is computed by the following equation.

$$\frac{C - C(r^*, q^*)}{C(r^*, q^*)} \times 100\%, \qquad (4.6.18)$$

where

(i) C is obtained from (4.4.17), and
(ii) (4.6.18) is negative when the postponement system results in lower average cost and vice versa.

The results of the 32 cases are summarized in Table 4.6.

4.6.1 An Overview of the Simulation Results

Based on the above results, only eight cases reported a lower total average cost when the postponement system is adopted. In other words, the system only had a

Table 4.6 A summary of the simulation results

Case	L	a	k	h	b	c	r^*	q^*	$C(r^*, q^*)$	C	(%)
1	1	3	0.1	1	0.1	3	0	3	5.417	4.725	−12.8
2	1	3	0.1	1	0.1	15	0	3	23.42	22.73	−3.0
3	1	3	0.1	1	15	3	1	6	8.19	27.08	230.5
4	1	3	0.1	1	15	15	1	6	26.19	45.08	72.1
5	1	3	15	1	0.1	3	0	7	10.37	15.90	53.3
6	1	3	15	1	0.1	15	0	7	28.37	33.90	19.5
7	1	3	15	1	15	3	1	9	11.94	38.25	220.2
8	1	3	15	1	15	15	1	9	29.94	56.25	87.8
9	1	10	0.1	1	0.1	3	0	10	17.20	15.59	−9.4
10	1	10	0.1	1	0.1	15	0	10	77.20	75.59	−2.1
11	1	10	0.1	1	15	3	4	22	27.84	90.09	223.6
12	1	10	0.1	1	15	15	4	22	87.84	150.1	70.9
13	1	10	15	1	0.1	3	0	14	24.04	29.14	21.2
14	1	10	15	1	0.1	15	0	14	84.04	89.14	6.1
15	1	10	15	1	15	3	4	25	32.04	103.6	223.5
16	1	10	15	1	15	15	4	25	92.04	163.6	77.8
17	2	3	0.1	1	0.1	3	0	6	5.804	4.875	−16.0
18	2	3	0.1	1	0.1	15	0	6	23.80	22.88	−3.9
19	2	3	0.1	1	15	3	2	12	11.26	49.58	340.2
20	2	3	0.1	1	15	15	2	12	29.26	67.58	130.9
21	2	3	15	1	0.1	3	0	7	9.382	16.05	71.1
22	2	3	15	1	0.1	15	0	7	27.38	34.05	24.4
23	2	3	15	1	15	3	2	13	13.39	60.75	353.9
24	2	3	15	1	15	15	2	13	31.39	78.75	150.9
25	2	10	0.1	1	0.1	3	0	20	18.55	16.09	−13.3
26	2	10	0.1	1	0.1	15	0	20	78.55	76.09	−3.1
27	2	10	0.1	1	15	3	10	36	36.01	165.1	358.4
28	2	10	0.1	1	15	15	7	42	97.58	225.1	130.7
29	2	10	15	1	0.1	3	0	20	22.28	29.64	33.0
30	2	10	15	1	0.1	15	0	20	82.28	89.64	8.9
31	2	10	15	1	15	3	8	44	40.21	178.6	344.2
32	2	10	15	1	15	15	8	44	100.2	238.6	138.1

0.25 probability to outperform the non-postponement system. The maximum saving was 16% (case 17), while the minimum saving was only 2.1% (case 10). The saving is relatively smaller than those cases when a non-postponement system is preferred. These eight cases have two characteristics in common: low fixed ordering cost and backorder cost. Particularly, by comparing case 1 with case 17, it was found that when lead-time increases, the total cost saving also increases, given that other parameters remain unchanged. Another finding was that when demand variability (measured by a) becomes larger, the saving yielded becomes smaller. It was supported by comparing case 1 with case 9, case 2 with case 10, case 17 with case 25, and case 18 with case 26. In sum, a postponement system performs better when fixed ordering cost k, variable cost c and backorder cost b are small, demand variability a is low and lead-time is long.

Among those 24 cases that showed the non-postponement system is preferred, eight of them led to an absolute cost advantage as the percentage difference was at least 220%. They were case 3, case 7, case 11, case 15, case 19, case 23, case 27 and case 31. In fact, they share two characteristics in common: low variable cost and high backorder cost. The maximum saving was 358.4% when adopting a non-postponement system. It is a solid evidence to support that this system yields more total average cost savings in a wide variety of cases (75% in our study). A brief investigation of all possible impacts of the parameters on the average cost of both postponement system and non-postponement system are explained in detail in the next section.

4.6.2 Impacts of Parameters on Average Cost

4.6.2.1 Impacts of Lead-time L

In the non-postponement system, longer lead-time induces the system to order more at one time and to maintain a higher safety stock level. It implies higher r and q values, and a higher inventory holding cost. Also, a longer lead-time means a higher backorder risk because it creates more uncertainties to future demand and supply. Therefore, the average backorder cost increases so does the total average cost. On the other hand, the average backorder cost for the postponement system is proportional to L. As both system costs depend on L, it is worth evaluating when a postponement system outweighs a non-postponement system. In Section 6, it has been proved that it is more advantageous to adopt a postponement system when $L = 0$, except k is large. When $L \geq 1$, a postponement system is still preferred when k and b are small (0.1 in our study).

4.6.2.2 Impacts of Demand Variability a

Demand variability is measured by the maximum demand per period a as the demand distribution is uniformly distributed. As stated before, a postponement system performs better when coping with high demand variability, given k and b is small.

4.6.2.3 Impacts of Inventory Holding Cost h

Remember h is the base value for calculating appropriate k and b based on different ratios. Its relative impacts are explained in terms of k and b. For example, $h = 1$ is relatively high if $k = 0.1$ and relatively low if $k = 15$.

4.6.2.4 Impacts of Fixed Ordering Cost k

It is the most significant factor for addressing the advantage of a non-postponement system. Among the sixteen cases when the fixed ordering cost was high ($k = 15$), there was no case reported that the postponement system is more cost effective. It is identical to our findings when $L = 0$ in Section 6 that when k is large, it is more cost effective to adopt a non-postponement system. The major reason is that when k is large, there is a higher incentive to order more at one time in order to avoid such high cost. However, there is no way for a postponement system to adjust the ordering pattern since it orders every period.

4.6.2.5 Impacts of Backorder Cost *b*

According to our findings, the postponement system is highly sensitive to the backorder cost because all orders received are fully backlogged. In contrast, only a proportion of customer orders is backlogged in the non-postponement system. In facing high backorder cost, the non-postponement system is more cost effective.

4.6.2.6 Impacts of Variable Cost *c*

c is a special parameter for analyzing the cost impact. When c increases, while other parameters remain unchanged, both $C(r^*, q^*)$ and C increase significantly. In case the postponement system results in a lower average cost, the percentage in saving decreases as c increases. On the other hand, if a non-postponement system results in a lower average cost, the percentage saved decreases as c increases. In fact, this point can be explained by revisiting equation (4.6.18). Since $c\overline{D}$ is identical for both $C(r^*, q^*)$ and C, it is cancelled out in the nominator but it remains in the denominator. Therefore, it leads to a lower percentage no matter it is positive or negative. But it does not change the sign.

Based on our analysis, a postponement system outperforms a non-postponement system when $L = 0$ and k is relatively smaller than a and h, or both k and b are relatively smaller than a and h when $L \geq 1$. Otherwise, a non-postponement system should be adopted. In sum, this comparison can be viewed as a trade-off between h against b and k.

4.7 Concluding Remarks

The objective of this chapter is to evaluate a postponement system based on a total average cost function. To achieve it, an optimal total average cost (when $L = 0$) or a near optimal total average cost (when $L \geq 1$) of a lot size-reorder point system was computed and was used for comparison basis. In both systems, it was assumed that customer demand is uniformly distributed and discrete in time, and order lead-time is constant. Optimization technique was employed when lead-time is zero and simulation technique was used when lead-time is non-zero. Our results indicate that the non-postponement system is more cost effective in a wide variety of cases in dealing with a single end-product under stochastic demand.

There are some limitations in this chapter. The first one is the assumption that there is no more than one outstanding order for the non-postponement system. It may not be valid in practice. Besides, we simplify the postponement system to a pure make-to-order system of a single end-product so that there is no point of product differentiation. It limits the capability of a postponement system to deal with product variety. To be more equitable, a multi-product model is developed in the next chapter so that there is a point of product differentiation in a supply chain. Also we attempt to evaluate the postponement system with other demand distributions such as Poisson or normal so that more evidence can be collected to maintain a high completeness of this study.

Moreover, our proposed heuristic for finding an optimal total average cost does not guarantee the global optimal solution can be found. Its speed is relatively slow as the lead-time is long and the demand variability is high. In our simulation, the maximum iterations required were 35 (case 32). It is anticipated that more computation time is needed when lead-time and demand variability increase. An alternative strategy is to start at finding a good r^* with a fixed q. Then fix r^* and relax q for approaching a near optimal solution. It is possible because we found that r converges more rigorously. Besides, the algorithm for computing the probability distribution is not yet confirmed to be applicable when lead-time is more than two periods. One future research direction is to generalize this algorithm to be used when lead-time is more than two periods.

Chapter 5
Simulation of a Two-End-Product Postponement System

In the last chapter, we have shown that the cost benefits of a postponement system are limited when it is implemented to a single end-product supply chain system whose customer demands are discrete and uniformly distributed. To a great extent, a lot size-reorder point system outperforms a postponement system when it is operated in its optimal or near optimal total average cost at steady state based on our results. However, it is argued that a postponement system may outperform a lot size-reorder point system if the supply chain system offers more than one end-product. This view is coherent with our findings in Chapters 2 and 3. Besides, the analysis would be more valuable if both Poisson and normal distributions are considered. In fact, it is more equitable to compare the two systems by an experimental approach instead of their long-run steady states because the steady state may be reached only after infinite periods or a very long time. In order to maintain completeness of our study, a more dynamic system is developed by simulation technique in this chapter.

In this chapter we conduct simulation experiments of a two-end-product supply chain, for which customer demands are discrete and independent. Customer demands follow a uniform, Poisson or normal distribution. Two simulation models, namely one is a postponement system while the other is a non-postponement system, are designed for comparing their performance and total cost after t periods. Given a set of (r, q) policies and a demand distribution, the postponement system outperforms the non-postponement system in terms of average order frequency, average on-hand inventory, average backorder and average fill-rate. Thus, this system provides some cost benefits when the net postponement cost is low.

This chapter is organized as follows. The proposed models and assumptions are presented in Section 5.1. Details of the simulation run are discussed in Section 5.2. In Section 5.3 and 5.4, simulation results for both performance indicators and total average costs are analyzed. Some concluding remarks are given in Section 5.5.[1]

[1] The following discussion in this chapter is largely based on the ideas and results presented in Wan [125].

T.C.E. Cheng et al., *Postponement Strategies in Supply Chain Management*, International Series in Operations Research & Management Science 143,
DOI 10.1007/978-1-4419-5837-2_5,

5.1 Proposed Model and Assumptions

There are three parties, namely a supplier, a manufacturer and a group of customers in our supply chain. We assume the model is discrete in time. This time the manufacturer offers two end-products to customers. The end-products are similar but their demands are independent of each other. Like our study in the previous chapter, the manufacturer is facing two options in offering the end-products: to use a lot size-reorder point system and to use a zero stock system. For simplicity, we denote the first one as a non-postponement system and the latter one is referred to as a postponement system. In the non-postponement system, the manufacturer orders the two end-products separately from the supplier by sending two independent orders. These ordering decisions are controlled by the specified reorder point and order quantity of each end-product (Fig. 5.1). On the other hand, the manufacturer can order the work-in-process inventory (WIP) and transform them into finished goods when demands are known in the beginning of each period. The customization process is assumed to be simple and quick so that the end-products can be shipped out in the same period. The ordering decision of the WIP is also controlled by its reorder point and order quantity (Fig. 5.2). Since there is no stock of end-products, we call this system the postponement system and the customization process is the product differentiation point.

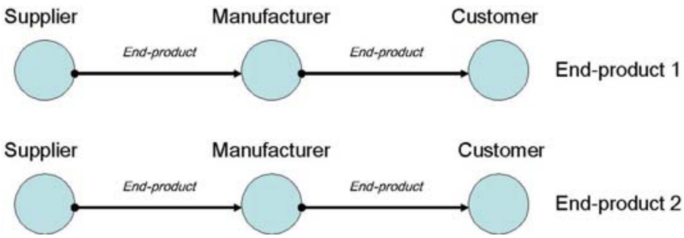

Fig. 5.1 Process flow of the non-postponement system

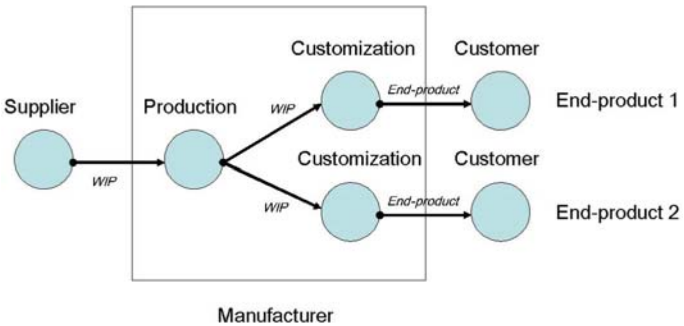

Fig. 5.2 Process flow of the postponement system

5.1.1 Notation

- i = end-product ($i = 1, 2$), and WIP ($i = 3$);
- t = Number of periods, $t > 0$;
- k = Fixed order cost, $k > 0$;
- c = Variable cost per unit, $c > 0$;
- h = Inventory holding cost per unit per period, $h > 0$;
- b = Backorder cost per unit per period, $b > 0$;
- α = Net postponement cost per unit per period;
- r_i = Reorder point for i, $r_i \geq 0$;
- q_i = Order quantity for i, $q_i \geq 0$;
- $I_i(t)$ = On-hand inventory for i at the end of period t, $I(t)_i \geq 0$;
- $B_i(t)$ = Accumulated backorder for i at the end of period t, $B(t)_i \geq 0$;
- $O_i(t)$ = Order quantity placed for i at the end of period t, $O(t)_i \geq 0$;
- $A_i(t)$ = Order arrived for i in the beginning of period t, $A(t)_i \geq 0$;
- $IL_i(t)$ = Inventory level for i at the end of period t, $IL_i(t) = I_i(t) - B_i(t)$;
- $IP_i(t)$ = Inventory position for i at the end of period t before placing order, $IP_i(t) = I_i(t) - B_i(t) +$ outstanding orders at time t;
- $S_i(t)$ = Order shipped out for i at the end of period t, $S(t)_i \geq 0$;
- $D_i(t)$ = Demand for i in the beginning of period t, $D(t)_i \geq 0$;
- L = Order lead-time for the end-product, $L \geq 0$;
- $\overline{I}_i$ = Average inventory for i per period, $\overline{I}_i \geq 0$;
- $\overline{B}_i$ = Average backorder for i per period, $\overline{B}_i \geq 0$;
- $\overline{OF}_i$ = Average order frequency for i per period, $\overline{OF}_i \geq 0$;
- $\overline{D}_i$ = Average demand for i per period, $\overline{D}_i \geq 0$;
- $\overline{F}_i$ = Average fill rate for i per period, $\overline{F}_i = \frac{\overline{D}_i - \overline{B}_i}{\overline{D}_i} \times 100\%$;
- $C(r, q)$ = Total average cost per period when q and r are given,$C(r, q) > 0$.

5.1.2 Model Assumptions

(i) Order lead-time L is constant for i.
(ii) For simplicity, one unit of WIP is used to make one unit of end-product 1 or end-product 2.
(iii) The fixed order cost, variable cost and backorder cost are identical for end-product 1 and end-product 2.
(iv) The inventory holding cost is identical for end-product 1, end-product 2 and WIP.
(v) The supplier has unlimited capacity.
(vi) Order $A_i(t)$ arrives in the beginning of period t with no delay, $A_i(0) = 0$.
(vii) Customer demand for end-product i, $(i = 1, 2)$ at period t, $D_i(t)$, is not known until the beginning of period t, $D(0)_i = 0$.
(viii) Customer demand for end-product i, $(i = 1, 2)$ at period t, $D_i(t)$, is handled in the beginning of period t.

(**ix**) Backorders are fulfilled immediately when there is enough inventory in the beginning of period t.

We recall that one advantage of a postponement system is that it can reduce demand variability by the risk-pooling effect (Zinn [129]). It can enjoy bulk purchase advantage and a lower inventory holding cost as the WIPs are less bulky and easier to handle. However, the cost associated with the customization process cannot be ignored (Lee and Billington [66]). In view of this fact, a simulation study is employed to analyze whether or not the postponement is more effective in terms of some performance indicators and a total average cost function. In this chapter, the five statements below are examined.

Statement 1: Postponement results in a lower average order frequency $\overline{OF}$.
Statement 2: Postponement results in a lower average inventory $\bar{I}$.
Statement 3: Postponement results in a lower average backorder $\overline{B}$.
Statement 4: Postponement results in a better average fill rate $\overline{F}$.
Statement 5: Postponement results in a lower total average cost $C(r,q)$ for some cost parameters.

The total average cost per period for the non-postponement system is

$$C(r_1,q_1)+C(r_2,q_2) = k(\overline{OF}_1+\overline{OF}_2)+c(\overline{D}_1+\overline{D}_2) \\ +h(\bar{I}_1+\bar{I}_2)+b(\overline{B}_1+\overline{B}_2). \qquad (5.1.1)$$

The total average cost per period for the postponement system is

$$C(r_3,q_3) = k\overline{OF}_3+(c+\alpha)\overline{D}_3+h\bar{I}_3+b\overline{B}_3, \qquad (5.1.2)$$

where α is the net postponement cost per unit per period.

5.2 Methodology

5.2.1 System Dynamics

The system dynamics for end-product 1, end-product 2 and WIP, numbered as 3, are identical to those expressed in Section 4.2. To recap, they are stated as follows.

$$IL_i(t) = IL_i(t-1)+A_i(t)-D_i(t),\ t=1,2,\ldots \qquad (5.2.3)$$

$$B_i(t) = -\min\{IL_i(t),0\},\ t=0,1,\ldots \qquad (5.2.4)$$

$$I_i(t) = \max\{IL_i(t),0\},\ t=0,1,\ldots \qquad (5.2.5)$$

$$IP_i(t) = IL_i(t) + \sum_{j=t+1-L}^{t-1} O_i(j),\ t = 0, 1, \ldots \tag{5.2.6}$$

and

$$O_i(j) = 0 \text{ for all } j < 0, \tag{5.2.7}$$

The reorder decision,

$$O_i(t) = A_i(t+L) = \begin{cases} q_i, \text{ if } IP_i(t) \le r_i \\ 0, \quad \text{otherwise.} \end{cases}$$

for $i = 1, 2, 3$.

