INTEGRATING MULTI-PERIOD QUANTITY FLEXIBILITY CONTRACTS WITH A CAPACITATED PRODUCTION AND INVENTORY PLANNING

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MEHMET KAYHAN

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF DOCTOR OF PHILOSOPHY
IN
INDUSTRIAL ENGINEERING

SEPTEMBER 2008

Approval of the thesis:

INTEGRATING MULTI-PERIOD QUANTITY FLEXIBILITY CONTRACTS WITH A CAPACITATED PRODUCTION AND INVENTORY PLANNING

submitted by **MEHMET KAYHAN** in partial fulfillment of the requirements for the degree of **Doctor of Philosophy in Industrial Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** ____________

Prof. Dr. Nur Evin Özdemirel
Head of Department, **Industrial Engineering** ____________

Prof. Dr. Sinan Kayalıgil
Supervisor, **Industrial Engineering Dept., METU** ____________

Prof. Dr. Nesim Erkip
Co-supervisor, **Industrial Engineering Dept., Bilkent Univ.** ____________

Examining Committee Members:

Assist. Prof. Dr. Seçil Savaşaneril
Industrial Engineering Dept., METU ____________

Prof. Dr. Sinan Kayalıgil
Industrial Engineering Dept., METU ____________

Prof. Dr. Nesim Erkip
Industrial Engineering Dept., Bilkent University ____________

Assoc. Prof. Dr. Emre Berk
Department of Management, Bilkent University ____________

Assist. Prof. Dr. İsmail Serdar Bakal
Industrial Engineering Dept., METU ____________

Date: ____________

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: MEHMET KAYHAN

Signature :

ABSTRACT

INTEGRATING MULTI-PERIOD QUANTITY FLEXIBILITY CONTRACTS WITH A CAPACITATED PRODUCTION AND INVENTORY PLANNING

Kayhan, Mehmet
Ph.D., Department of Industrial Engineering
Supervisor : Prof. Dr. Sinan Kayalıgil
Co-Supervisor : Prof. Dr. Nesim Erkip

September 2008, 192 pages

This research introduces a general approach for integrating a probabilistic model of the changes in the committed orders with an analytical model of production and inventory planning under multi-period Quantity Flexibility contracts. We study a decentralized structure where a capacitated manufacturer, capable of subcontracting, serves multiple contract buyers who actually perform forecasts on a rolling horizon basis. We model the evolution of buyers' commitments as a multiplicative forecast evolution process accommodating contract revision limits. A finite Markovian approximation to this sophisticated evolution model is introduced for facilitating the associated complex probability modeling. We implement computational dynamic programming and introduce an effective approach for reducing state-space dimensionality building upon our forecast evolution structure. Computational investigation demonstrates how the manufacturer benefits from the existence of order commitments and subcontracting option by analyzing the interplay of decisions.

Keywords: Quantity Flexibility Contracting, Order Commitment Evolution, Capacitated Production Inventory Planning, Stochastic Dynamic Programming

ÖZ

ÇOK DÖNEMLİ MİKTAR ESNEKLİĞİ KONTRATLARI İLE BÜTÜNLEŞİK KAPASİTELİ ÜRETİM VE ENVANTER PLANLAMA

Kayhan, Mehmet
Doktora, Endüstri Mühendisliği Bölümü
Tez Yöneticisi : Prof. Dr. Sinan Kayalıgil
Ortak Tez Yöneticisi : Prof. Dr. Nesim Erkip

Eylül 2008, 192 sayfa

Bu çalışma, çok dönemli Miktar Esnekliği kontratları altındaki üretim-envanter planlarının, sipariş taahhüt güncellemelerinin olasılık modeli ile entegrasyonuna genel bir yaklaşım sunmaktadır. Merkezi olmayan karar yapısı altında, kapasitesi sınırlı aynı zamanda fason üretimi teslim alabilen bir imalatçının, talep tahminlerini yuvarlanan ufuk bazlı belirleyen kontratlı alıcılarla çalıştığı bir sistem ele alınmıştır. Sipariş taahhüt güncellemelerinin zaman içinde nasıl geliştiği martingale talep tahmin evrim modeli ile modellenmiş ve kontrat revizyon limitleri bu modele dahil edilmiştir. İlgili karmaşık olasılık modelini kolaylaştırmak amacıyla, bu gelişmiş tahmin evrim modeline Markov rassal süreç bazlı yaklaştırma geliştirilmiştir. Rassal dinamik programlama çözüm yaklaşımı benimsenmiş ve tahmin evrim modeli temel alınarak, ilgili durum uzayı boyutlarını etkili şekilde azaltan bir yaklaşım sunulmuştur. Hızlı ve verimli çalışan bu rassal dinamik programlama kullanılarak, imalatçının sipariş taahhütleri ve fason üretim seçeneği varlığından nasıl fayda sağlayabileceği meselesi, imalatçının ilgili kararları arasındaki karşılıklı etkileşim analiz edilerek, deneysel olarak incelenmiştir.