Two flow charts that help to explain the flow of the non-postponement system and the postponement system are illustrated in Figs. 5.3 and 5.4, respectively. Note that for the non-postponement system, there should be two independent system flows, each of which is shown as Fig. 5.3, for end-product 1 and end-product 2, while there is only one for the postponement system (Fig. 5.4).

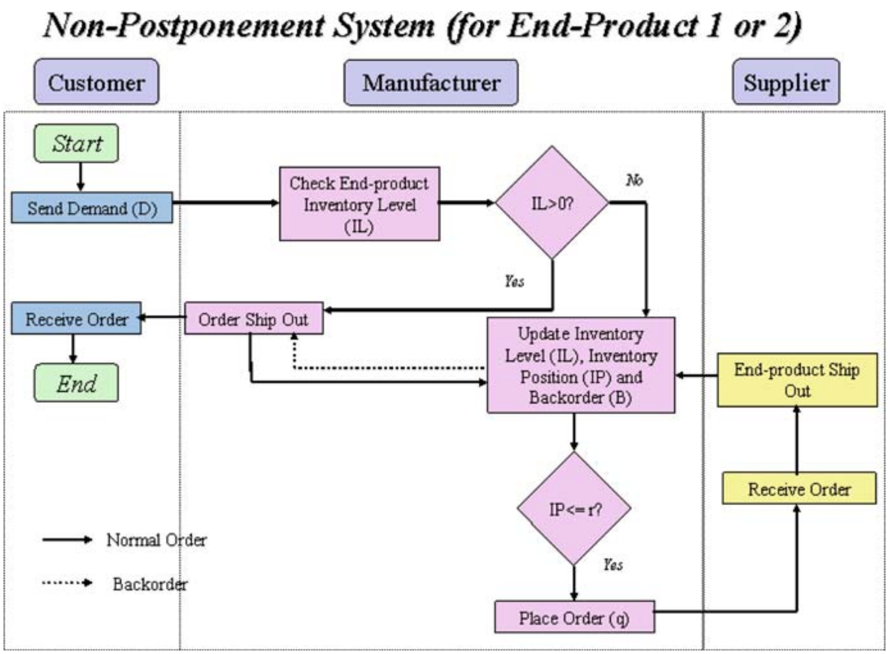

Fig. 5.3 Flow chart for the non-postponement system

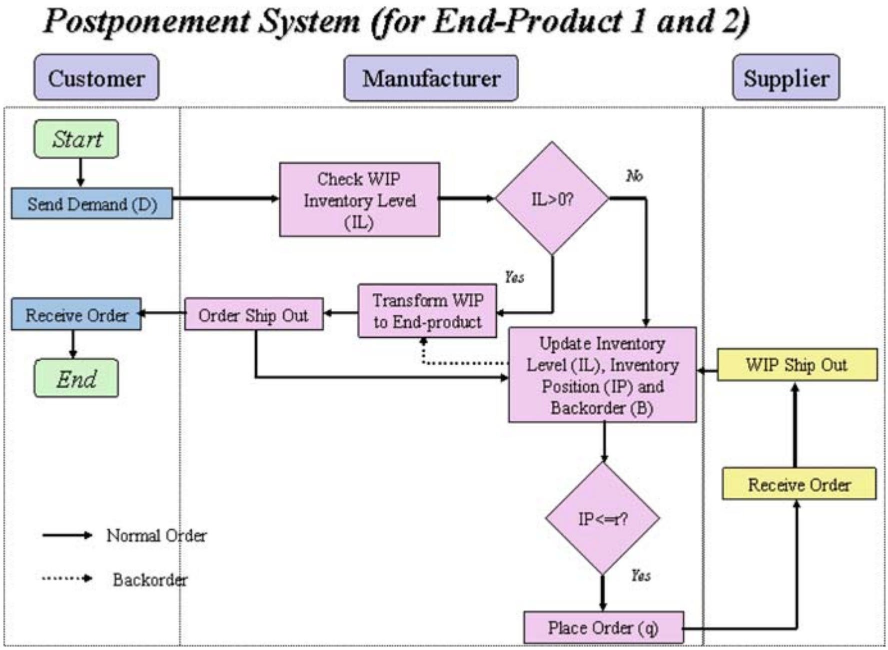

Fig. 5.4 Flow chart for the postponement system

5.2.2 The Simulation Model

Two simulation programs, one developed for modelling a postponement system and the other used for evaluating a non-postponement system, were developed based on the above system dynamics and assumptions. The purpose of the simulation programs is to collect the performance indicators, namely average order frequency, average on-hand inventory, average backorder and average fill rate, from each system after a sizable computational run (100+t periods in our study), where t =1,000. The first 100 periods were system warm-up periods and they were not recorded for analysis. It could avoid the initial instability of the program. We define a simulation run as an event when the two programs are run.

5.2.3 Customer Demand Distribution

For each simulation run, four sets of customer demands were generated randomly from the Excel random number generator based on a specified distribution and its parameters such as mean and standard deviation. In our study, customer demands followed a uniform distribution, a Poisson distribution, or a normal distribution in each simulation run. The four sets of customer demands covered both high and low demand variabilities, in terms of demand range, for end-product 1 and end-product 2 respectively. We let L_i and H_i, respectively, to represent a low demand range and a high demand range. The combinations of customer demands were (L_1, L_2), (H_1, H_2), (H_1, L_2) and (L_1, H_2). The aim of this arrangement is to compare if demand pattern affects the performance indicators under the same distribution. Moreover, the simulation programs were run with different order lead-times L so as to evaluate how stable our findings were when L is prolonged. Values of L under study were $L = 1, 2, 3, 5$, 10 and 20 periods.

5.2.4 Order Quantity and Reorder Point

In order to conduct a fair comparison, the same order quantity and reorder point were used for both simulation programs. However, the choice for each simulation run was arbitrary as we did not attempt to study the optimal total average cost. Instead, we wished to generalize our findings. For each combination, we used an (r_1, q_1) policy and an (r_2, q_2) policy, respectively, for end-product 1 and 2 when they were ordered separately. On the other hand, we used an $(r_1 + r_2, q_1 + q_2)$ policy for the postponement system. Further we assumed that the order quantity and reorder point were proportional to L so as to maintain a high fill rate. The (r, q) policy employed under different demand distributions is summarized in Table 5.1.

5.2.5 Summary of Parameters

The values of the cost parameters, number of periods and order lead-times for our simulation study are listed below.

h: 10 (base value)
k: 5 (low fixed order cost), 10 (moderate fixed order cost), 20 (high fixed order cost)
b: 5 (low variable cost), 10 (moderate variable cost), 20 (high variable cost)
c: 10
α: $0, 0.5, 1$
L: $1, 2, 3, 5, 10, 20$
t: 1000

According to our finding in Chapter 5, c does not affect the results of our analysis so that it was assumed to be a constant. Like what we have done in the last chapter, we used h as our base value in this simulation. For each demand distribution (uniform, Poisson or normal), there were 2 (simulation programs)$\times$27 (cost combinations)$\times$4 (demand combinations)$\times$6 (order lead-times)=1,296 sets of data.

5.2.6 Initial Conditions

For simplicity, we assume $D_i(0) = O_i(0) = A_i(0) = B_i(0) = S_i(0) = 0$ for $i = 1, 2, 3$. $I_i(0) = IL_i(0) = IP_i(0) = 50$ if the demand range for end-product i is low, $i = 1, 2$, or $I_i(0) = IL_i(0) = IP_i(0) = 500$ if the demand range for end-product i

Table 5.1 (r, q) policy in different distributions

	Uniform distribution		Poisson distribution		Normal distribution	
End-product demand distribution	1 $U(0, a_1)$	2 $U(0, a_2)$	1 $P(\lambda_1, 1)$	2 $P(\lambda_2, 1)$	1 $N(\mu_1, \sigma_1)$	2 $N(\mu_2, \sigma_2)$
Policy	(r_1, q_1)	(r_2, q_2)	(r_1, q_1)	(r_2, q_2)	(r_1, q_1)	(r_2, q_2)
Non-postponement	$(0.5a_1L, a_1L)$	$(0.5a_2L, a_2L)$	$(\lambda_1L, 2\lambda_1L)$	$(\lambda_2L, 2\lambda_2L)$	$(\mu_1L, 2\mu_1L)$	$(\mu_2L, 2\mu_2L)$
Postponement	$[0.5(a_1 + a_2)L, (a_1 + a_2)L]$		$[(\lambda_1 + \lambda_2)L, 2(\lambda_1 + \lambda_2)L]$		$[(\mu_1 + \mu_2)L, 2(\mu_1 + \mu_2)L]$	

is high, $i = 1, 2$. Thus, $I_3(0) = IL_3(0) = IP_3(0)$=100, 550 or 1,000 based upon different demand combinations.

5.3 Simulation Results for Non-cost Parameters

5.3.1 Uniform Distribution

Four sets of customer demands, namely L_1, H_1, L_2, H_2, were generated based on the parameters in Table 5.2. Further, the lot size-reorder points for the non-postponement system and the postponement system are summarized in Table 5.3. The simulation results are shown in Tables A.1 and A.2 in Appendix A.

Table 5.2 Customer demand sets (Uniform Distribution)

Data set	Distribution
L_1	D_1:$U(0, 10)$
H_1	D_1:$U(0, 100)$
L_2	D_2:$U(0, 10)$
H_2	D_2:$U(0, 100)$

Table 5.3 Lot size-reorder points (r, q) for the non-postponement system and the postponement system (Uniform distribution)

Combination	End-product 1	End-product 2	WIP
L_1, L_2	$(5L, 10L)$	$(5L, 10L)$	$(10L, 20L)$
H_1, H_2	$(50L, 100L)$	$(50L, 100L)$	$(100L, 200L)$
L_1, H_2	$(5L, 10L)$	$(50L, 100L)$	$(55L, 110L)$
H_1, L_2	$(50L, 100L)$	$(5L, 10L)$	$(55L, 110L)$

According to Tables A.1 and A.2, it was found that the results are insensitive to order lead-time L in general. Based on these two tables, it is clear that the saving from the average order frequency is 50%. Thus, *Statement 1* is supported. However, this is the maximum saving as we treated those situations when end-product 1 and end-product 2 are ordered in the same period as two independent orders. In fact, they belong to one order, especially when there is only one supplier such as our supply chain.

With reference to Fig. A.1 (Appendix A), it reveals that 75% of the cases supported that there is improvement on average inventory by adopting a postponement system. Among the four demand combinations, the saving in inventory is the most significant when both demand ranges for D_1 and D_2 are low and the order lead-time L is shorter (this saving is 9.1% smaller than that of the non-postponement system). Thus, *Statement 2* is supported.

Besides, the saving in average backorder is satisfactory (Table A.2 in Appendix A). Only one exceptional case reported that the postponement system results a higher average backorder. It is relatively better when both demand ranges for D_1 and D_2 are low or high. The minimum improvement reached 40% in these two data sets, while the maximum saving was only 26.7% in the other two data sets. Hence, *Statement 3* is supported.

Since there were improvements in average inventory and average backorder, the average fill rate was found to be improved in twenty-three cases with one exceptional case (Fig. A.3 in Appendix A). Thus, *Statement 4* is supported. The most significant improvement was revealed from the data sets (L_1, L_2) and (H_1, H_2). Generally speaking, the postponement system results in a more outstanding performance than the non-postponement system, especially when demand ranges for D_1 and D_2 are both low or high.

5.3.2 Poisson Distribution

Four sets of customer demands, namely L_1, H_1, L_2, H_2, were generated based on the parameters in Table 5.4. Further, the lot size-reorder points for the non-postponement system and the postponement system are summarized in Table 5.5. The simulation results are shown in Tables B.1 and B.2 in Appendix B.

Table 5.4 Customer demand sets (Poisson Distribution)

Data set	Distribution
L_1	D_1:$P(10,1)$
H_1	D_1:$P(100,1)$
L_2	D_2:$P(10,1)$
H_2	D_2:$P(100,1)$

Table 5.5 Lot size-reorder points (r,q) for the non-postponement system and the postponement system (Poisson distribution)

Combination	End-product 1	End-product 2	WIP
L_1, L_2	$(10L, 20L)$	$(10L, 20L)$	$(20L, 40L)$
H_1, H_2	$(100L, 200L)$	$(100L, 200L)$	$(200L, 400L)$
L_1, H_2	$(10L, 20L)$	$(100L, 200L)$	$(110L, 220L)$
H_1, L_2	$(100L, 200L)$	$(10L, 20L)$	$(110L, 220L)$

Unlike the findings reported in Section 5.3.1, this time the performance indicators are dependent of the lead-time L. Again, it is obvious that *Statement 1* is supported as the average order frequency has dropped by about 50% by adopting the postponement system (Tables B.1 and B.2). Based on Fig. B.1 in Appendix B, the saving in the average inventory is generally satisfactory, except for the data set (L_1, H_2) for which only 33% of this statement is true. As the saving depends on the lead-time L, it is higher when L is small, while it becomes insignificant when $L \geq 10$. In general, there were 79% of cases that reported that the postponement system results in a lower average inventory. Thus, *Statement 2* is supported.

The difference in the average backorder between the two systems depends more on lead-time than on demand range (Fig. B.2 in Appendix B). But there is no trend as L is prolonged. It is noted that the average backorder of the non-postponement system is roughly equal to that of the postponement system when $L = 5$ for all data sets. One possible reason is that the lot size-reorder point policy for the non-postponement system provides a good fit for all data sets when $L = 5$. However, twenty-one cases supported the result that the postponement system leads to a lower average backorder. Thus, *Statement 3* is supported.

The most remarkable average fill rate improvement takes place when both λ_1 and λ_2 are low (Fig. B.3 in Appendix B). For the other three data sets, only slight improvements are revealed. In general, *Statement 4* is supported.

5.3.3 Normal Distribution I

Four sets of customer demands, namely L_1, H_1, L_2, H_2, were generated based on the parameters in Table 5.6. Since the normal distribution includes negative values, we simply treat them as zero during the random number generation process. The lot size-reorder points for the non-postponement system and the postponement system are summarized in Table 5.7. The simulation results are shown in Tables C.1 and C.2 in Appendix C.

Table 5.6 Customer demand sets (Normal Distribution I)

Data Set	Distribution
L_1	D_1:$N(10,5)$
H_1	D_1:$N(100,50)$
L_2	D_2:$N(10,5)$
H_2	D_2:$N(100,50)$

Table 5.7 Lot size-reorder points (r,q) for the non-postponement system and the postponement system (Normal Distribution I)

Combination	End-product 1	End-product 2	WIP
L_1, L_2	$(10L, 20L)$	$(10L, 20L)$	$(20L, 40L)$
H_1, H_2	$(100L, 200L)$	$(100L, 200L)$	$(200L, 400L)$
L_1, H_2	$(10L, 20L)$	$(100L, 200L)$	$(110L, 220L)$
H_1, L_2	$(100L, 200L)$	$(10L, 20L)$	$(110L, 220L)$

Based on the simulation results (Tables C.1 and C.2), it is clear that *Statement 1* is supported as the average order frequency decreases by 50% on average. However, the difference in the average inventory between the non-postponement system and the postponement system is affected by both lead-time L and mean demands μ_1 and μ_2. (Fig. C.1 in Appendix C). When $L \geq 10$, the difference is close to zero, while when lead-time is short, say $L = 1$, the postponement system yields a lower average inventory. The postponement system results in a lower inventory when both μ_1 and μ_2 are low ($\mu_1 = \mu_2 = 10$ in our study). In contrast, high μ_1 and μ_2 give the worst performance among the four data sets, in which only two cases reported savings. Generally speaking, 67% of the cases supported that the postponement system yields a lower average inventory. Thus, *Statement 2* is supported.

The overall results support that the postponement system associated with a lower average backorder as there was only one exceptional case (Fig. C.2 in Appendix C). Hence, *Statement 3* is supported. Among the four data sets, the postponement system outweighs the non-postponement in coping with the average backorder especially when both μ_1 and μ_2 are equal. That is, $\mu_1 = \mu_2 = 10$ or $\mu_1 = \mu_2 = 100$ in our study.

Although postponement is not excellent in maintaining a lower average inventory, the average fill rate is remarkably better than the non-postponement system (Fig. C.3 in Appendix C). Again, there was only one exceptional case to disagree with this finding. Thus, *Statement 4* is supported. The improvement increases as lead-time L increases, particularly when both μ_1 and μ_2 are equal. In sum, the postponement system is highly preferable when L is long and mean customer demands μ_1 and μ_2 are similar.

5.3.4 Normal Distribution II

In describing a normal distribution, customer demand variability can also be explained by its standard deviation instead of its mean, as standard deviation measures the absolute variability of customer demands (Simchi-Levi et al. [105]). In our case, customer demands for the two end-products can fluctuate highly even though they have the same mean. Therefore, our attention is paid to evaluate whether or not a different conclusion is made when they have smaller standard deviations. In what follows, $\sigma_1 = \sigma_2 = 2$ when $\mu_1 = \mu_2 = 10$, and $\sigma_1 = \sigma_2 = 20$ when $\mu_1 = \mu_2 = 100$. Four customer sets are presented in Table 5.8. For simplicity, the lot size-reorder points for the non-postponement system and the postponement system are identical to those used in Section 5.3.3 (see Table 5.7).

Table 5.8 Customer demand sets (Normal Distribution II)

Data Set	Distribution
L_1	$D_1{:}N(10,2)$
H_1	$D_1{:}N(100,20)$
L_2	$D_2{:}N(10,2)$
H_2	$D_2{:}N(100,20)$

The simulation results are shown in Tables D.1 and D.2 in Appendix D. According to these two tables, it is revealed that the average order frequency has dropped by about 50% for all sets of simulation results. The saving was roughly equal to that reported in Section 5.3.3. Thus, *Statement 1* is supported. However, there were nearly half of the cases found to support that the non-postponement is more cost effective, especially when the two customer demand sets have different means (Fig. D.1 in Appendix D). The total number of cases that reported a saving dropped from 67 to 58% when the standard deviation became smaller. It gives an impression that there was no significant difference in choosing a postponement system to obtain a reduction in average inventory, although *Statement 2* is supported.

The improvement in the average backorder was still satisfactory when the standard deviations of customer demands became smaller. With reference to Fig. D.2 in Appendix D, it is shown that there were only three exceptional cases disagreed with the fact that the postponement system led to a reduction in average backorder. More specifically, the postponement has an absolute advantage to outperform the non-postponement system when customer demands for both end-products 1 and 2 are equal. That is, $\mu_1 = \mu_2$. Hence, *Statement 3* is supported. It is consistent with our finding in Section 5.3.3, in which standard deviations of customer demands are higher.

Again, there were only three cases that departed from the observation that the postponement system results in better average fill rates (Fig. D.3 in Appendix D). In other words, the overall system performance has been improved, especially when both customer demands are equal. Thus, *Statement 4* is supported. Although the four statements are supported, the postponement system yields more savings when the standard deviations of the two demands are relatively larger (see Section 5.3.3).

Based on our findings, the postponement system is preferred when the end-products have similar demand patterns and larger standard deviations, though it generally results in a better performance when customer demands follow a uniform, Poisson or normal distribution.

5.4 Simulation Results for Cost Parameters

It is found that the total average cost for the postponement system and the non-postponement is insensitive to average backorder and average order frequency as they are too small in comparison with the average demand and average inventory, given that the maximum b and k are only twice of the values of c and h. Although there are remarkable savings from average backorder and average order frequency according to our previous findings, their associated cost savings may not be sufficient to cover the net postponement cost per unit α. Thus, an analysis based on total average cost is meaningful on this ground.

We compare the two system costs by a cost difference percentage as follows.

$$C = \frac{(5.1.2) - (5.1.1)}{(5.1.1)} \times 100\%.$$

If $C < 0$, then it implies that the postponement system results in a lower total average cost per period, and vice versa. Since there are 648 data sets for each distribution, an average C is presented for each demand combination [(L_1, L_2), (H_1, H_2), (H_1, L_2), (L_1, H_2)], lead-time ($L = 1, 2, 3, 5, 10, 20$) and α ($\alpha = 0, 0.5, 1$). A summary of the simulation results are shown in Table E.1 in Appendix E.

It is interesting to note that when α=0, there were about 77% of our data reports a cost saving with the postponement system, regardless of the demand distribution. However, when $\alpha = 1$, the percentage of the data reports having a cost saving is below 21%. In other words, α is very sensitive to the total average costs of the two system. When α=0.5, the number of cases that reported a saving dropped below 50%. The worst case is when customer demands have smaller standard deviations (see Normal Distribution II from Table E.1 in Appendix E). There were only three cases that agreed with the fact that the postponement system results in a lower total average cost. Furthermore, it dropped to one case when α=1. Among the four distributions, it is found that the postponement system performs better under uniform distribution and Poisson distribution. If the data from the Normal Distribution II is discarded, the postponement system is preferred if the mean customer demands for end-products 1 and 2 are equal and relatively low (L_1, L_2) when α=0.5. Otherwise, the non-postponement system should be adopted. For those customer demands that follow a Poisson distribution, or a normal distribution with larger standard deviations, this conclusion is still applied for longer lead-time (say 5 or 10 periods in our study). Whereas it is only true for those uniform customer demands with shorter lead-time (say 1,2 or 3 periods). In general, one reason that limits the cost saving from the postponement system when lead-time is long is that the system does not account for demand variability between the two end-products. In fact, the demands of the two end-products may be dependent because they are similar from customers' point of view. Therefore, they may have negative covariance such that the demand standard deviation for the postponement system is smaller. Thus, it might result a lower $C(r_3, q_3)$, given the mean customer demand for the postponement system is unchanged. When α=1, there is no strong reason to opt for the postponement system in all distributions according to our findings. Thus, *Statement 5* is supported only when α=0, or α=0.5, the mean customer demands for end-products 1 and 2 are similar and low, and lead-time is short.

It is argued that α is greater than zero in a strong sense. In fact, the net postponement cost can be minimized in three ways. The first one is from joint ordering. Recall that the average order frequency drops by 50% for all demand distributions. It implies we order more at each time period. It raises the bargaining power to ask for a discounted k due to bulk purchase. Besides, since WIP is less bulky than the end-products, more number of WIP can be shipped and thus, a lower k is obtained. Another cost improvement strategy is the variable cost. In general, the variable cost is lower if WIP is ordered instead of end-products. Moreover, the WIP is easier to manage and forecast so that all h, q and r can be improved. These savings can be used to balance the extra customization process cost so that the net postponement cost is low or even negative. It makes our simulation study more applicable in handling real life scenarios.

5.5 Concluding Remarks

In this chapter, the relative effectiveness between a postponement system and a non-postponement system was studied. We assumed two end-products are offered in a supply chain. Customer demands are discrete in time period and they follow a uniform distribution, a Poisson distribution or a normal distribution. In the non-postponement system, end-products are ordered separately from a single supplier, while in the postponement system, WIPs are ordered from the supplier and are transformed into end-products when demands are known. Order decisions in both systems follow some lot size-reorder point policies. Two simulation programs were modelled to collect the key performance indicators, including average order frequency, average inventory, average backorder and average fill rate, of the two systems. Also, a cost analysis was conducted by putting the values of these indicators and their cost parameters into two total average cost functions. By comparing the performance indicators and the difference in total average cost between the two systems, it was shown that the postponement system is more effective when the customer demands have similar demand patterns (independent of the choice of demand distribution), larger standard deviations (for normal distribution) and the net postponement cost is relatively low (maximum 5% of the variable cost in our study). However, there are limitations for this study. The first one is the generation of data sets. For each distribution, only four sets of customer demands were drawn. In fact, a larger sample size can reduce bias and make the analysis more general. It is preferred in future research studies. Another limitation is that this study only considered the supply of two end-products. The simulation model can be extended to support up to n end-products so that the extent to which n affects the choice of the postponement system can be fully evaluated.

Chapter 6
Application of Postponement: Examples from Industry

In this chapter we report on two case studies of applying postponement strategy in industry. The first case is a study of a Hong Kong based toaster manufacturing company, which has successfully implemented postponement strategy. We present a summary of how postponement strategy was implemented in its supply chain and elaborate on all the benefits arising from the implementation of postponement. We also discuss the implications of postponement for its supply chain. In the second case an empirical analysis by Chiou et al. [28] is introduced. They empirically examine the application of postponement strategy in Taiwanese information technology (IT) firms. First, the four types of form postponements are examined. Second, the factors affecting the adoption of different form postponement strategies are explored. The managerial and practical implications are also discussed.

This chapter is organized as follows. In Section 6.1 a case study form Hong Kong are discussed. In Section 6.2 the case of Taiwan's information technology industry is addressed. Some concluding remarks are given in Section 6.3.

6.1 A Case Study from Hong Kong

A survey was conducted by interviewing the managing director of an electric toaster manufacturer whose headquarters is based in Hong Kong.[1] All the production processes of the manufacturer are located in Guangdong province, China. The company was chosen because it has implemented postponement strategies for more than ten years. It provides some useful insights into the implementation of postponement strategies in a real-life situation. Moreover, both advantages and disadvantages of the strategy can be evaluated in comparison with our theoretical findings.

[1] The following discussion in this Section is largely based on the ideas and results presented in Wan [125].

T.C.E. Cheng et al., *Postponement Strategies in Supply Chain Management*, International Series in Operations Research & Management Science 143,
DOI 10.1007/978-1-4419-5837-2_6,

6.1.1 An Overview of the Company

The manufacturer produces more than 120 different models of toasters for 11 key customers (in terms of brands) from the US, Europe, Japan and China. It has more than 1,500 employees, who can produce 12,000 units per day. The company operates in a batch production mode. The number of toasters required per order ranges from 1,000 to 3,000 units and the customer is expected to wait between 2 and 4 weeks for the delivery of products. It is noted that the manufacturing capacity is sufficient to handle 12 orders at the same time. If there is enough capacity to handle an order, the average order lead-time is 9 days when work-in-process inventory (WIP) is available. Otherwise, it takes 15 days on average to complete the order. The order is fulfilled when it is delivered to a predetermined dock in Guangdong. The breakdowns of the lead-time is shown in Table 6.1.

Table 6.1 Lead-time breakdowns for toaster offering

Process	Average lead-time required (days)
From raw material to WIP	6
From WIP to finished products (assembly)	1
Gift box ordering (after order receipt)	7
Packing and pre-shipment inspection	1
Ship to dock	1
Order lead-time (no WIP)	$6 + 7 + 1 + 1 = 15$
Order lead-time (WIP available)	$7 + 1 + 1 = 9$

Table 6.1 reveals that the company should keep more WIP so that the lead-time can be 6 days shorter. It is one of the reasons to motivate the company to implement the postponement strategy. Besides, customer demand fluctuates highly because orders are generated from different companies, which sell their products to different markets. In addition, customers always demand highly customized products so there are few repeat orders. This demand pattern makes both the make-to-stock and make-to-order operational modes impossible to smooth the production. It drives the company to consider adopting a postponement strategy, in which some operations are run in a make-to-stock mode, while others are operated in a make-to-order mode. The postponement strategy implementation is presented in the next section.

6.1.2 Implementation of Postponement

Before adopting a postponement strategy, the company employed a pure make-to-order mode to fulfill customer orders. In short, it transformed raw material to finished goods only after customer order was received. It resulted in a highly unpredictable production schedule.

The company implemented two postponement strategies, namely a form postponement strategy and a pull postponement strategy, to improve the production schedule. The company carried out a form postponement by part standardization.

Parts, such as heaters, power cords and modules, are make-to-stock based on demand forecast and all related information is provided to customers during the product design process. It not only reduces the production time but also maintains a high degree of product variety. When actual orders are received, these standardized parts, together with other customized parts such as gift boxes and outer shells, are transformed to finished goods in a timely way. It helps the company to become more responsive in fulfilling customers' requirements at a lower cost by this seamless process. A detailed process flow chart of the postponement system is illustrated in Fig. 6.1.

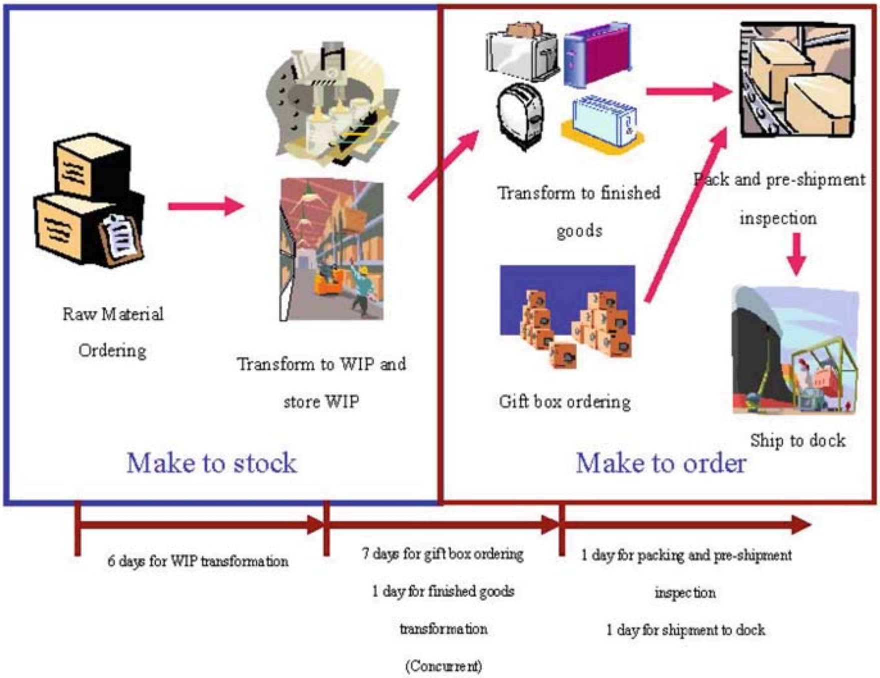

Fig. 6.1 A process flow chart of the toaster company

6.1.3 Benefits of Using Postponement

The benefits experienced by the toaster company after 10 years of implementation of postponement are summarized as follows.

(**i**) Lead time has reduced from 15 to 9 days.
(**ii**) Cost of raw material has decreased by 5% due to bulk purchase.
(**iii**) Although the reduction in raw materials inventory is only 10% (see Table 6.2), the number of items handled has dropped by 33% because the majority

of them have been transformed to WIP (common module). It enables the company to respond to different customer demands quickly with a lower raw materials inventory.

(iv) Set-up time has dropped by 15% because the manufacturing of WIP is a make-to-stock process based on demand forecast. In fact, the WIP is suitable for producing nearly all kinds of toasters. It helps to smooth the process and reduce the set-up time.

(v) Overall machine utilization is over 85%.

(vi) Order fulfillment is near 100%.

(vii) Make-to-stock in upstream production smoothes the whole supply chain process in facing high demand fluctuations. Since all make-to-stock production is based upon demand forecast, the company can have a better manpower schedule to handle demand fluctuations.

(viii) A lower cost has been achieved in expanding the product range because only those customization processes and parts are reconfigured.

(ix) It has become easier to maintain a low WIP inventory level by fine-tuning the forecast system and manpower.

(x) Since the average order lead-time has been reduced to 8 days, customers can place their orders more frequently. From the manufacturer's point of view, although the order quantity is lower, the demand fluctuation is lower as well. It implies that production is more stable. From the customers' point of view, they are more satisfied because this manufacturer can adjust their orders in a timely and responsive manner instead of having a large batch of unwanted products.

Table 6.2 Change in inventory before and after postponement implementation

Inventory	Before postponement (%)	After postponement (%)	Change (%)
Raw material	80	70	−10
WIP	10	15	5
Finished goods	10	15	5
Total	100	100	0

6.1.4 Implications

Although postponement has helped the company to achieve a low inventory level and a low cost, it is not a perfect strategy to solve all supply chain problems. Currently the toaster company still faces the following three problems.

(i) Recall from Table 6.1 that the gift box ordering process requires 7 days as it is a make-to-order process. Attention should be paid to this process for the further reduction of lead-time.

(ii) Customer orders are discontinuous and so demand fluctuations are still high. In fact, the production quantity of an order is equal to several months' sales for many key customers.

(iii) Logistics postponement is not economical in this study because the production cost in China is about the cheapest in the world. The cost saving in shipping a smaller cubic volume and weight cannot offset the extra production cost for shifting the packaging process to local distribution centers in the US or Europe.

6.2 The Case of Taiwanese Information Technology Industry[2]

In this section, we introduce the case study of Taiwanese information technology (IT) firms. The IT firm was selected because IT products are characterized by high product values, short product life cycles, and high demands for customization. Taiwan was selected because it is one of the largest producers of IT products and is the largest original equipment manufacturing (OEM) partner for US and Japan. Furthermore, Taiwanese firms have taken up the responsibility of providing global logistics.

6.2.1 The Hypothesis

The goal of this case study is multi-fold. First, the four types of form postponements proposed by Zinn and Bowersox [130] are empirically examined. Second, the factors affecting the adoption of different form postponement strategies are explored.

6.2.1.1 Four Types of Form Postponement Strategies

Zinn and Bowersox [130] first classified form postponement strategies into four types: labeling, packaging, assembly, and manufacturing. The validity of the classification is assessed. Then the strength of the relationships among the four postponement strategies is examined in the following hypothesis.

Hla: If a firm adopts a packaging postponement strategy, it is more likely to adopt a labeling postponement strategy as well.

Hlbl: If a firm adopts an assembly postponement strategy, it is more likely to adopt a packaging postponement strategy as well.

Hlb2: If a firm adopts an assembly postponement strategy, it is more likely to adopt a labeling postponement strategy as well.

Hlcl: If a firm adopts a manufacturing postponement strategy, it is more likely to adopt an assembly postponement strategy as well.

---

[2] The following discussion in this section is an excerpt from Chiou et al. [28].

Hlc2: If a firm adopts a manufacturing postponement strategy, it is more likely to adopt a packaging postponement strategy as well.

Hlc3: If a firm adopts a manufacturing postponement strategy, it is more likely to adopt a labeling postponement strategy as well.

6.2.1.2 Factors Affect Postponement Strategy

Based on the interviews with industry professionals and evidences from past studies (Bowersox and Morash [15], Closs et al. [30], Feitzinger and Lee [41], Pagh and Cooper [91], van Hoek et al. [122]), four major product/demand characteristics are identified as candidate drivers of form postponement strategies. These drivers are demand for customization, modularity in construction, product value, and product life cycle. The characteristics are examined with the following hypothesis.

H2a: Firms whose products are characterized by strong customer demand for customization are more likely to use modular product designs.

H2b: Firms whose products are more modular in design are more likely to practice assembly postponement.

H2c: Firms whose products are characterized by higher consumer demand for customization are more likely to practice assembly postponement.

H3a: Firms whose products' key components are more expensive to carry are more likely to practice labeling postponement.

H3b: Firms whose products' key components are more expensive to carry are more likely to practice packaging postponement.

H3c: Firms whose products' key components are more expensive to carry are more likely to practice assembly postponement

H3d: Firms whose products' key components are more expensive to carry are more likely to practice manufacturing postponement.

H4: Firms whose products are characterized by a shorter product life cycle are more likely to practice manufacturing postponement.

6.2.2 Methodology

Confirmatory factor analysis (CFA) is used to validate the dimensionality of form postponement strategies, while path analysis is used to examine the relationships between product/demand characteristics and the adoption of postponement strategies.

Six hundred Taiwanese IT firms were selected as targets for this study. Survey questions were designed to identify the characteristics and the type of form postponements associated with the firms' primary products. Of the 600 questionnaires distributed, 102 were completed and returned (17%). Sample statistics are presented in Table 6.3.

Table 6.3 The background statistics of the sample

Demographic statistics average sales	87.5 million USD (2.1–61,000)[a]
Average of sales growth	40.1% (2–400%)
Average No. of employees	736 (3–20,000)
Average percentage of export sales	47% (8–100%)
Average of foreign production over total production	32.5% (5–100%)
Average No. of overseas warehouses	1.58(1–10)

[a]Numbers in parentheses indicate range of response.
Source: Chiou et al. [28]

6.2.3 Results

First, it is concluded that form postponement is better modeled as four distinct constructs: labeling, packaging, assembly, and manufacturing postponements.

Afterwards, it is concluded that hypotheses *Hla*, *Hlbl*, *Hlb2*, and *Hlcl* are accepted while *Hlc2* and *Hlc3* are not. The rejection of *Hlc2* and *Hlc3* indicates that the average Taiwanese IT firm practicing manufacturing postponement, is not more or less likely to also practice labeling or packaging postponement.