Anahtar Kelimeler: Miktar Esnekliği Anlaşması, Sipariş Taahhüt Evrimi, Kapasiteli Üretim Envanter Planlama, Rassal Dinamik Programlama

To my beloved family

ACKNOWLEDGMENTS

I would like to express my deep gratitude to my supervisors Prof. Dr. Sinan Kayalıgil and Prof. Dr. Nesim Erkip for their valuable guidance and support. They contributed a lot to my study with their great insights, perspectives, experience and encouraging attitudes. I am fully aware that I can still learn a lot from you.

I sincerely appreciate Seçil Savaşaneril, Emre Berk and İsmail Serdar Bakal for serving on my dissertation committee and for their valuable contributions and helpful suggestions.

Most special thanks go to my beloved family. My mother Gülsen, my father Cemil, my sister Özlem, my brother Ali Murat and my little princess Yaren never stopped being supportive. My Princess deserves sincere appreciation for encouragements and everlasting emotional support she provided whenever it is actually needed. I could not have accomplished this work, among many other things, without them. I am quite aware of your unconditional love, endless support and great understanding.

Finally, I am truly grateful to all my friends and colleagues. One way or another they made this process much more easy. I am thankful to all those who gave me the support in difficult times and the possibility to complete this study.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

LIST OF ABBREVIATIONS

B
: the number of contract buyers involved.

N
: the number of periods with replenishment decision.

L
: the length of the manufacturer's replenishment lead-time.

H
: the length of commitment horizon.

$d^b_{s,s+k}$
: the order commitment submitted by buyer b at the end of period s for the amount to be ordered in period $s+k$, $k = 1, 2, \ldots, H$. It is denoted by $D^b_{s,s+k}$ when it is random.

$d^b_{s+k,s+k}$
: the eventual order realization from buyer b in period $s+k$. It is denoted by $D^b_{s+k,s+k}$ when it is random.

$\mathbf{d^b_s} = [\, d^b_{s-1,s+k-1},\ \forall\, k \in \{1, 2, \ldots, H\}\,]$
: the vector of order commitments submitted by buyer b at the end of period $s-1$. It is denoted by $\mathbf{D^b_s}$ when it is random.

$\mathbf{d_s} = [\, \mathbf{d^b_s},\ \forall\, b \in \{1, 2, \ldots, B\}\,]$
: the vector of all order commitments available to the manufacturer at beginning of period s. It is denoted by $\mathbf{D_s}$ when it is unknown.

μ_{D^b}
: the expected value of $D^b_{s,s}$.

$\varepsilon^b_{s,k}$
: the k-step order commitment update from buyer b at the end of period s (made to $d^b_{s-1,s+k-1}$, $k = 1, 2, \ldots, H+1$). It is defined as the difference of logarithms and has the units of $\ln d^b_{s-1,s+k-1}$.

$\tilde{\varepsilon}^b_{s,k}$
: the k-step intended order commitment update of buyer b in period s.

$\boldsymbol{\mathcal{E}_s^b} = \left[\, \varepsilon_{s,k}^b,\ \forall\, k \in \{1,\, 2,\, \ldots,\, H+1\} \right]$

the vector of commitment updates from buyer b at the end of period s.

$\boldsymbol{\mathcal{E}_s} = [\, \boldsymbol{\mathcal{E}_s^b},\ \forall\, b \in \{1,\, 2,\, \ldots,\, B\}\,]$

the vector of all commitment updates made at the end of period s.