Finally, it is concluded that hypotheses *H2a*, *H2b*, *H2c* and *H3a* are supported at the 95% significance level while *H3b* and *H4* are accepted at 90% significance; *H3c* and *H3d* are not accepted. The acceptance of hypotheses *H2a* and *H2b* indicates that customization is likely to induce modular product designs and modular product designs are likely to induce assembly postponement. Hypothesis *H2c* is accepted indicating that customization is also likely to induce assembly postponement directly. Hypothesis *H3a* and *H3b* are accepted while *H3c* and *H3d* are not accepted indicating that products with high component costs appear to benefit from labeling and/or packaging postponement strategy but not from manufacturing postponement or assembly postponement. Hypothesis *H4* is accepted indicating that product life cycle has a significantly positive effect on manufacturing postponement implementation.

6.2.4 Implications

This study provides evidence that form postponement strategies are practiced widely by Taiwan IT firms and the four types of form postponements as proposed by Zinn and Bowersox fit the postponement strategies practiced by Taiwan's IT firms.

In addition, these results indicate that the experience in implementing one postponement strategy may have lowered the cost of implementing another postponement strategy. The experience should provide better forecasts of the benefits and costs of implementing postponement strategies, thus lowering the risk of implementation.

Furthermore, the results confirm the suspicion that there is a natural causal relationship between postponement strategies. Products characterized by high customer demand for customization appear to benefit from assembly postponement as well as implementing more modular designs. Products, which are modular in design, appear to benefit from assembly postponement as well. Also products with expensive key components appear to benefit from labeling and packaging postponement, but not necessarily assembly and manufacturing postponement. Finally, products which have short product life cycles, appear to benefit from manufacturing postponement.

6.3 Concluding Remarks

In this chapter we report on two case studies of applying postponement strategy in industry. The first case is a study of a Hong Kong based toaster manufacturing company, which has successfully implemented postponement strategy. We present a summary of how postponement strategy was implemented in its supply chain and elaborate on all the benefits arising from the implementation of postponement. We also discuss the implications of postponement for its supply chain. In the second case study we introduce an empirical analysis of the application of postponement strategy in Taiwanese information technology (IT) firms. This study provides evidence that form postponement strategies are practiced widely by IT firms in Taiwan. In addition, four major product/demand characteristics are identified as candidate drivers of form postponement strategies.

From the two case studies, the following conclusions can be obtained.

- Postponement strategies are practiced widely by some firms with certain characteristics.
- Although postponement can help a company to achieve a low inventory level and a low cost under certain circumstances, it is not a perfect strategy to solve all supply chain problems.

Chapter 7
Conclusions, Implications and Future Research Directions

7.1 Conclusions

The objective of this book was to address the benefits of adopting a pull postponement strategy, which is enabled with a form postponement strategy, in a supply chain from a manufacturer's point of view. In order to provide some general insights that can be applied to a wide spectrum of scenarios, we developed four postponement models based on different customer demand distributions, inventory policies and supply chain compositions. In each model, a total average cost function and some performance indicators were defined and they were used to compare with those of a non-postponement model. Our research showed that the postponement strategy is more beneficial when dealing with more than one end-product. A summary of our findings are presented as follows.

There are new research opportunities in studying postponement by deterministic models such as EOQ and EPQ. Also, there is little research that studies postponement when customer demands follow a uniform distribution. Our book fills these two gaps.

The first postponement model is EOQ-based. The model aims at evaluating whether or not ordering n imperishable end-products jointly results in a lower EOQ and a lower total average cost. By optimization, our results showed that postponement yields a lower EOQ and a lower total average cost, provided that the end-products have same backorder costs, different backorder costs or a combination of both. Furthermore, we develop an EOQ-based model with perishable items to evaluate the impact of item deterioration rate on inventory replenishment policies. Our theoretical analysis and computational results show that for perishable products with a constant deterioration rate, a postponement strategy can yield a lower total average cost when the constant deterioration rate is small (Chapter 2).

Afterwards, we develop two EPQ-based models with and without stockout to examine the impact of postponement. We separate a supply chain into two processes. The first process is the production of a generic product that is the core module of n end-products, while the second process is to customize the generic products to different end-products. We formulate the total average cost functions of the two scenarios for producing and keeping n end-products in a supply chain, in which

T.C.E. Cheng et al., *Postponement Strategies in Supply Chain Management*, International Series in Operations Research & Management Science 143,
DOI 10.1007/978-1-4419-5837-2_7,

their demands are known and deterministic. Our findings indicated that: (1) in the scenario without stockout, postponement results in a lower total average cost when customer demands of the n end-products are fulfilled after production is complete; (2) in the scenario with planned backorder, postponement leads to a lower optimal total average cost when the demand is met after production is complete and the planned backorder costs are the same for all the end-products (Chapter 3).

The third model is similar to the first model we have described. This time a lot size-reorder point (r, q) policy replaces the EOQ ordering decision because customer demands are stochastic (uniformly distributed). Also, there is only one end-product in the supply chain. The purpose of this study was to analyze whether or not postponement is still more cost effective in stochastic environment. By optimization, we showed that postponement is more cost effective when the order lead-time is zero and the fixed ordering cost is relatively low. When order lead-time is greater than zero, postponement loses its cost advantage in a wide variety of cases by simulation technique (Chapter 4).

The last model attempts to confirm our findings from the third model by expanding it to offer two end-products. We considered a number of general situations instead of optimums and a net postponement cost was added so as to generalizing our view. The simulation results showed that postponement generally yields a lower total average cost under different customer demand distributions such as uniform, Poisson and normal, provided that customer demands have similar patterns and the net postponement cost is low (Chapter 5). Table 7.1 summarizes our findings in this book.

Our analysis would be more complete by giving two real life examples to highlight the benefits of postponement in practice. By interviewing a Hong Kong manufacturer, it was found that postponement offers substantial cost savings through lead-time and inventory reductions. By surveying the Taiwanese information technology firms, the relationships between product/demand characteristics and the adoption of postponement strategies are examined. It was found that postponement strategies are practiced widely by IT firms in Taiwan. The conclusions obtained from the two case studies support our theoretical results (Chapter 6).

7.2 Implications and Further Research Directions

As mentioned, there are totally four postponement strategies, namely pull, logistics, form and price postponement, according to recent research studies. However, our book only focuses on pull and form postponement models. In fact, our models can be extended to examine the benefits of adopting a logistics postponement. One way is to build some distribution centers near the point of purchase so that the customization process can be handled there. It involves adding other cost parameters such as distribution cost, fixed cost, variable cost and inventory holding cost of the distribution centers to our total average cost functions. It may provide some saving opportunities because generic products are less bulky to ship (see Lee et al. [67],

Table 7.1 Summary of book findings

Chapter	Number of end-products	Customer demand	Inventory model	Methodology	Results
2	n	Deterministic	EOQ	Optimization	Postponement results in a lower total average cost and EOQ for imperishable products. For perishable products with a deterioration rate θ, postponement is more cost effective when θ is small.
3	n	Deterministic	EPQ	Optimization	Without stockout, postponement results in a lower total average cost when the demands are met after production is complete. With planned backorders, postponement is more cost effective when the demands are met after production is complete and the backorder costs are equal.
4	1	Stochastic (uniform)	(r, q) and zero stock	Optimization and simulation	Postponement is more cost effective when order lead-time is zero. Postponement is not as cost effective as the (r, q) policy when dealing with one end-product.
5	2	Stochastic (uniform, Poisson and normal)	(r, q)	Simulation	Postponement results in a lower total average cost when the two end-products have similar demand patterns and the net postponement cost is low.

Ackerman [2] and Twede et al. [114]). The saving can be used to offset the extra spending on standardization and customization in the system-wide cost function of a supply chain.

Besides, the impacts on a price postponement can be evaluated by maximizing a profit function instead of minimizing a total average cost because it involves a pricing strategy that is dependent of customer demand. Of course, an ideal strategy is to formulate a system-wide profit function that involves all relevant costs incurred from adopting different combinations and levels of postponement in multiple points of product differentiations (Garg and Tang [44] studied two differentiation points).

Our total average cost functions consist of fixed ordering/production cost, variable cost, inventory holding cost and backorder cost. Perhaps more costs are relevant if the system becomes globalized. One of these costs is the custom and duty costs (Lee and Billington [66]). They are associated with the charges for moving parts and end-products across different national boundaries. In fact, the calculation of these custom and duty costs is rather complicated as they involve profound knowledge of international trading laws and trading policies. From a global supply chain perspective, custom and duty costs must be taken into account whenever a logistics postponement decision is made, although most models have neglected these costs.

Apart from considering purely on costs and benefits, further analysis based on quality issues is required, especially when some processes are shifted to distribution centers or third parties.

Finally, the book addressed the impact of adopting a pull postponement strategy on a manufacturer in a supply chain. One potential future research direction is to study the impact of postponement on the entire supply chain with deterministic customer demands.

Appendix A
Simulation Results (Uniform Distribution)

Table A.1 Simulation results 1 (Uniform Distribution)

Data	Average demand	Average inventory	Average backorder	Average order frequency	Average fill rate (%)
L = 1					
L1	4.961	5.847	0.149	0.497	97.00
L2	4.969	5.624	0.168	0.497	96.62
H1	49.769	52.06	2.43	0.497	95.12
H2	50.395	51.595	1.893	0.504	96.24
LL (without postponement)	9.93	11.471	0.317	0.994	96.81
LL (with postponement)	9.93	10.428	0.184	0.496	98.15
Change (%)	0.00	9.09	41.96	60.10	1.34
HH (without postponement)	100.164	103.655	4.323	1.001	95.68
HH (with postponement)	100.164	102.271	2.139	0.501	97.86
Change (%)	0.00	–1.34	–50.52	–49.95	2.18
HL (without postponement)	54.738	57.684	2.598	0.994	95.25
HL (with postponement)	54.738	56.57	1.904	0.497	96.52
Change (%)	0.00	–1.93	–26.71	–50.00	1.27
LH (without postponement)	55.356	57.442	2.042	1.001	96.31
LH (with postponement)	55.356	57.328	1.728	0.503	96.88
Change (%)	0.00	–0.20	–15.38	–49.75	0.57
L = 2					
L1	4.961	10.869	0.141	0.248	97.16
L2	4.969	11.041	0.185	0.249	96.28
H1	49.769	102.364	2.034	0.249	95.91
H2	50.395	101.454	2.052	0.252	95.93
LL (without postponement)	9.93	21.91	0.326	0.497	96.72
LL (with postponement)	9.93	21.219	0.175	0.248	98.24
Change (%)	0.00	–3.15	–46.32	–50.10	1.52
HH (without postponement)	100.164	203.818	4.086	0.501	95.92
HH (with postponement)	100.164	203.352	1.82	0.25	98.18
Change (%)	0.00	–0.23	–55.46	–50.10	2.26
HL (without postponement)	54.738	113.405	2.219	0.498	95.95
HL (with postponement)	54.738	112.035	1.929	0.249	96.48
Change (%)	0.00	–1.21	–13.07	–50.00	0.53
LH (without postponement)	55.356	112.323	2.193	0.5	96.04
LH (with postponement)	55.356	111.867	2.257	0.252	95.92
Change (%)	0.00	–0.41	2.92	–49.60	–0.12
L = 3					
L1	4.961	15.691	0.203	0.166	95.91
L2	4.969	15.76	0.154	0.166	96.90
H1	49.769	152.396	1.666	0.166	96.65
H2	50.395	150.142	2.44	0.168	95.16
LL (without postponement)	9.93	31.451	0.357	0.332	96.40
LL (with postponement)	9.93	30.344	0.18	0.165	98.19

T.C.E. Cheng et al., *Postponement Strategies in Supply Chain Management*,
International Series in Operations Research & Management Science 143,
DOI 10.1007/978-1-4419-5837-2,

Table A.1 (continued)

Data	Average demand	Average inventory	Average backorder	Average order frequency	Average fill rate (%)
Change (%)	0.00	–3.52	–49.58	–50.30	1.78
HH (without postponement)	100.164	302.538	4.106	0.334	95.90
HH (with postponement)	100.164	297.774	2.042	0.167	97.96
Change (%)	0.00	–1.57	–50.27	–50.00	2.06
HL (without postponement)	54.738	168.156	1.82	0.332	96.68
HL (with postponement)	54.738	169.162	1.416	0.166	97.41
Change (%)	0.00	0.60	–22.20	–50.00	0.74
LH (without postponement)	55.356	165.833	2.643	0.334	95.23
LH (with postponement)	55.356	165.404	2.334	0.168	95.78
Change (%)	0.00	–0.26	–11.69	–49.70	0.56

Table A.2 Simulation results 2 (Uniform Distribution)

Data	Average demand	Average inventory	Average backorder	Average order frequency	Average fill rate (%)
L = 5					
L1	4.961	25.688	0.21	0.099	95.77
L2	4.969	25.534	0.198	0.1	96.02
H1	49.769	250.314	2.184	0.099	95.61
H2	50.395	256.118	2.116	0.101	95.80
LL (without postponement)	9.93	51.222	0.408	0.199	95.89
LL (with postponement)	9.93	51.447	0.183	0.099	98.16
Change (%)	0.00	0.44	–55.15	–50.25	2.27
HH (without postponement)	100.164	506.432	4.3	0.2	95.71
HH (with postponement)	100.164	507.507	2.375	0.1	97.63
Change (%)	0.00	0.21	–44.77	–50.00	1.92
HL (without postponement)	54.738	275.848	2.382	0.199	95.65
HL (with postponement)	54.738	277.061	1.845	0.099	96.63
Change (%)	0.00	0.44	–22.54	–50.25	0.98
LH (without postponement)	55.356	281.806	2.326	0.2	95.80
LH (with postponement)	55.356	276.117	1.837	0.101	96.68
Change (%)	0.00	–2.02	–21.02	–49.50	0.88
L = 10					
L1	4.961	51.444	0.116	0.05	97.66
L2	4.969	50.847	0.211	0.049	95.75
H1	49.769	508.342	1.212	0.05	97.56
H2	50.395	491.972	2.97	0.05	94.11
LL (without postponement)	9.93	102.291	0.327	0.099	96.71
LL (with postponement)	9.93	102.126	0.162	0.049	98.37
Change (%)	0.00	–0.16	–50.46	–50.51	1.66
HH (without postponement)	100.164	1,000.314	4.182	0.1	95.82
HH (with postponement)	100.164	995.633	2.501	0.05	97.50

Table A.2 (continued)

Data	Average demand	Average inventory	Average backorder	Average order frequency	Average fill rate (%)
Change (%)	0.00	–0.47	–40.20	–50.00	1.68
HL (without postponement)	54.738	559.189	1.423	0.099	97.40
HL (with postponement)	54.738	556.338	1.172	0.05	97.86
Change (%)	0.00	–0.51	–17.64	–49.49	0.46
LH (without postponement)	55.356	543.416	3.086	0.1	94.43
LH (with postponement)	55.356	546.036	2.806	0.05	94.93
Change (%)	0.00	0.48	–9.07	–50.00	0.51
L = 20					
L1	4.961	101.082	0.054	0.025	98.91
L2	4.969	101.194	0.258	0.025	94.81
H1	49.769	1,001.791	1.661	0.025	96.66
H2	50.395	1,013.586	2.584	0.025	94.87
LL (without postponement)	9.93	202.276	0.312	0.05	96.86
LL (with postponement)	9.93	201.732	0.168	0.025	98.31
Change (%)	0.00	–0.27	–46.15	–50.00	1.45
HH (without postponement)	100.164	2,015.377	4.245	0.05	95.76
HH (with postponement)	100.164	2,027.159	2.027	0.025	97.98
Change (%)	0.00	0.58	–52.25	–50.00	2.21
HL (without postponement)	54.738	1,102.985	1.919	0.05	96.49
HL (with postponement)	54.738	1,083.868	1.802	0.025	96.71
Change (%)	0.00	–1.73	–6.10	–50.00	0.21
LH (without postponement)	55.356	1,114.668	2.638	0.05	95.23
LH (with postponement)	55.356	1,111.164	2.534	0.025	95.42
Change (%)	0.00	–0.31	–3.94	–50.00	0.19

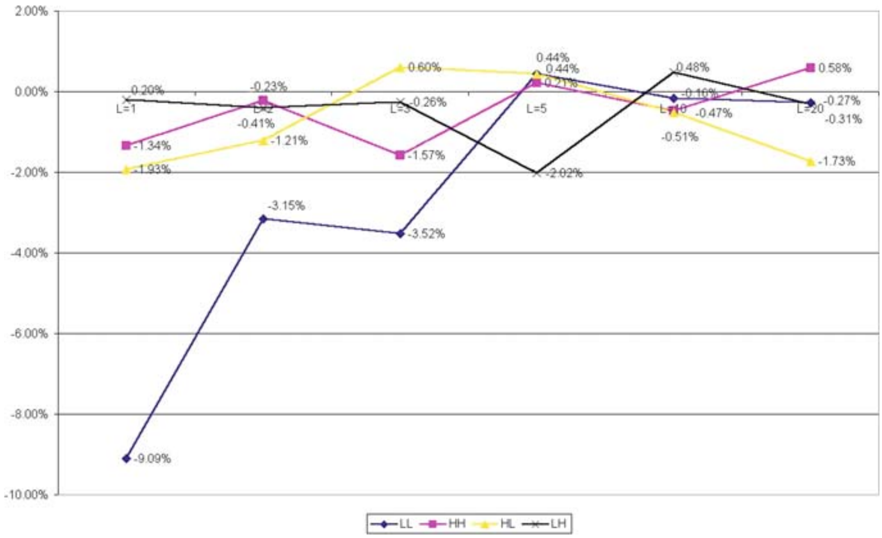

Fig. A.1 Difference in average inventory (Uniform Distribution)

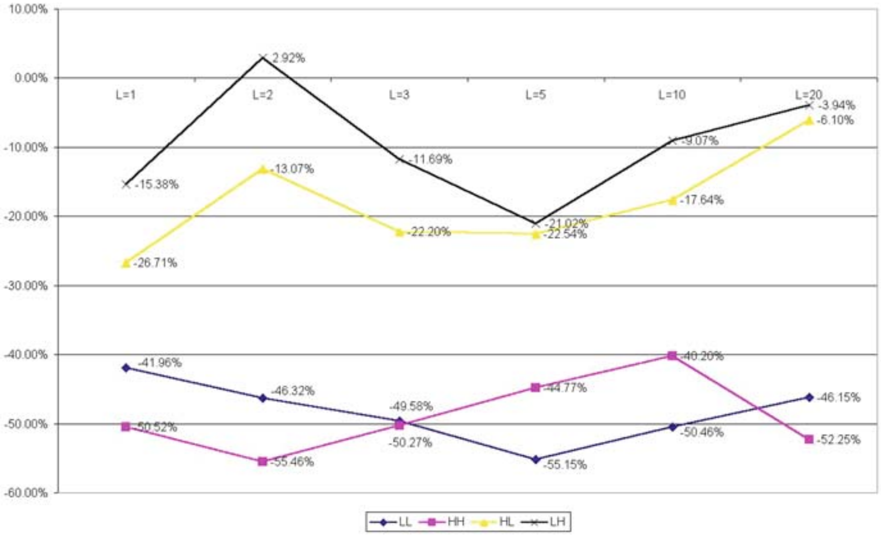

Fig. A.2 Difference in average backorder (Uniform Distribution)