$(\omega_k^b,\, \alpha_k^b)$

the lower and upper flexibility limits to the modification of k-step order commitments from buyer b. We have $\boldsymbol{\mathcal{W}^b} = \{\, \omega_k^b, \quad k = 1,\, 2,\, \ldots,\, H\,\}$ and $\boldsymbol{\mathcal{A}^b} = \{\, \alpha_k^b, \quad k = 1,\, 2,\, \ldots,\, H\,\}$.

$\mu_{\varepsilon_k^b}$

the expected value of the kth component of $\boldsymbol{\mathcal{E}_s^b}$. We have the mean vector $\boldsymbol{\mu_{\mathcal{E}^b}} = [\, \mu_{\varepsilon_k^b} = 0,\ \forall\, k \in \{1,\, 2,\, \ldots,\, H+1\}\,]$.

$\Sigma_{\boldsymbol{\mathcal{E}}}$

the variance-covariance matrix of $\boldsymbol{\mathcal{E}_s}$. The variance of the kth component of $\boldsymbol{\mathcal{E}_s^b}$, and the covariance between the kth component of $\boldsymbol{\mathcal{E}_s^b}$ and the lth component of $\boldsymbol{\mathcal{E}_s^r}$ are denoted by $\sigma^2_{\varepsilon_k^b}$ and $\sigma_{\varepsilon_k^b,\, \varepsilon_l^r}$, respectively.

$f_{\boldsymbol{\mathcal{E}_s}}(\cdot \,|\, \boldsymbol{\mathcal{W}},\, \boldsymbol{\mathcal{A}})$

the multivariate probability function of $\boldsymbol{\mathcal{E}_s}$. $F_{\boldsymbol{\mathcal{E}_s}}(\cdot \,|\, \boldsymbol{\mathcal{W}},\, \boldsymbol{\mathcal{A}})$ is its cumulative probability.

$\mathbf{R_s} = [\, e^{\boldsymbol{\mathcal{E}_s^b}},\ \forall\, b \in \{1,\, 2,\, \ldots,\, B\}\,]$

the vector of all multiplicative commitment updates made at the end of period s. It has the units of $d_{s,s}^b$.

$\mathcal{F}_s$

the smallest σ-field generated by all observed events up to time s (like order commitments and realized orders) such that $\mathcal{F}_s \subseteq \mathcal{F}_{s+1}$.

U_k^b

the sum of logarithmic changes in successive order commitments submitted by buyer b over the following k-period time interval, $k = \{1,\, 2,\, \ldots,\, H+1\}$

M

the number of disjoint sub-intervals in discretization.

$\aleph_k^b$

the set of discrete states corresponding to U_k^b, labeled $1,\, 2,\, \ldots,\, M$.

$\aleph_{k,m}^b$

the midpoint value of the sub-interval for state m, with the exception for the first and final states which represent lower and upper censoring points, respectively.

$f_{U_k^b}(\cdot \,|\, \boldsymbol{\mathcal{W}^b},\, \boldsymbol{\mathcal{A}^b})$

the probabilities of U_k^b in each of M categories. $F_{U_k^b}(\cdot \,|\, \boldsymbol{\mathcal{W}^b},\, \boldsymbol{\mathcal{A}^b})$ is its cumulative distribution.

$\mathbf{JTP_{j,k}}$

the $(k-j)$-step joint transition probability matrix $(M^B \times M^B)$ of the multivariate Markov chain $\{(U_k^1, \ldots, U_k^B),\ 1 \leqslant k \leqslant H+1\}$.

$\mathbf{JP_1}$

the initial probability distribution of the multivariate Markov chain.

q_s

replenishment amount placed in period s for delivery in period $s+L$.

I_s

inventory position before ordering in period s.

TI_s

inventory position after ordering in period s, $TI_s \geqslant I_s$.

NI_s

net inventory level (i.e., on-hand inventory or backorders) to be carried over in period s.

K

finite per-period capacity for in-house production.

h

unit cost of carrying inventory per period.

π_b

unit backorder penalty per period for buyer b. We have $\pi_1 > \pi_2$.

c_{pi}

unit cost of in-house production.

c_{ps}

unit cost of subcontracting. We have $c_{ps} \geqslant c_{pi}$.