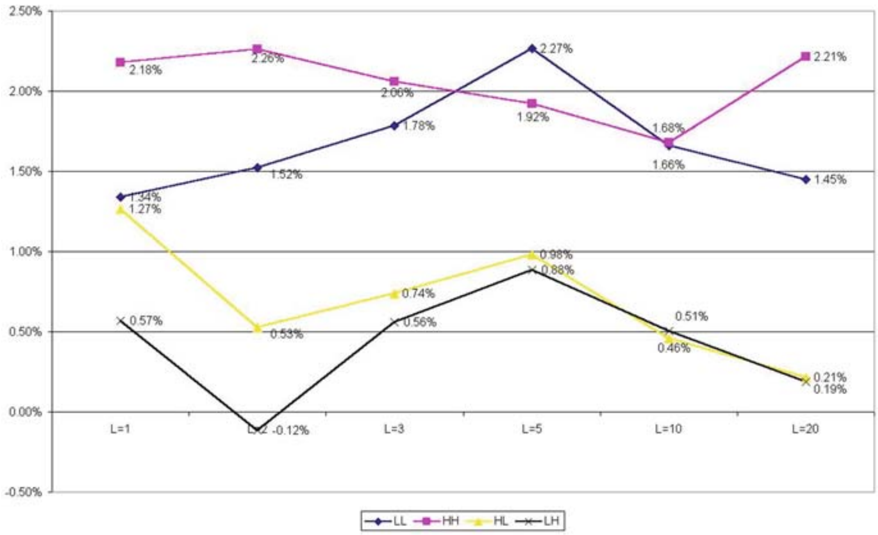

Fig. A.3 Improvement in average fill rate (Uniform Distribution)

Appendix B
Simulation Results (Poisson Distribution)

Table B.1 Simulation results 1 (Poisson Distribution)

Data	Average demand	Average inventory	Average backorder	Average order frequency	Average fill rate (%)
L = 1					
L1	9.954	10.777	0.094	0.498	99.06
L2	10.01	10.692	0.142	0.5	98.58
H1	100.305	104.872	0.122	0.502	99.88
H2	100.522	97.846	0.159	0.502	99.84
LL (non-postponement)	19.964	21.469	0.236	0.998	98.82
LL (postponement)	19.964	20.711	0.098	0.499	99.51
Change (%)	0.00	–3.53	–58.47	–50.00	0.69
HH (non-postponement)	200.827	202.718	0.281	1.004	99.86
HH (postponement)	200.827	196.006	0.169	0.503	99.92
Change (%)	0.00	–3.31	–39.86	–49.90	0.06
HL (non-postponement)	110.315	115.564	0.264	1.002	99.76
HL (postponement)	110.315	108.985	0.125	0.502	99.89
Change (%)	0.00	–5.69	–52.65	–49.90	0.13
LH (non-postponement)	110.476	108.623	0.253	1	99.77
LH (postponement)	110.476	109.978	0.148	0.502	99.87
Change (%)	0.00	1.25	–41.50	–49.80	0.10
L = 2					
L1	9.954	20.974	0.091	0.248	99.09
L2	10.01	20.804	0.114	0.251	98.86
H1	100.305	204.688	0.138	0.25	99.86
H2	100.522	197.417	0.13	0.251	99.87
LL (non-postponement)	19.964	41.778	0.205	0.499	98.97
LL (postponement)	19.964	40.558	0.105	0.249	99.47
Change (%)	0.00	–2.92	–48.78	–50.10	0.50
HH (non-postponement)	200.827	402.105	0.268	0.501	99.87
HH (postponement)	200.827	397.608	0.171	0.251	99.91
Change (%)	0.00	–1.12	–36.19	–49.90	0.05
HL (non-postponement)	110.315	225.492	0.252	0.501	99.77
HL (postponement)	110.315	217.462	0.142	0.25	99.87
Change (%)	0.00	–3.56	–43.65	–50.10	0.10
LH (non-postponement)	110.476	218.391	0.221	0.499	99.80
LH (postponement)	110.476	221.485	0.115	0.251	99.90
Change (%)	0.00	1.42	–47.96	–49.70	0.10
L = 3					
L1	9.954	31.032	0.129	0.166	98.70
L2	10.01	30.993	0.123	0.167	98.77
H1	100.305	304.466	0.116	0.167	99.88
H2	100.522	297.068	0.181	0.167	99.82
LL (non-postponement)	19.964	62.025	0.252	0.333	98.74

Table B.1 (continued)

Data	Average demand	Average inventory	Average backorder	Average order frequency	Average fill rate (%)
LL (postponement)	19.964	60.905	0.092	0.166	99.54
Change (%)	0.00	–1.81	–63.49	–50.15	0.80
HH (non-postponement)	200.827	601.534	0.297	0.334	99.85
HH (postponement)	200.827	595.96	0.123	0.167	99.94
Change (%)	0.00	–0.93	–58.59	–50.00	0.09
HL (non-postponement)	110.315	335.459	0.239	0.334	99.78
HL (postponement)	110.315	327.695	0.155	0.167	99.86
Change (%)	0.00	–2.31	–35.15	–50.00	0.08
LH (non-postponement)	110.476	328.1	0.31	0.333	99.72
LH (postponement)	110.476	330.643	0.153	0.167	99.86
Change (%)	0.00	0.78	–50.65	–49.85	0.14

Table B.2 Simulation results 2 (Poisson Distribution)

Data	Average demand	Average inventory	Average backorder	Average order frequency	Average fill rate (%)
L = 5					
L1	9.954	51.773	0.07	0.1	99.30
L2	10.01	51.199	0.069	0.1	99.31
H1	100.305	504.216	0.066	0.101	99.93
H2	100.522	497.047	0.16	0.1	99.84
LL (non-postponement)	19.964	102.972	0.139	0.2	99.30
LL (postponment)	19.964	100.461	0.128	0.1	99.36
Change (%)	0.00	–2.44	–7.91	–50.00	0.06
HH (non-postponement)	200.827	1,001.263	0.226	0.201	99.89
HH (postponement)	200.827	987.317	0.28	0.101	99.86
Change (%)	0.00	–1.39	23.89	–49.75	–0.03
HL (non-postponement)	110.315	555.415	0.135	0.201	99.88
HL (postponement)	110.315	551.419	0.139	0.101	99.87
Change (%)	0.00	–0.72	2.96	–49.75	0.00
LH (non-postponement)	110.476	548.82	0.23	0.2	99.79
LH (postponement)	110.476	546.298	0.208	0.1	99.81
Change (%)	0.00	–0.46	–9.57	–50.00	0.02
L = 10					
L1	9.954	101.312	0.109	0.05	98.90
L2	10.01	99.877	0.147	0.05	98.53
H1	100.305	1,000.279	0.129	0.05	99.87
H2	100.522	983.228	0.341	0.05	99.66
LL (non-postponement)	19.964	201.189	0.256	0.1	98.72
LL (postponment)	19.964	199.152	0.219	0.05	98.90
Change (%)	0.00	–1.01	–14.45	–50.00	0.19
HH (non-postponement)	200.827	1,983.507	0.47	0.1	99.77
HH (postponement)	200.827	1,983.327	0.29	0.05	99.86

Table B.2 (continued)

Data	Average demand	Average inventory	Average backorder	Average order frequency	Average fill rate (%)
Change (%)	0.00	–0.01	–38.30	–50.00	0.09
HL (non-postponement)	110.315	1, 100.156	0.276	0.1	99.75
HL (postponement)	110.315	1, 097.006	0.126	0.05	99.89
Change (%)	0.00	–0.29	–54.35	–50.00	0.14
LH (non-postponement)	110.476	1, 084.54	0.45	0.1	99.59
LH (postponement)	110.476	1, 094.102	0.212	0.05	99.81
Change (%)	0.00	0.88	–52.89	–50.00	0.22
L = 20					
L1	9.954	200.371	0.168	0.024	98.31
L2	10.01	200.48	0.15	0.025	98.50
H1	100.305	1, 996.4	0.25	0.025	99.75
H2	100.522	2, 003.006	0.119	0.026	99.88
LL (non-postponement)	19.964	400.851	0.318	0.049	98.41
LL (postponment)	19.964	402.227	0.094	0.025	99.53
Change (%)	0.00	0.34	–70.44	–48.98	1.12
HH (non-postponement)	200.827	3, 999.406	0.369	0.051	99.82
HH (postponement)	200.827	3, 947.468	0.431	0.025	99.79
Change (%)	0.00	–1.30	16.80	–50.98	–0.03
HL (non-postponement)	110.315	2, 196.88	0.4	0.05	99.64
HL (postponement)	110.315	2, 183.976	0.296	0.025	99.73
Change (%)	0.00	–0.59	–26.00	–50.00	0.09
LH (non-postponement)	110.476	2, 203.377	0.287	0.05	99.74
LH (postponement)	110.476	2, 193.983	0.093	0.026	99.92
Change (%)	0.00	–0.43	–67.60	–48.00	0.18

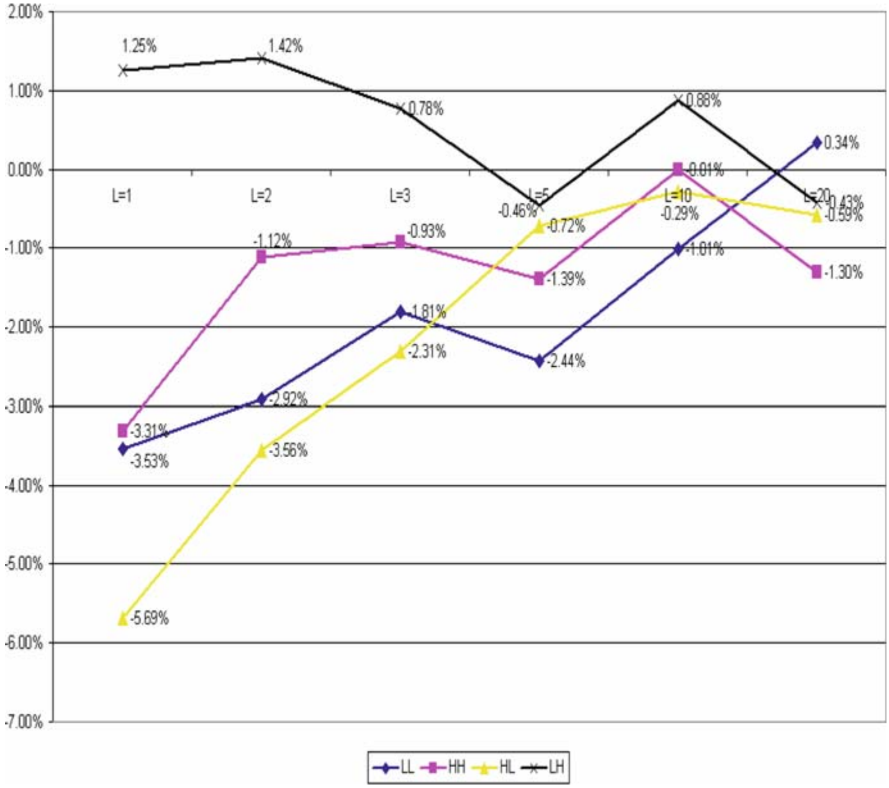

Fig. B.1 Difference in average inventory (Poisson Distribution)

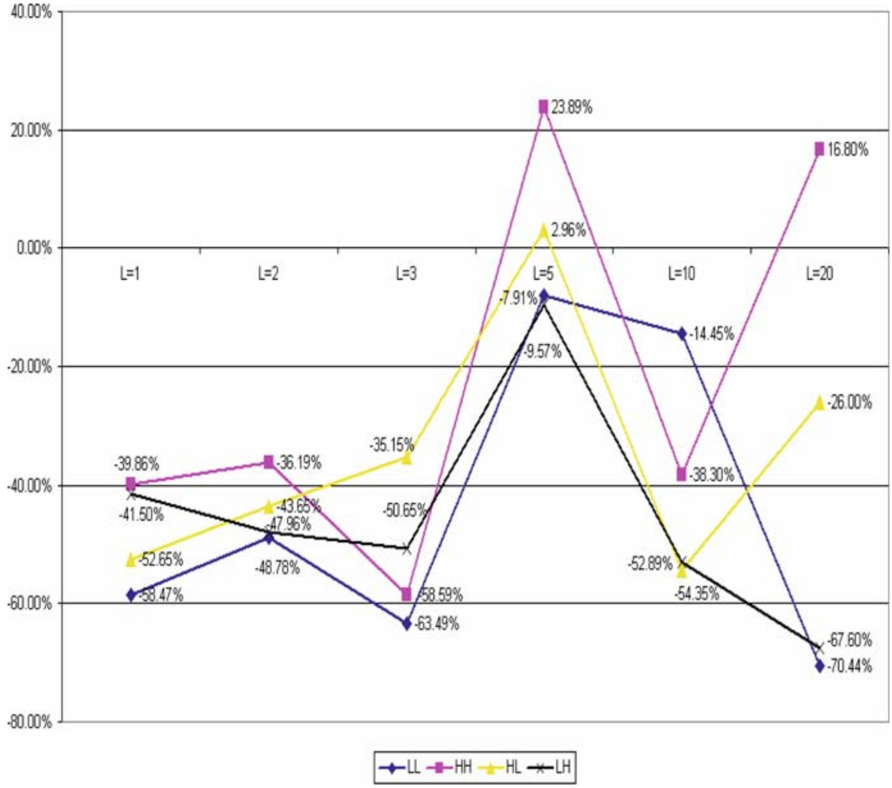

Fig. B.2 Difference in average backorder (Poisson Distribution)

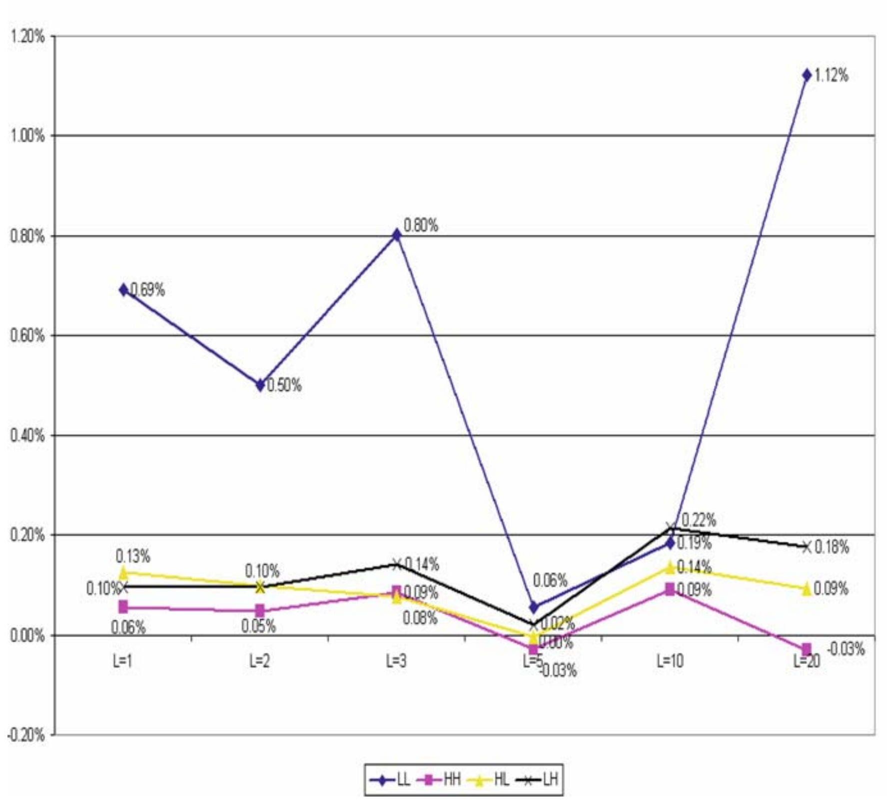

Fig. B.3 Improvement in average fill rate (Poisson Distribution)

Appendix C
Simulation Results (Normal Distribution I)

Table C.1 Simulation results 1 (Normal Distribution I)

Data	Average demand	Average inventory	Average backorder	Average order frequency	Average fill rate (%)
L = 1					
L1	10.033	10.663	0.303	0.501	96.98
L2	9.981	10.878	0.269	0.499	97.30
H1	100.045	103.47	3.488	0.501	96.51
H2	101.25	100.865	3.504	0.506	96.54
LL (non-postponement)	20.014	21.541	0.572	1	97.14
LL (postponement)	20.014	20.646	0.297	0.501	98.52
Change (%)	0.00	–4.15	–48.08	–49.90	1.37
HH (non-postponement)	201.295	204.335	6.992	1.007	96.53
HH (postponement)	201.295	206.539	3.196	0.503	98.41
Change (%)	0.00	1.08	–54.29	–50.05	1.89
HL (non-postponement)	110.026	114.348	3.757	1	96.59
HL (postponement)	110.026	113.906	2.655	0.5	97.59
Change (%)	0.00	–0.39	–29.33	–50.00	1.00
LH (non-postponement)	111.283	111.528	3.807	1.007	96.58
LH (postponement)	111.283	108.775	2.974	0.505	97.33
Change (%)	0.00	–2.47	–21.88	–49.85	0.75
L = 2					
L1	10.033	20.485	0.265	0.251	97.36
L2	9.981	20.891	0.322	0.249	96.77
H1	100.045	205.071	2.689	0.25	97.31
H2	101.25	199.159	2.598	0.253	97.43
LL (non-postponement)	20.014	41.376	0.587	0.5	97.07
LL (postponement)	20.014	40.673	0.244	0.25	98.78
Change (%)	0.00	–1.70	–58.43	–50.00	1.71
HH (non-postponement)	201.295	404.23	5.287	0.503	97.37
HH (postponement)	201.295	402.628	2.885	0.252	98.57
Change (%)	0.00	–0.40	–45.43	–49.90	1.19
HL (non-postponement)	110.026	225.962	3.011	0.499	97.26
HL (postponement)	110.026	222.161	2.45	0.25	97.77
Change (%)	0.00	–1.68	–18.63	–49.90	0.51
LH (non-postponement)	111.283	219.644	2.863	0.504	97.43
LH (postponement)	111.283	223.103	4.002	0.253	96.40
Change (%)	0.00	1.57	39.78	–49.80	–1.02
L = 3					
L1	10.033	30.239	0.339	0.167	96.62
L2	9.981	30.682	0.353	0.166	96.46
H1	100.045	295.628	3.846	0.167	96.16
H2	101.25	298.052	3.091	0.169	96.95
LL (non-postponement)	20.014	60.921	0.692	0.333	96.54