$\mathbf{1}(A_s)$

indicator function of the event $A_s = \{TI_s - I_s > K\}$, which is 1 if it is true, 0 otherwise.

$PC_s(TI_s,\, I_s)$

the replenishment cost to be incurred for an amount $q_s = TI_s - I_s$.

$L_s(TI_s,\, \mathbf{d_s})$

the current L-period costs associated with inventory carrying and backorders, given that the inventory position is set to TI_s after the manufacturer's ordering in period s.

$J_s(TI_s,\, I_s,\, \mathbf{d_s})$

the current-period cost associated with period s, $PC_s(TI_s,\, I_s) + L_s(TI_s,\, \mathbf{d_s})$.

$Z^b_{[s,s+L)}$

the total of L order occurrences to be received from buyer b over the time interval $[s,\, s+L)$. We have $Z_{[s,s+L)} = \sum_b Z^b_{[s,s+L)}$.

$f_{Z_L^b}(\cdot \mid \mathbf{d_s})$

the probability function of $Z^b_{[s,s+L)}$, conditioned on the value of order commitment vector $\mathbf{d_s}$ available at the beginning of period s. $F_{Z_L^b}(\cdot \mid \mathbf{d_s})$ is its cumulative distribution function.

$\pi = \{TI_s,\ 1 \leqslant s \leqslant N\} \in \Pi$

an ordered set of replenishment decisions (a policy), there being one decision for each system state $(I_s,\ \mathbf{d_s})$ at the beginning of period s.

$G_1(I_1,\ \mathbf{d_1},\ \pi)$

the total expected cost over a finite time horizon of $N+L$ periods, given that the policy π is being used and the initial system state is observed as $(I_1,\ \mathbf{d_1})$.

$V_1(I_1,\ \mathbf{d_1})$

the *minimum* expected total cost from the beginning of period 1 to the end of period $N+L$, given that the system state is observed as $(I_1, \mathbf{d_1})$ at the beginning of period 1 and that an optimal decision is made in each period 1, 2, ..., N.

$G_s(TI_s,\ I_s,\ \mathbf{d_s})$

the *(suboptimal)* expected total cost from the beginning of period s to the end of period $N+L$, given that the system state is observed as $(I_s,\ \mathbf{d_s})$ at the beginning of period s.

TI_s^*

the value of TI_s at which the minimal value is attained by $G_s(TI_s,\ I_s,\ \mathbf{d_s})$.

$G_s^{inh}(TI_s,\ I_s,\ \mathbf{d_s})$

the expected total cost from period s through $N+L$, given that the system is in state $(I_s,\ \mathbf{d_s})$ and only the in-house capacity is used at the beginning of period s.

TI_s^{inh}

the minimizer of the cost function $G_s^{inh}(TI_s,\ I_s,\ \mathbf{d_s})$.

$G_s^{sub}(TI_s,\ I_s,\ \mathbf{d_s})$

the expected total cost from period s through $N+L$, given that the system is in state $(I_s,\ \mathbf{d_s})$ and both in-house and subcontract capacity are engaged at the beginning of period s.

TI_s^{sub}

the minimizer of the cost function $G_s^{sub}(TI_s,\ I_s,\ \mathbf{d_s})$.

β_k^b

the expected order after k updates from buyer b, given by $\mu_{D^b}\ e^{(\mu_{\varepsilon_1^b}+\cdots+\mu_{\varepsilon_k^b})}$. It has the same units as $Z^b_{[s,s+L)}$.

θ_k^b
: the proportion of β_k^b in the expected aggregate L-period buyer order, given by $\beta_k^b / \sum_{m=1}^{L} \beta_m^b$. It is dimensionless.

λ_k^b
: the weight factor of k-step commitment update, given by $\sum_{m=k}^{L} \theta_m^b$. It is dimensionless. $\lambda_1^b = 1$ and it decreases as k increases from 1 to L.

ϑ_L^b
: the weighted aggregate L-period commitment update. It denotes a weighted sum of commitment updates (on the logarithmic scale) to be received from buyer b over the time interval $[s,\, s+L)$, given by $\sum_{k=1}^{L} \lambda_k^b \varepsilon_{s,k}^b$. It has the units of $\ln\, Z_{[s,s+L)}^b$.

$f_{\vartheta_L^b}(\cdot)$
: the probability function of ϑ_L^b.

$\mathfrak{D}_{s,H}$
: an indicator of weighted effects by an observed order commitments $\mathbf{d_s}$, given by $\sum_b \sum_{k=1}^{H} \theta_k^b \ln\, d_{s-1,s+k-1}^b$.

$\widehat{I}_s$
: the modified inventory position before ordering in period s, as a result of state-space reduction.

$\widehat{V}_s(\cdot)$
: the modified cost-to-go function which is a function of the modified inventory position $\widehat{I}_s$ and the $\mathfrak{D}_{s,H}$ statistic.

$\widehat{TI}_s$
: the modified inventory position after ordering in period s.

$\widehat{J}_s(\widehat{TI}_s,\, \mathfrak{D}_{s,H})$
: the current-period costs associated with state $(\widehat{I}_s,\, \mathfrak{D}_{s,H})$ when action $\widehat{TI}_s$ is selected.

$\widehat{NI}_s$
: the modified net inventory level to be carried over in period s.

CO
: correlations of commitment updates across buyers and through time.

CV
: coefficient of variation for demand.

FL
: the level of flexibility limit per period. We have $FL = \omega_k^b = \alpha_k^b$, if not stated otherwise.

Δ_K
: the capacity slack representing the amount of excess capacity over the expected total of orders to be received from all the buyers per lead-time.

Δ_c
: the cost differential between the in-house production and subcontracting, $c_{ps} - c_{pi}$.

Δ_π
: the backorder-to-holding cost ratio.

CI_H^+
: the manufacturer's percentage cost saving over the minimal-commitment base case ($H = 1$ and $FL = \infty$).

CI_H^-
: the manufacturer's percentage cost increase over the minimal-flexibility base case ($H = k$ and $FL = 0.01$).

CI_K
: the manufacturer's percentage cost improvement over the capacity base case ($\Delta_K = 0$).

TI_{dev}
: the mean order-up-to deviation denoting average ratio of optimal order-up-to positions to the mean order per lead-time.

TI_{ins}
: the order-up-to instability denoting the average absolute deviation between optimal order-up-to positions, in fraction of the mean order per lead-time, of consecutive decision periods.

CU
: the capacity utilization denoting the extent to which the manufacturer actually uses his in-house production capacity.

φ
: the fill rate, as type-2 service level, denoting the expected proportion of total realized order over all the buyers that is satisfied immediately from the manufacturer's finished-goods inventory.

$\daleth^b(\boldsymbol{\mathcal{A}^b})$
: the expected loss due to limited flexibility buyer b may suffer when she is offered the set $\boldsymbol{\mathcal{A}^b} = \{\, \alpha_k^b,\ \ k = 1, 2, \ldots, H \,\}$ of flexibility limits.

$\boldsymbol{\Theta}$
: the 2×2 unknown matrix $\left[\begin{smallmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{smallmatrix}\right]$ of moving average parameters for the bivariate ARIMA(0, 1, 1) demand process.

$\boldsymbol{\eta_s}$
: the 2×1 random vector $\left[\begin{smallmatrix} \eta_s^1 \\ \eta_s^2 \end{smallmatrix}\right]$ of disturbances in period s for the bivariate ARIMA(0, 1, 1) demand process.

$\Sigma_{\boldsymbol{\eta}}$
: the covariance matrix of random disturbances $\boldsymbol{\eta_s}$. The variance of the disturbances corresponding to the bth buyer and the covariance between the disturbances corresponding to the bth and rth buyers are denoted by $\sigma^2_{\eta^b}$ and σ_{η^b,η^r}, respectively.

F^b_{s+j}
: the first-order EWMA forecast for the order quantity to be submitted by buyer b in period $s + j$, made after observing the order of period s.

q_s^{ma}
: the replenishment order placed in period s for delivery in period $s + L$ under the ARIMA-based inventory management.