Table C.1 (continued)

Data	Average demand	Average inventory	Average backorder	Average order frequency	Average fill rate (%)
LL (postponement)	20.014	60.988	0.279	0.167	98.61
Change (%)	0.00	0.11	–59.68	–49.85	2.06
HH (non-postponement)	201.295	593.68	6.937	0.336	96.55
HH (postponement)	201.295	595.024	4.081	0.168	97.97
Change (%)	0.00	0.23	–41.17	–50.00	1.42
HL (non-postponement)	110.026	326.31	4.199	0.333	96.18
HL (postponement)	110.026	336.909	2.798	0.166	97.46
Change (%)	0.00	3.25	–33.37	–50.15	1.27
LH (non-postponement)	111.283	328.291	3.43	0.336	96.92
LH (postponement)	111.283	324.889	3.268	0.168	97.06
Change (%)	0.00	–1.04	–4.72	–50.00	0.15

Table C.2 Simulation results 2 (Normal Distribution I)

Data	Average demand	Average inventory	Average backorder	Average order frequency	Average fill rate (%)
L = 5					
L1	10.033	50.894	0.394	0.1	96.07
L2	9.981	51.644	0.295	0.099	97.04
H1	100.045	505.674	3.092	0.1	96.91
H2	101.25	503.021	3.46	0.101	96.58
LL (non-postponement)	20.014	102.538	0.689	0.199	96.56
LL (postponement)	20.014	101.29	0.341	0.1	98.30
Change (%)	0.00	–1.22	–50.51	–49.75	1.74
HH (non-postponement)	201.295	1,008.695	6.552	0.201	96.75
HH (postponement)	201.295	1,010.098	3.955	0.1	98.04
Change (%)	0.00	0.14	–39.64	–50.25	1.29
HL (non-postponement)	110.026	557.318	3.387	0.199	96.92
HL (postponement)	110.026	549.165	2.534	0.1	97.70
Change (%)	0.00	–1.46	–25.18	–49.75	0.78
LH (non-postponement)	111.283	553.915	3.854	0.201	96.54
LH (postponement)	111.283	553.194	2.333	0.101	97.90
Change (%)	0.00	–0.13	–39.47	–49.75	1.37
L = 10					
L1	10.033	101.547	0.247	0.05	97.54
L2	9.981	100.852	0.403	0.05	95.96
H1	100.045	1,009.158	3.576	0.05	96.43
H2	101.25	996.001	5.44	0.051	94.63
LL (non-postponement)	20.014	202.399	0.65	0.1	96.75
LL (postponement)	20.014	200.651	0.302	0.05	98.49
Change (%)	0.00	–0.86	–53.54	–50.00	1.74
HH (non-postponement)	201.295	2,005.159	9.016	0.101	95.52
HH (postponement)	201.295	1,989.497	3.354	0.051	98.33
Change (%)	0.00	–0.78	–62.80	–49.50	2.81
HL (non-postponement)	110.026	1110.01	3.979	0.1	96.38

Table C.2 (continued)

Data	Average demand	Average inventory	Average backorder	Average order frequency	Average fill rate (%)
HL (postponement)	110.026	1,108.647	3.216	0.05	97.08
Change (%)	0.00	–0.12	–19.18	–50.00	0.69
LH (non-postponement)	111.283	1,097.548	5.687	0.101	94.89
LH (postponement)	111.283	1,096.416	4.355	0.051	96.09
Change (%)	0.00	–0.10	–23.42	–49.50	1.20
L = 20					
L1	10.033	199.512	0.212	0.025	97.89
L2	9.981	195.954	0.505	0.025	94.94
H1	100.045	1,988.154	2.572	0.025	97.43
H2	101.25	1,979.427	6.866	0.025	93.22
LL (non-postponement)	20.014	395.466	0.717	0.05	96.42
LL (postponement)	20.014	397.468	0.319	0.025	98.41
Change (%)	0.00	0.51	–55.51	–50.00	1.99
HH (non-postponement)	201.295	3,967.581	9.438	0.05	95.31
HH (postponement)	201.295	3,997.116	2.973	0.025	98.52
Change (%)	0.00	0.74	–68.50	–50.00	3.21
HL (non-postponement)	110.026	2,184.108	3.077	0.05	97.20
HL (postponement)	110.026	2,181.633	2.602	0.025	97.64
Change (%)	0.00	–0.11	–15.44	–50.00	0.43
LH (non-postponement)	111.283	2,178.939	7.078	0.05	93.64
LH (postponement)	111.283	2,198.638	4.377	0.025	96.07
Change (%)	0.00	0.90	–38.16	–50.00	2.43

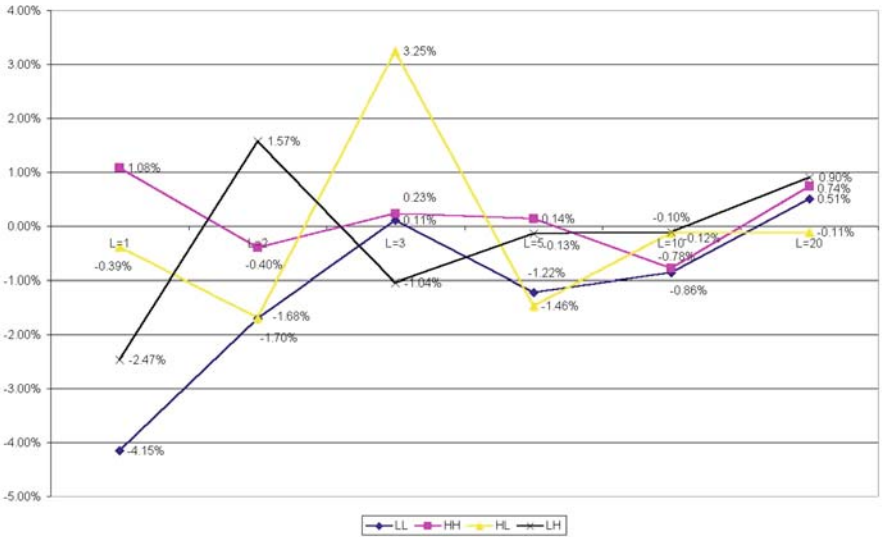

Fig. C.1 Difference in average inventory (Normal Distribution I)

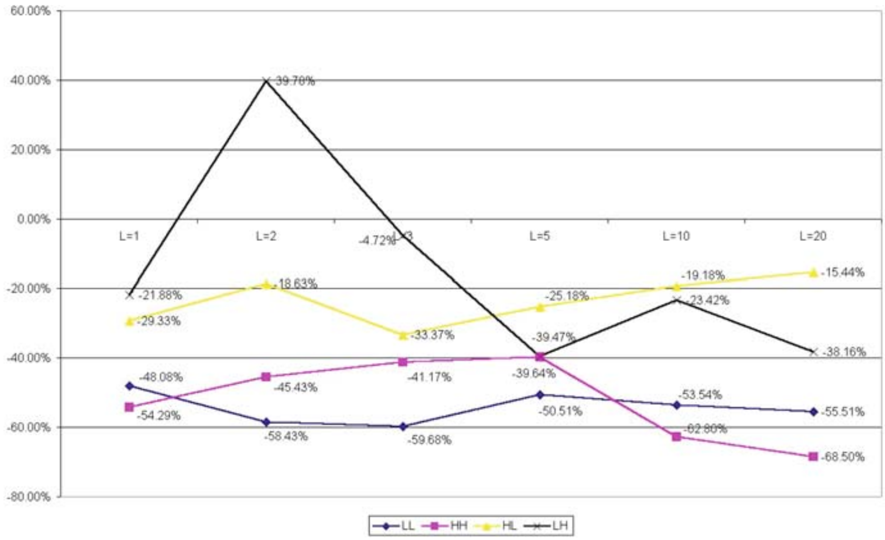

Fig. C.2 Difference in average backorder (Normal Distribution I)

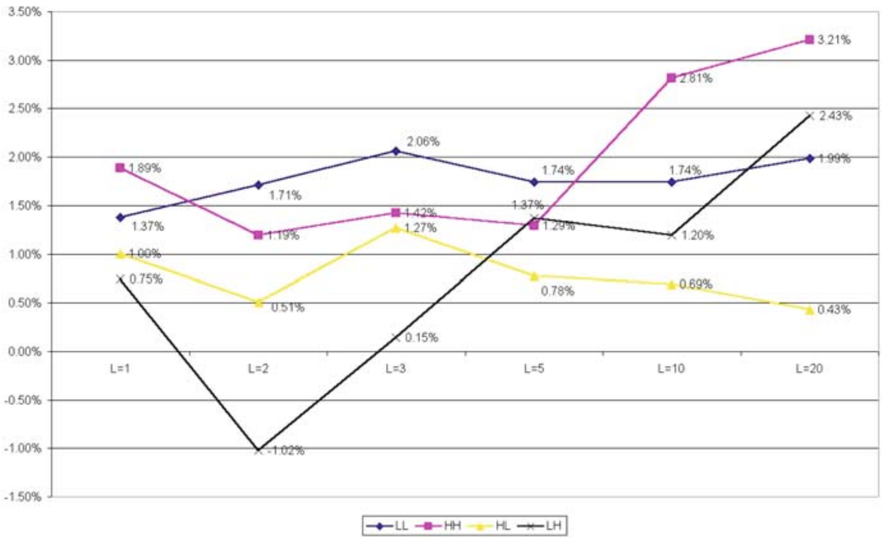

Fig. C.3 Improvement in average fill rate (Normal Distribution I)

Appendix D
Simulation Results (Normal Distribution II)

Table D.1 Simulation results 1 (Normal Distribution II)

Data	Average demand	Average inventory	Average backorder	Average order frequency	Average fill rate (%)
L = 1					
L1	10.041	10.481	0.027	0.502	99.73
L2	9.979	10.571	0.013	0.499	99.87
H1	99.435	101.695	0.57	0.497	99.43
H2	100.134	101.964	0.445	0.501	99.56
LL (non-postponement)	20.02	21.052	0.04	1.001	99.80
LL (postponement)	20.02	20.607	0.035	0.501	99.83
Change (%)	0.00	–2.11	–12.50	–49.95	0.02
HH (non-postponement)	199.569	203.659	1.015	0.998	99.49
HH (postponement)	199.569	202.283	0.439	0.499	99.78
Change (%)	0.00	–0.68	–56.75	–50.00	0.29
HL (non-postponement)	109.414	112.266	0.583	0.996	99.47
HL (postponement)	109.414	112.655	0.392	0.497	99.64
Change (%)	0.00	0.35	–32.76	–50.10	0.17
LH (non-postponement)	110.175	112.445	0.472	1.003	99.57
LH (postponement)	110.175	113.113	0.44	0.501	99.60
Change (%)	0.00	0.59	–6.78	–50.05	0.03
L = 2					
L1	10.041	20.392	0.018	0.251	99.82
L2	9.979	20.395	0.037	0.249	99.63
H1	99.435	203.56	0.435	0.248	99.56
H2	100.134	199.819	0.7	0.251	99.30
LL (non-postponement)	20.02	40.787	0.055	0.5	99.73
LL (postponement)	20.02	40.44	0.028	0.25	99.86
Change (%)	0.00	–0.85	–49.09	–50.00	0.13
HH (non-postponement)	199.569	403.379	1.135	0.499	99.43
HH (postponement)	199.569	402.705	0.461	0.25	99.77
Change (%)	0.00	–0.17	–59.38	–49.90	0.34
HL (non-postponement)	109.414	223.955	0.472	0.497	99.57
HL (postponement)	109.414	220.392	0.549	0.248	99.50
Change (%)	0.00	–1.59	16.31	–50.10	–0.07
LH (non-postponement)	110.175	220.211	0.718	0.502	99.35
LH (postponement)	110.175	218.602	0.549	0.251	99.50
Change (%)	0.00	–0.73	–23.54	–50.00	0.15
L = 3					
L1	10.041	30.529	0.035	0.167	99.65
L2	9.979	30.857	0.019	0.167	99.81
H1	99.435	306.63	0.305	0.166	99.69
H2	100.134	299.315	0.596	0.167	99.40
LL (non-postponement)	20.02	61.386	0.054	0.334	99.73

Table D.1 (continued)

Data	Average demand	Average inventory	Average backorder	Average order frequency	Average fill rate (%)
LL (postponement)	20.02	60.889	0.037	0.167	99.82
Change (%)	0.00	–0.81	–31.48	–50.00	0.08
HH (non-postponement)	199.569	605.945	0.901	0.333	99.55
HH (postponement)	199.569	603.071	0.427	0.166	99.79
Change (%)	0.00	–0.47	–52.61	–50.15	0.24
HL (non-postponement)	109.414	337.487	0.324	0.333	99.70
HL (postponement)	109.414	336.636	0.193	0.166	99.82
Change (%)	0.00	–0.25	–40.43	–50.15	0.12
LH (non-postponement)	110.175	329.844	0.631	0.334	99.43
LH (postponement)	110.175	332.321	0.528	0.167	99.52
Change (%)	0.00	0.75	–16.32	–50.00	0.09

Table D.2 Simulation results 2 (Normal Distribution II)

Data	Average demand	Average inventory	Average backorder	Average order frequency	Average fill rate (%)
L = 5					
L1	10.041	50.03	0.056	0.1	99.44
L2	9.979	50.925	0.027	0.1	99.73
H1	99.435	503.589	0.664	0.099	99.33
H2	100.134	503.154	0.435	0.1	99.57
LL (non-postponement)	20.02	100.955	0.083	0.2	99.59
LL (postponement)	20.02	101.198	0.026	0.101	99.87
Change (%)	0.00	0.24	–68.67	–49.50	0.28
HH (non-postponement)	199.569	1006.743	1.099	0.199	99.45
HH (postponement)	199.569	1007.193	0.549	0.099	99.72
Change (%)	0.00	0.04	–50.05	–50.25	0.28
HL (non-postponement)	109.414	554.514	0.691	0.199	99.37
HL (postponement)	109.414	551.08	0.357	0.099	99.67
Change (%)	0.00	–0.62	–48.34	–50.25	0.31
LH (non-postponement)	110.175	553.184	0.491	0.2	99.55
LH (postponement)	110.175	552.369	0.576	0.1	99.48
Change (%)	0.00	–0.15	17.31	–50.00	–0.08
L = 10					
L1	10.041	100.403	0.029	0.05	99.71
L2	9.979	100.82	0.022	0.049	99.78
H1	99.435	1,014.223	0.298	0.049	99.70
H2	100.134	993.792	1.073	0.05	98.93
LL (non-postponement)	20.02	201.223	0.051	0.099	99.75
LL (postponement)	20.02	200.991	0.019	0.05	99.91
Change (%)	0.00	–0.12	–62.75	–49.49	0.16
HH (non-postponement)	199.569	2,008.015	1.371	0.099	99.31
HH (postponement)	199.569	2,024.898	0.254	0.05	99.87

Table D.2 (continued)

Data	Average demand	Average inventory	Average backorder	Average order frequency	Average fill rate (%)
Change (%)	0.00	0.84	–81.47	–49.49	0.56
HL (non-postponement)	109.414	1,115.043	0.32	0.098	99.71
HL (postponement)	109.414	1,103.296	0.373	0.049	99.66
Change (%)	0.00	–1.05	16.56	–50.00	–0.05
LH (non-postponement)	110.175	1,094.195	1.102	0.1	99.00
LH (postponement)	110.175	1,096.984	0.691	0.05	99.37
Change (%)	0.00	0.25	–37.30	–50.00	0.37
L = 20					
L1	10.041	199.633	0.059	0.026	99.41
L2	9.979	200.243	0.045	0.025	99.55
H1	99.435	2010.22	0.295	0.025	99.70
H2	100.134	1,991.487	0.768	0.025	99.23
LL (non-postponement)	20.02	399.876	0.104	0.051	99.48
LL (postponement)	20.02	398.22	0.048	0.025	99.76
Change (%)	0.00	–0.41	–53.85	–50.98	0.28
HH (non-postponement)	199.569	4,001.707	1.063	0.05	99.47
HH (postponement)	199.569	4,012.991	0.347	0.025	99.83
Change (%)	0.00	0.28	–67.36	–50.00	0.36
HL (non-postponement)	109.414	2,210.463	0.34	0.05	99.69
HL (postponement)	109.414	2,220.788	0.265	0.025	99.76
Change (%)	0.00	0.47	–22.06	–50.00	0.07
LH (non-postponement)	110.175	2,191.12	0.827	0.051	99.25
LH (postponement)	110.175	2,192.591	0.698	0.025	99.37
Change (%)	0.00	0.07	–15.60	–50.98	0.12

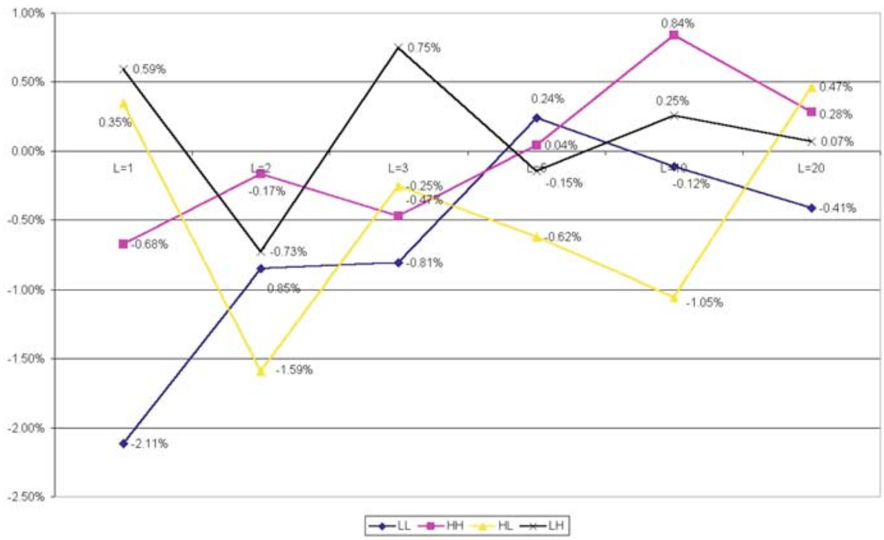

Fig. D.1 Difference in average inventory (Normal Distribution II)