CHAPTER 1

INTRODUCTION

Decentralized decisional structure with demand uncertainty is a reality for supply chains in many industrial contexts. The entities in such supply chains observe only the local information structures available under some perceived risks. They pursue different and possibly adversarial objectives, and maximize their own performance metrics. It is a *traditional practice* to use a set of corporate rules which are mainly based on performance metrics in coordinating the corresponding operational decisions. Although these rules are intended to align the incentives of the entities, they may not always effectively compensate the entities for the risks that they assume. Consequently, manufacturing companies may shoulder the largest burden of risk, as they must invest in capacity and inventory in advance in the face of unreliable demand information caused especially by distortions. Downstream entities, on the other hand, may not be the solely responsible for the quality of demand information to be confronted by the manufacturing companies, and hence, they do not assume a position to reduce the risks on their own. Lee et al. (1997) provide several industrial examples of these inefficiencies and analyze their potential causes.

Traditionally this situation is mitigated by price and quota adjustments. However, their success have not proven to be sufficient. For instance, Lariviere and Porteus (2001) study price-only arrangements and conclude that they alone cannot coordinate a decentralized supply chain due to double marginalization. Gerchak and Wang (2004), on the other hand, show that wholesale price plus buy-back arrangements can achieve coordination in an assembly system. The bottom line in this stream of research is that pricing strategies (and other traditional practices) alone are not sufficient for the efficient operation of the supply chain. More comprehensive schemes for coordination among the supply chain entities are also required.

In this context, the way in which the *relationships between entities* are structured is a major direction for improvement in better aligning the incentives and more equitably sharing the risks across a supply chain. With a growing need to survive in a highly volatile and dynamic business environment, these relationships are becoming more dependent on factors like *commitment* to cooperate on a common purpose making each entity better off and *flexibility* to effectively respond to changing market conditions, as opposed to traditional relationships based mainly on price discounting and quota arrangements. This change of focus requires the entities to look beyond the flexible organization to the *flexible supply chain*. As a consequence, proactive means of reducing supply chain uncertainty are becoming an important consideration in these relationships, instead of simply reactive means of coping with uncertainty. This necessitates research directed toward a better understanding of the drives and constraints that the entities face with flexibility capabilities.

In this research, within a general class of mechanisms for structuring the relationships across a supply chain, we are particularly interested in *contractual agreements*. Contracting in general enables a downstream supply chain entity to purchase products from an upstream entity for a specified period of time under specific terms and conditions defining how the contract is implemented. Contractual agreements in supply chain management range from simple price-only contracts to more sophisticated coordinating contracts possibly with returns policies (see Tsay et al., 1999 for a review of literature on supply contracts). The use of supply contracts is frequent in several industries such as consumer electronics, apparel, fast moving consumer goods, automotive and electronic industrial markets. These sectors are characterized by the level of competition and relatively large volume demand with variety and rather short lead-times. Contracts are essentially intended to improve cooperation between the contract participants, and the generally accepted perspective is becoming to constitute an improved supply chain flexibility and responsiveness. Sánchez and Pérez (2005), for instance, analyze the relationship between the dimensions of supply chain flexibility and firm performance in a sample of automotive suppliers.

The type of supply contract to be implemented depends on the operating environment and the underlying behavioral dynamics. In this research, we particularly focus on the supply contracts containing terms of *quantity commitments* and *flexi-*

bility. The quantity commitments with flexibility is particularly effective when the demand environment is highly uncertain and changing. We assume that quantity commitments are made on *periodical orders* to be purchased in a number of future periods. The flexibility, on the other hand, defines the limits on the range of allowable changes to these commitments and their frequency. This structure is referred to as *quantity flexibility (QF) contract* (see Eppen and Iyer 1997, Anupindi and Bassok 1998, Tsay and Lovejoy 1999, Tsay 1999, Sethi et al. 2004, Wu 2005). We preferred to study the QF contracts as quantity flexibility is an important concept for providing a wider insight into flexibility in the supply chain, not just in the first-tier relationships and the manufacturing function. It can provide some level of stability for the upstream entities and assist the downstream entities in responding to market demand fluctuations.

The QF structure is essentially intended to make contract participants better off. It appeals to the downstream entity since it places restrictions on the risk level she is prepared to bear, passing some portion of the expected cost associated with demand uncertainty on to upstream entities. The upstream entity, on the other hand, is interested in this structure either when the marginal cost of his production is low, but the fixed costs are high or when there are other competitive manufacturers with uncertainty about which manufacturer the downstream entity will select in the short run.