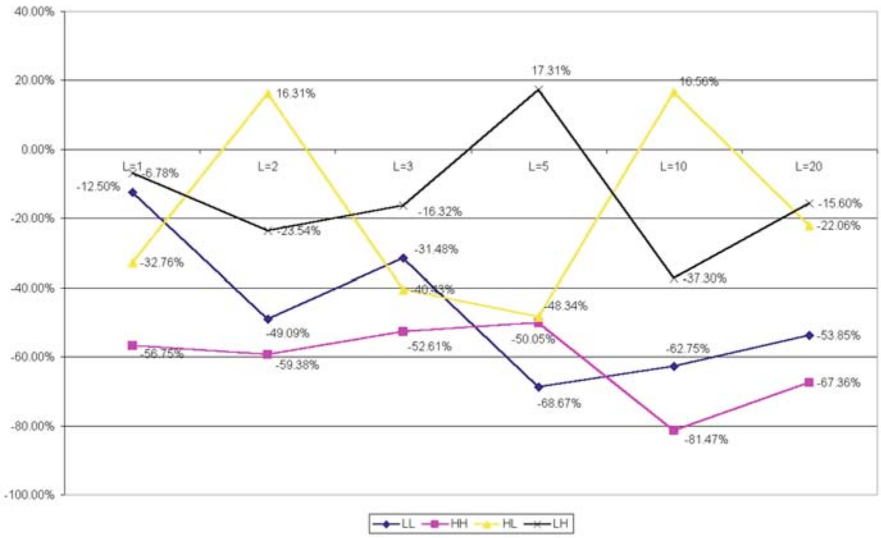

Fig. D.2 Difference in average backorder (Normal Distribution II)

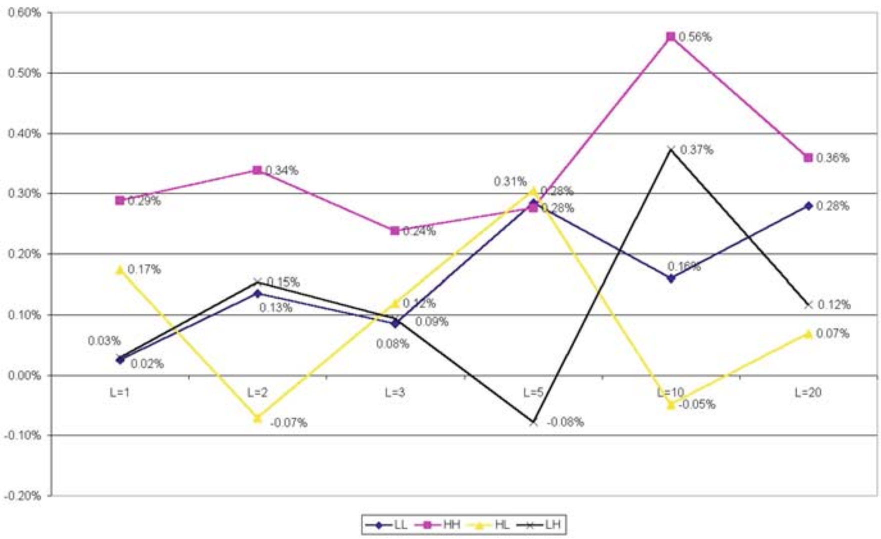

Fig. D.3 Improvement in average fill rate (Normal Distribution II)

Appendix E
Simulation Results for Cost Analysis

Table E.1 Simulation results for cost analysis

Data	Lead-	Uniform distribution			Poisson distribution			Normal distribution I			Normal distribution II		
set	time	$\alpha = 0$ (%)	$\alpha = 0.5$(%)	$\alpha = 1$(%)	$\alpha = 0$(%)	$\alpha = 0.5$(%)	$\alpha = 1$(%)	$\alpha = 0$(%)	$\alpha = 0.5$(%)	$\alpha = 1$(%)	$\alpha = 0$(%)	$\alpha = 0.5$(%)	$\alpha = 1$(%)
LL	1	–7.7	–5.6	–3.4	–3.5	–1.2	1.2	–4.1	–1.8	0.5	–2.4	–0.1	2.3
LL	2	–3.5	–2.0	–0.5	–2.6	–1.0	0.6	–2.2	–0.6	1.0	–1.1	0.5	2.2
LL	3	–3.6	–2.4	–1.2	–1.8	–0.6	0.6	–0.7	0.5	1.7	–0.9	0.4	1.6
LL	5	–0.2	0.6	1.4	–2.1	–1.3	–0.5	–1.4	–0.6	0.2	0.1	0.9	1.7
LL	10	–0.4	0.1	0.5	–1.0	–0.5	–0.1	–1.0	–0.5	–0.1	–0.1	0.3	0.8
LL	20	–0.3	–0.1	0.1	0.3	0.5	0.7	0.4	0.6	0.8	–0.4	–0.2	0.1
HH	1	–2.1	0.2	2.6	–1.8	0.6	3.1	–0.7	1.8	4.2	–0.6	1.8	4.3
HH	2	–1.1	0.5	2.1	–0.8	0.9	2.5	–0.8	0.9	2.5	–0.3	1.4	3.0
HH	3	–1.8	–0.6	0.7	–0.8	0.5	1.8	–0.3	1.0	2.2	–0.4	0.8	2.0
HH	5	–0.2	0.6	1.4	–1.2	–0.3	0.5	–0.1	0.7	1.5	0.0	0.8	1.6
HH	10	–0.6	–0.2	0.3	0.0	0.4	0.9	–1.0	–0.6	–0.1	0.7	1.2	1.6
HH	20	0.4	0.7	0.9	–1.2	–1.0	–0.8	0.5	0.8	1.0	0.2	0.5	0.7
HL	1	–2.1	0.2	2.6	–3.2	–0.8	1.6	–1.0	1.4	3.8	–0.2	2.3	4.7
HL	2	–1.2	0.4	2.0	–2.5	–0.9	0.8	–1.4	0.2	1.8	–1.1	0.5	2.1
HL	3	0.2	1.4	2.6	–1.8	–0.6	0.7	2.0	3.2	4.5	–0.3	1.0	2.2
HL	5	0.1	1.0	1.8	–0.6	0.2	1.0	–1.4	–0.6	0.3	–0.6	0.2	1.1
HL	10	–0.5	–0.1	0.4	–0.3	0.2	0.6	–0.2	0.3	0.7	–1.0	–0.5	–0.1
HL	20	–1.7	–1.4	–1.2	–0.6	–0.3	–0.1	–0.1	0.1	0.3	0.4	0.7	0.9
LH	1	–0.9	1.5	3.8	0.3	2.8	5.3	–1.9	0.6	3.0	0.0	2.5	4.9
LH	2	–0.4	1.2	2.8	0.8	2.5	4.2	1.3	3.0	4.7	–0.6	1.0	2.7
LH	3	–0.4	0.8	2.0	0.5	1.8	3.0	–0.9	0.4	1.7	0.5	1.7	3.0
LH	5	–1.9	–1.1	–0.2	–0.4	0.4	1.3	–0.4	0.4	1.3	–0.1	0.7	1.5
LH	10	0.4	0.8	1.3	0.8	1.2	1.7	–0.2	0.2	0.7	0.2	0.6	1.1
LH	20	–0.3	–0.1	0.2	–0.4	–0.2	0.1	0.7	1.0	1.2	0.1	0.3	0.5

References

1. Abboud, N. E. 2001, A discrete-time Markov production-inventory model with machine breakdowns, *Computers and Industrial Engineering*, 39, 95–107.
2. Ackerman, K. B. 1997, Postponement, *Practical Handbook of Warehousing*, 4th Edition, Chapman and Hall, New York.
3. Alderson, W. 1957, Marketing efficiency and the principle of postponement, *Marketing Behavior and Executive Action*, Richard D. Irwin, Homewood, IL pp. 423–427.
4. Arcelus, F. J., Shah, N. H. and Srinivasan, G. 2003, Retailer's pricing, credit and inventory policies for deteriorating items in response to temporary price/credit incentives, *International Journal of Production Economics*, 81, 1, 153–162.
5. Arnold, J. 1998, *Introduction to Materials Management*, 3*rd* Edition, Prentice-Hall, Upper Saddle River, New Jersey.
6. Aviv, Y. and Federgruen, A. 2001, Capacitated multi-item inventory systems with random and seasonally fluctuating demands: implications for postponement strategies, *Management Science*, 47, 4, 512–531.
7. Axsäter, S. 1997, Simple evaluation of echelon stock (R, Q) policies for two-level inventory systems, *IIE Transactions*, 29, 661–669.
8. Badinelli, R. D. 1992, A model for continuous-review pull policies in serial inventory systems, *Operations Research*, 40, 142–156.
9. Baker, K. R., Magazine, J. and Nuttle, H. L. W. 1986, The effect of commonality on safety stock in a simple inventory model, *Management Science*, 32, 8, 982–988.
10. Ballou, R. H. 1989, Heuristics: rule of thumb for logistics decision making, *Journal of Business Logistics*, 10, 1, 122–132.
11. Barkan, P. 1991, Strategic and tactical benefits of simultaneous engineering, *Design Management Journal*, spring, 39–42.
12. Beamon, B. M. 1998, Supply chain design and analysis: models and methods, *International Journal of Production Economics*, 55, 3, 281–294.
13. Bertrand, J. W. M. 1985, Multiproduct optimal batch sizes with in-process inventories and multi work centers, *IIE Transactions*, 17, 2, 157–163.
14. Bowersox, D. J. and Closs, D. J. 1996, *Logistical Management, the Integrated Supply Chain Process*, Macmillan Publishing Company, New York.
15. Bowersox, D. J. and Morash, E. A. 1989, The integration of marketing flows in channels of distribution, *European Journal of Marketing*, 23, 2, 58–67.
16. Bronson, R. 1989, *Schaum's Outline of Theory and Problems of Matrix Operations*, McGraw-Hill, New York.
17. Brown, A. O., Lee, H. L. and Petrakian, R. 2000, Xilinx improves its semiconductor supply chain using product and process postponement, *Interfaces*, 30, 4, 65–80.
18. Brown, G., Keegan, J., Vigus, B. and Wood, K. 2001, The Kellogg company optimizes production, inventory, and distribution, *Interfaces*, 31, 6, 1–15.

19. Browne, J., Harhen, J. and Shivnan, J. 1996, *Production Management Systems: An Integrated Perspective*, 2*nd* Edition, Addision-Wesley, Harlow, England.
20. Browne, S. and Zipkin, P. 1991, Inventory models with continuous stochastic demands, *The Annals of Applied Probability*, 1, 419–435.
21. Bucklin, L. P. 1965, Postponement, speculation and the structure of distribution channels, *Journal of Marketing Research*, February, 26–31.
22. Byrne, M. D. 1990, Multi-item production lot sizing using a search simulation approach, *Engineering Costs and Production Economics*, 19, 1–3, 307–311.
23. Calvin III, R. K. and Miller, W. H. 1989, Design for manufacturing, *SRC Newsletter*, 7, 1–3.
24. Cardenas-Barron, L. E. 2001, The economic production quantity with shortage derived algebraically, *International Journal of Production Economics*, 70, 3, 289–292.
25. Chen, C. K. and Min, K. J. 1994, A multi-product EOQ model with pricing consideration-T. C. E. Cheng's model revisited, *Computers and Industrial Engineering*, 26, 4, 787–794.
26. Cheng, T. C. E. 1990, An EOQ model with pricing consideration. *Computers and Industrial Engineering*, 18, 4, 529–534.
27. Cheng, T. C. E. and Podolsky, S. 1996, *Just-in-time Manufacturing: An Introduction*, 2*nd* Edition, Chapman and Hall, London.
28. Chiou, J. S., Wu, L. Y. and Hsu, J. C. 2002, The adoption of form postponement strategy in a global logistics system: the case of Taiwanese information technology industry, *Journal of Business Logistics*, 23, 107–124.
29. Christopher, M. 2000, The agile supply chain: competing in volatile markets, *Industrial Marketing Management*, 29, 1, 37–44.
30. Closs D. J., Roath A. S., Goldsby T. J., Eckert J. A. and Swartz S. M. 1998, An empirical comparison of anticipatory and response-based supply chain strategies, *International Journal of Logistics Management*, 9, 2, 21–34.
31. Cochran, J. K. and Kim, S. S. Optimal junction point location and inventory levels in serial hybrid push/pull production system, *International Journal of Production Research*, 36, 4, 1141–1155.
32. Collier, D. A. 1982, Aggregate safety stock levels and component part commonality, *Management Science*, 28, 11, 1296–1303.
33. Corry, P. and Kozan, E. 2002, Optimizing a horizontally integrated push/pull hybrid production system in a foundry, in Kozan, E. and Ohuchi, A. (editors), *Operations Search/Management Science at Work*, Kluwer Academic Publishers, Boston/Dordrecht/London.
34. Covert, R. B. and Philip, G. S. 1973, An EOQ model with Weibull distribution deterioration, *AIIE Transactions*, 5, 323–326.
35. De Hann, J., Yamamoto, M. and Lovink, G. 2001, Production planning in Japan: rediscovering lost experiences or new insight? *International Journal of Production Economics*, 71, 1–3, 101–109.
36. Dye, C. Y. and Ouyang, L. Y. 2005, An EOQ model for perishable items under stock-dependent selling rate and time-dependent partial backlogging, *European Journal of Operational Research*, 163, 3, 776–783.
37. Eilon, S. 1985, Multi-product batch production on a single machine-a problem revisited, *OMEGA, the International Journal of Management Science*, 13, 5, 453–468.
38. Ernst, R. and Kamrad, B. 2000, Evaluation of supply chain structures through modularization and postponement, *European Journal of Operational Research*, 124, 495–510.
39. Federgruen, A. and Zheng, Y. S. 1992, An efficient algorithm for computing an optimal *(r,q)* policy in continuous review stochastic inventory systems, *Operations Research*, 40, 808–813.
40. Federgruen, A. and Zipkin, P. 1984, Approximations of dynamic multi-location production and inventory problems, *Management Science*, 30, 69–84.
41. Feitzinger, E. and Lee, H. L. 1997, Mass customization at Hewlett-Packard: the power of postponement, *Harvard Business Review*, January–February, 116–121.

42. Feller, W. 1968, *An Introduction to Probability Theory and Its Applications*, Volume 1, 3*rd* Edition, John Wiley and Sons, New York.
43. Garg, A. and Lee, H. L. 1998, Managing product variety: an operations perspective, in Tayur, S., Ganesham, R. and Magazine, M. (editors), *Quantitative Models for Supply Chain Management*, Kluwer Academic Publishers, Boston/Dordrecht/London.
44. Garg, A. and Tang, C. S. 1997, On postponement strategies for product families with multiple points of differentiation, *IIE Transactions*, 29, 8, 641–650.
45. Ghare, P. M. and Schrader, G. F. 1963, A model for exponentially decaying inventory, *Journal of Industrial Engineering*, 14, 238–243.
46. Goyal, S. K. 1984, Determination of economic production quantities for a two-product single machine system, *International Journal of Production Research*, 22, 1, 121–126.
47. Goyal, S. K. and Giri, B. C. 2001, Recent trends in modeling of deteriorating inventory, *European Journal of Operational Research*, 134, 1–16.
48. Goyal, S. K. and Gunasekaran, A. 1995, An integrated production-inventory-marketing model for deteriorating items. *Computers and Industrial Engineering*, 28, 4, 755–762.
49. Goyal, S. K., Gunasekaran, A., Martikainen, T. and Yli-Olli, P. 1993, Determining economic batch sizes for multiple items in a multilevel manufacturing system, *International Journal of Operations and Production Management*, 13, 12, 79–93.
50. Graman, G. A. and Magazine, M. J. 2002, A numerical analysis of capacitated postponement, *Production and Operations Management*, 11, 3, 340–357.
51. Hadley, G. and Whitin, T. M. 1963, *Analysis of Inventory Systems*, Prentice-Hall, Englewood Cliffs, New Jersey.
52. Heskett, J. L. 1976, *Marketing*, Macmillan Publishing Co., Inc., New York.
53. Hillier, F. S. and Lieberman, G. J. 2001, *Introduction to Operations Research*, 7*th* Edition, McGraw-Hill, Singapore.
54. Hoekstra, S. J. and Romme, J. H. J. M. 1985, *Direction Towards Integral Logistics*, Kluwer, Deventer, The Netherlands.
55. Hoekstra, S. J. and Romme, J. H. J. M. 1992, *Integral Logistics Structures: developing Customer-Oriented Goods Flow*, McGraw-Hill, London.
56. Huang, C. W. and Lo, C. P. 2003, Using postponed manufacturing to reconfigure the supply chain in the desktop personal company industry: the case of Taiwan, *International Journal of Management*, 20, 241–256.
57. Jahre, M. 1995, Household waste collection as a reverse channel: a theoretical perspective, *International Journal of Physical Distribution and Logistics Management*, 25, 2, 39–55.
58. Jensen, P. A. and Bard, J. F. 2003, *Operations Research: Models and Methods*, John Wiley and Sons, USA.
59. Johnson, M. E. and Anderson, E. 2000, Postponement strategies for channel derivatives, *The International Journal of Logistics Management*, 11, 1, 19–35.
60. Kahn, B. E. 1998, Dynamic relationship with customers: high-variety strategies, *Academy of Marketing Science Journal*, 26, 1, 45–53.
61. Kanchanasuntorn, K. and Techanitisawad, A. 2006, An approximate periodic model for fixed-life perishable products in a two-echelon inventory-distribution system, *International Journal of Production Economics*, 100, 1, 101–115.
62. Kennedy, D., Elliott, B. and Carbone, F. 1998, Lean synchronous manufacturing: "hard-wiring" production to demand, in Gattorna, J. (editor), *Strategic Supply Chain Alignment*, Gower, England.
63. Lampel, L. and Mintzberg, H. 1996, Customizing customization, *Sloan Management Review*, 38, 21–30.
64. Lee, H. L. 1996, Effective inventory and service management through product and process redesign, *Operations Research*, 44, 1, 151–159.
65. Lee, H. L. 1998, Postponement for mass customization: satisfying customer demands for tailor-made products, in Gattorna, J. (editor), *Strategic Supply Chain Alignment*, Gower, England.

66. Lee, H. L. and Billington, C. 1994, Designing products and processes for postponement, in Sriram, D. and Eastman, C. (editors), *Management of Design: Engineering and Management Perspectives*, Kluwer Academic Publishers, Boston.
67. Lee, H. L., Billington, C. and Carter, B. 1993, Hewlett-Packard gains control of inventory and service through design for localization, *Interfaces*, 23, 4, 1–11.
68. Lee, H. L., Padmanabhan, V. and Whang, S. 1997, The bullwhip effect in supply chain, *Sloan Management Review*, 38, 3, 93–102.
69. Lee, H. L. and Tang, C. S. 1997, Modeling the costs and benefits of delayed product differentiation, *Management Science*, 43, 1, 40–53.
70. Lee, H. L. and Tang, C. S. 1998, Variability reduction through operations reversal, *Management Science*, 44, 2, 162–172.
71. Li, J., Cheng, T. C. E. and Wang, S. Y. 2007, Analysis of postponement strategy for perishable items by EOQ-based models, *International Journal of Production Economics*, 107, 1, 31–38.
72. Li, J., Wang, S. Y. and Cheng, T. C. E. 2008, Analysis of postponement strategy by EPQ-based models with planned backorder, *OMEGA, the International Journal of Management Science*, 36, 5, 777–788.
73. Li, S., Ragu-Nathan, B., Ragu-Nathan, T. S. and Rao, S. S. 2006, The impact of supply chain management practices on competitive advantage and organizational performance, *OMEGA, the International Journal of Management Science*, 34, 2, 107–124.
74. Ma, S., Wang, W. and Liu, L. 2002, Commonality and postponement in multistage assembly systems, *European Journal of Operational Research*, 142, 3, 523–538.
75. MacDuffie, J. P., Sethuraman, K. and Fisher, M. L. 1996, Product variety and manufacturing performance: evidence from the international automotive assembly plant study, *Management Science*, 42, 3, 350–369.
76. Martin, M., Hausman, W. and Ishii, K. 1999, Design for variety, in Ho, T. H. and Tang, C. S. (editors), *Product Variety Management: Research Advances*, Kluwer Academic Publishers, Boston/Dordrecht/London.
77. Matheus, P. and Gelders, L. 2000, The (R, Q) inventory policy subject to a compound Poisson demand pattern, *International Journal of Production Economics*, 68, 3, 307–317.
78. McClain, J. O., Maxwell, W. L., Muckstadt, J. A., Thomas, L. J. and Weiss, E. N. 1984, Comment on "Aggregate safety stock levels and component part commonality", *Management Science*, 30, 6, 772–774.
79. Melchiors, P. 2002, Calculating can-order policies for the joint replenishment problem by the compensation approach, *European Journal of Operational Research*, 141, 587–595.
80. Moinzadeh, K. and Lee, H. L. 1989, Approximate order quantities and reorder points for inventory systems where orders arrive in two shipments, *Operations Research*, 37, 277–287.
81. Moinzadeh, K. and Nahmias, S. 1988, A continuous review model for an inventory system with two supply modes, *Management Science*, 34, 761–773.
82. Nahmias, S. 1982, Perishable inventory theory: a review, *Operations Research*, 30, 680–708.
83. Nahmias, S. 2001, *Production and Operations Analysis*, 4*th* Edition, McGraw-Hill International Edition, New York.
84. Ng, C. T., Li, L. Y. O. and Chakhlevitch, K. 2001, Coordinated replenishments with alternative supply sources in two-level supply chains, *International Journal of Production Economics*, 73, 227–240.
85. Nicoll, A. D. 1994, Integrating logistics strategies, *Annual International Conference Proceedings-American Production and Inventory Control Society*, pp. 590–594.
86. Nocturne, D. J. 1973, Economic ordering frequency for several items jointly replenished, *Management Science*, 19, 9, 1093–1096.
87. O'Grady, P. J. and Byrne, M. D. 1987, Lot sizing in a practical manufacturing system, *Engineering Costs and Production Economics*, 13, 1, 13–26.
88. O'Grady, P. and Byrne, M. 1989, Adaptive lot sizing (with case study), *Engineering Costs and Production Economics*, 15, 397–404.

89. Olhager, J. 1994, On the positioning of the customer order decoupling point, Proceedings from the *1994 Pacific Conference on Manufacturing*, Jakarta, pp. 1093–1100.
90. Padmanabhan, G. and Vrat, P. 1995, EOQ models for perishable items under stock dependent selling rate, *European Journal of Operational Research*, 86, 281–292.
91. Pagh, J. D. and Cooper, M. C. 1998, Supply chain postponement and speculation strategies: how to choose the right strategy, *Journal of Business Logistics*, 19, 2, 13–33.
92. Parlar, M. and Perry, D. 1995, Analysis of a (Q, r, T) inventory policy with deterministic and random yields when future supply is uncertain, *European Journal of Operational Research*, 84, 431–443.
93. Pine II, B. J. 1993, *Mass Customization: The New Frontier in Business Competition*, Harvard Business School Press, Boston.
94. Porter, M. 1980, *Competitive Strategy: Techniques for Analyzing Industries and Competitors*, The Free Press, New York.
95. Raafat, F. 1983, A lot size inventory model for deteriorating items, in *American Institute for decision sciences twelfth annual meeting-western regional conference-proceedings and Abstracts*, pp. 141–143.
96. Raafat, F. 1988, An inventory model with a monotonically increasing deterioration rate function, in *American Institute for decision sciences seventeenth annual meeting -western regional conference-proceedings and Abstracts*, pp. 8–10.
97. Raafat, F. 1991, Survey of literature on continuously deterionrating inventory models, *Jounal of the Operational Research Society*, 42, 27–37.
98. Render, B., Stair, R. M. Jr. and Hanna, M. 2003, *Quantitative Analysis for Management*, Prentice-Hall, New Jersey.
99. Ronald, R., Yang, G. K., and Chu, P. 2004, Technical note: the EOQ and EPQ models with shortages derived without derivatives, *International Journal of Production Economics*, 92, 2, 197–200.
100. Ross, S. M. 1993, *Introduction to Probability Models*, 5*th* Edition, Academic Press, London.
101. Schroeder, R. 1993, *Operations Management: Decision Making in the Operations Function*, 4*th* Edition, McGraw-Hill, New York.
102. Shah, Y. K. 1977, An ordered level lot size for deteriorating items, *AIIE Transactions*, 9, 108–112.
103. Shu, F. T. 1971, Economic ordering frequency for two items jointly replenished, *Management Science*, 17, 6, 406–410.
104. Silver, E. A. 1976, A simple method of determining order quantities in joint replenishments under deterministic demand, *Management Science*, 22, 12, 1351–1361.
105. Simchi-Levi, D., Kaminsky, P. and Simchi-Levi, E. 2000, *Designing and Managing the Supply Chain, Concepts, Strategies, and Case Studies*, McGraw-Hill, New York.
106. Song, X. P., Cai, X. Q. and Chen, J. 2005, Studies on interaction and coordination in supply chains with perishable products: a review, in Chan, C. K. and Lee, H. W. J. (editors), *Successful Strategies in Supply Chain Management*, Idea Group Publishing, Hershey, PA pp. 222–248.
107. South China Morning Post (SCMP), 2002, *BoC Pledges 500 Shares to Every Applicant*, 22 July 2002.
108. South China Morning Post (SCMP), 2002, *Institutions Happy with Price of BoC*, 23 July 2002.
109. Stalk, G. Jr. 1988, Time-the next source of competitive advantage, *Harvard Business Review*, 66, 4, 45–51.
110. Tadikamalla, P. R. 1978, An EOQ model for items with Gamma distributed deterioration, *AIIE Transactions*, 10, 100–103.
111. Taha, H. A. 2003, *Operations Research: An introduction*, 7*th* Edition, Prentice-Hall, New Jersey.
112. Tersine, R. J. 1994, *Principles of Inventory and Material Management*, 4*th* Edition, Prentice-Hall, New Jersey.

113. Tully, S. 1993, The module corporation, *Fortune*, 8 February, 52–56.
114. Twede, D., Clarke, R. H. and Tait, J. A. 2000, Packaging postponement: a global packaging strategy, *Packaging Technology and Science*, 13, 3, 105–115.
115. Ulrich, K., Randall, T., Fisher, M. and Reibstein, D. 1999, Managing product variety, in Ho, T. H. and Tang, C. S. (editors), *Product Variety Management: Research Advances*, Kluwer Academic Publishers, Boston/Dordrecht/London.
116. van Dijk, S., van der Vorst, J. and Beulens, A. 2003, Chicken run-the poultry supply chain, in Johnston, R., Chambers, S., Harland, C., Harrison, A. and Slack, N. (editors), *Cases in Operations Management*, 3*rd* Edition, Prentice Hall, UK.
117. van Donk, D. P. 2001, Make to stock or make to order: the decoupling point in the food processing industries, *International Journal of Production Economics*, 69, 297–306.
118. van Hoek, R. 1997, Postponed manufacturing: a case study in the food supply chain, *Supply Chain Management*, 2, 2, 63–75.
119. van Hoek, R. I. 1998, Logistics and virtual integration: postponement, outsourcing and the flow of information, *International Journal of Physical Distribution and Logistics Management*, 28, 7, 508–523.
120. van Hoek, R. I. 2000, Role of third party logistics services in customization through postponement, *International Journal of Service Industry Management*, 11, 4, 374–387.
121. van Hoek, R. I. 2000, The role of third-party logistics providers in mass customization, *International Journal of Logistics Management*, 11, 37–46.
122. van Hoek, R. I., Commandeur, H. R. and Vos, B. 1998, Reconfiguring logistics systems through postponement strategies, *Journal of Business Logistics*, 19, 1, 33–54.
123. van Hoek, R. I., Vos, B. and Commandeur, H. R. 1999, Restructuring European supply chains by implementing postponement strategies, *Long Range Planning*, 32, 5, 505–518.
124. van Mieghem, J. A. and Dada, M. 1999, Price versus production postponement: capacity and competition, *Management Science*, 45, 12, 1631–1649.
125. Wan, J. C. L. 2004, Postponement Strategy in Supply Chain Management (M.Phil. Thesis), Department of Logistics, The Hong Kong Polytechnic University, Hong Kong.
126. Whitin, T. M. 1957, *Theory of Inventory Management*, Princeton University Press, New Jersey.
127. Yan, H. and Cheng, T. C. E. 1998, Optimal production stopping and restarting times for an EOQ model with deteriorating items, *Journal of Operational Research Society*, 49, 1288–1295.
128. Yeh, K. H. and Chu, C. H. 1991, Adaptive strategies for coping with product variety decisions, *International Journal of Operations and Production Management*, 11, 8, 35–47.
129. Zinn, W. 1990, Developing heuristics to estimate the impact of postponement on safety stock, *The International Journal of Logistics Management*, 1, 2, 11–16.
130. Zinn, W. and Bowersox, D. J. 1988, Planning physical distribution with the principle of postponement, *Journal of Business Logistics*, 9, 2, 117–136.
131. Zipkin, P. H. 2000, *Foundations of Inventory Management*, International Edition, McGraw-Hill, Boston.

About the Authors

T. C. Edwin Cheng Prof. T. C. Edwin Cheng is Chair Professor of Management in the Department of Logistics and Maritime Studies, The Hong Kong Polytechnic University. He is concurrently the Siyuan Chair Professor at Nanjing University in China. He obtained a B.Sc.[Eng](First Class Honours) from the University of Hong Kong, an M.Sc. from the University of Birmingham, and a Ph.D. and an Sc.D. from the University of Cambridge. He has previously taught in Canada, England and Singapore. Prof. Cheng's research interests are in Operations Management and Operations Research. He has published over 420 papers in such journals as *Management Science*, *Operations Research* and *MIS Quarterly*, and co-authored five books. He received the Outstanding Young Engineer of the Year Award from the Institute of Industrial Engineers, USA, in 1992, and the Croucher Senior Research Fellowship (the top science award in Hong Kong) in 2001. He was named one of the *Most Cited Scientists* in "All Fields", in "Computer Science", and in "Engineering" over the period 1999-2009 by the *ISI Web of Knowledge* in 2009. According to the *ISI Web of Science*, he has attained an *h*-index of 23 (i.e., having produced 23 publications each attracting 23 or more citations). He has received over 2,500 non-self citations as of April 2009. One of his papers published in the *European Journal of Operational Research* in 1989 was voted one of 30 influential articles published in that journal since its inception. A paper of the *Asia-Pacific Journal of Operational Research* in 2008 ranked him as the most productive and the highest-cited researcher in Operations Research/Management Science in Asia. A paper to appear in the *International Journal of Production Economics* ranked him as the most productive researcher and one of the top 15 *h*-index researchers in Production/Operations Management in the world in the past half-century. Over the years, Prof. Cheng has secured more than HK$30M in research funds from both public and private funding agencies to support his diverse research programmes.

Jian Li Jian Li received his Ph.D. degree in Management Sciences and Engineering from Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences in 2007. He is currently an associate professor of School of Economics and Management of Beijing University of Chemical Technology. In the past 4 years, he received several awards and honors including "President Prize of Academy of Mathematics and Systems Science" awarded by the

Academy of Mathematics and Systems Science, CAS and "President Prize of Chinese Academy of Sciences" awarded by the Chinese Academy of Sciences. He has published about 10 papers in journals including International Journal of Production Economics and OMEGA. His research interests include supply chain management and risk analysis.

C. L. Johnny Wan Johnny Wan received both his B.A.(First Class Honours) and M.Phil. degrees from The Hong Kong Polytechnic University in 1999 and 2004, respectively. Since his M.Phil. studies, he has developed a strong interest in studying postponement strategy from both mathematical and practical perspectives. He has recently completed his Ph.D. studies at the University of Cambridge, UK, focusing on the development of the supply chain strategy formulation process. He reported part of his doctoral research findings at two conferences in 2008.

Shouyang Wang (corresponding author) Prof. Shouyang Wang received his Ph.D. degree in Operations Research from Institute of Systems Science, Chinese Academy of Sciences in 1986. He is currently a Bairen distinguished professor of Management Science at Academy of Mathematics and Systems Science of Chinese Academy of Sciences and the Lotus distinguished professor management science of Hunan University at Changsha. He is also an adjunct professor of over 20 universities in the world. He is the editor-in-chief, an area editor or a co-editor of 12 journals including Information and Management, International Journal of Information Technology and Decision Making, Pacific Journal of Optimization, and Journal of Management Systems. He was/is a guest editor for a special issue/volume of over 10 journals including European Journal of Operational Research, Optimization, Engineering and Optimization, Annals of Operations Research, Dynamics of Continuous, Discrete and Impulsive Systems, International Journal of Information Technology and Decision Making, IIE Transactions, and Journal of Industrial and Management Optimization. He has published 10 monographs and over 150 journal papers. His current research interests include financial engineering, forecasting, knowledge management and decision analysis.

Index

B
Backorders, 11, 15, 20–21, 62, 64, 67
Branding postponement, 7
Bullwhip effect, 4

C
Case study, 11, 13, 17, 125, 129, 132, 134
Customer order decoupling point, 4–5

D
Demand-driven, 4, 82
Demand-pull, 4, 7
Demand-push, 4
Demand rate, 21, 34, 38–39, 45, 47, 51, 56, 63–64
Descriptive models, 11, 13
Deterioration rate, 19, 32–33, 39–41
Deterministic models, 11, 41, 133

E
Economies of scale, 1, 3, 10, 41
Economies of scope, 1, 3, 19, 41
EOQ model, 14–15, 20–21, 32, 41
EPQ model, 15, 41, 44–45, 62, 67
Exponential distribution, 34

F
Fill rate, 13–14, 111–112, 114, 118–121, 123, 140, 146, 150, 154
Flexible supply chain, 3
Forecast-driven, 4, 82
Form postponement, 7, 17, 33, 63, 125–126, 129, 130

H
Heuristic models, 11–12
Hybrid postponement strategy, 9

I
Independent system, 20, 24, 26, 30, 33, 38–39, 44, 47, 50–52, 55–56, 59, 78, 113
Inventory management, 2, 9, 11, 41

J
JIT principle, 9

L
Labeling postponement, 7, 13, 129–132
Late customization, 3
Lead-time, 4–5, 9, 12, 16–17, 21, 81–82, 84, 115, 126, 134
Logistics postponement, 4, 6–7, 9–10, 13, 17, 41, 129, 134, 136
Lot size-reorder point model, 16

M
Make-to-order, 4–5, 9, 81, 102–103, 107, 126, 128
Make-to-stock, 4, 6, 126–128
Markov chain, 16–17, 81, 83–86, 89
Mass customization, 1–2, 7, 10
Mass production, 10–11
Modular design, 2, 132

N
Non-postponement system, 15, 26, 62–66, 70–72, 75–76, 78–79, 81–84, 100, 105–107, 109–110, 113–114, 117–118
Normal distribution, 11, 109, 114, 119–123

O
Optimization technique, 19, 43, 81, 107

P
Packaging postponement, 7, 13, 129–132
Partial postponement strategy, 9, 11

Perishable product, 32–33, 41, 133
Physical supply chain, 4–5
Place postponement, 7
Point of product differentiation, 3, 5, 11–12, 20, 44, 82, 108
Poisson distribution, 16, 114, 118, 122–123
(r, q) policy, 16, 81, 84, 134
Postponement, 1, 3–4, 15, 17, 19, 42, 80, 125
Postponement system, 20, 22–24, 26, 30, 38, 44, 46, 48, 50–51, 53, 81–82, 84, 102–104, 109–110, 112, 114, 116, 118, 120, 122
Price postponement, 8–9, 17, 134, 136
Product postponement, 7
Product variety, 1–3, 12, 108, 127
Pull postponement, 4–8, 17–20, 25, 29, 41, 81, 126, 133
Push–pull boundary, 4–5, 17

R
Risk-pooling, 3, 19, 112

S
Service level, 9, 12–14, 16, 33
Simulation technique, 81, 107, 109, 134
Standardization, 1–3
Stochastic models, 11, 17, 41, 81
Stockout, 33, 43, 62, 78, 133–134
Supply chain, 1–6, 8–10, 17, 19, 43, 81, 110, 132, 136

T
Time postponement, 7, 13

U
Uniform distribution, 117, 123, 133

V
Virtual supply chain, 4–5

W
WIP, 6, 9, 11, 13, 15, 110–112, 122–123, 126