The supply chain setting that will be analyzed in this research is as follows. We consider a decentralized supply chain that consists of a single manufacturer and multiple buyers. The manufacturer produces and sells a single item to the buyers, who in turn serve an end market with stochastic market demands. The buyers are differentiated by their logistical and service requirements. Hence, the manufacturer offers each buyer separate multi-period QF contracts. There exist some restrictions on the maximum amount that can be produced internally. But, it is also possible to subcontract the production with an outside supplier. Subcontracting arrangements are more costly than internal production, but with an infinite-supply opportunity. Both in-house and subcontract replenishment quantities are assumed to arrive only after a known, constant lead-time. We address a more detailed description of the problem environment in Chapter 2. Therein, we describe the organizational environment as well as the buyers' ordering and manufacturer's planning operations that provide activities serving as the underpinning in managing the aggregate production and inventory planning under QF contracts.

The problem is characterized by the presence of random commitment evolution with limited revision flexibility, on the demand side, and production capacity restrictions with an option of subcontracting, on the supply side (we do not assume any contractual relation with the subcontractor although this will constitute an interesting issue to study). These are two complementary features that contribute to the supply chain flexibility. Even though a considerable amount of research has been devoted to flexibility considerations, most of it has been confined to internal manufacturing flexibility. A growing body of literature has begun to recognize the importance of flexibility capabilities of the entire supply chain. Vickery et al. (1999), for instance, empirically examine the dimensions of supply chain flexibility and their relationships with the environmental uncertainty, business performance, and functional interfaces. Stevenson and Spring (2007) provide a comprehensive review of the related literature. However, the issue of analyzing flexibility capabilities of the supply chain in the context of QF contracts and its impacts on system performance still offers a valuable research opportunity.

In this context, we study the interactions of the manufacturer and his contract buyers from the perspective of aggregate production and inventory planning. We do not attempt to make an explicit specification of the buyers' objective function and constraints. Rather, the focus of our investigation will be on the decision problems that arise for the manufacturer. We preferred to do this as the manufacturer faces a more involved and complicated decision. It can be thought as a first step to start the overall analysis. We think if the manufacturer's alternatives and consequences are analyzed well, a centralized scheme can later be studied extensively.

In the aggregate production and inventory planning framework under QF contracts, each buyer submits order commitments to the manufacturer for a number of future periods once every period. As new information becomes available to the buyers, they are allowed to update order commitments over time in accordance with the contracts, which eventually (i.e., at the end of update series) become the realized orders. In turn, the manufacturer delivers finished goods into stock according to forecast-driven production planning in the face of order commitments with stochasticity in their updates. This is like the manufacturer having a forecasting engine although it is the buyers who actually perform the forecasts.

QF contracting-type relationships represent one structured way of communicating early information on uncertain future demands. This is a form of *advance order information* (see Karaesmen et al. 2002 and Gallego and Özer 2001). The value of learning about buyer orders in advance is considerable in many industries. It improves the manufacturer's understanding of buyers' orders, and in turn enhances supply-side flexibility capabilities of the supply chain. Advance order information under QF contracts is in the form of periodical quantity commitments on orders to be submitted by the buyers. The QF contract requires the buyer to commit to purchase a minimum amount specified for a particular period while the manufacturer guarantees an upside coverage to the buyer of a certain percentage above her minimum commitment. Previous quantity commitments are allowed to be updated, as more information is available to the buyer from one period to the next, and the flexibility limits stipulated in the contract govern these updates. This requires the manufacturer to strike a balance between the quality of order information and the costs of production and inventory processes. Hence, developing methods for an effective use of early order information in the planning process becomes a major issue.

The related literature has focused primarily on contract design and incentives, information structures and its implications, methods for coordination and mitigating system inefficiencies, and the competition issues. However, the issue of integrating order information updates with planning operations remains as an important research area. The integrated analysis may assist the manufacturer in making better capacity and materials procurement decisions. It can lessen the effects of order variability throughout the supply chain. This may in turn lead to production quantities fluctuating less from one period to another and a superior customer service.

The basic motivations behind our research are several in the light of the aforementioned issues. (*i*) First, we are concerned with how the manufacturer can determine an appropriate scheme of QF contracts to be offered to the buyers. This necessitates addressing the parameter-setting problem for QF contracts. (*ii*) Second, we are interested in how the manufacturer's capacity investment decision making can benefit from the presence of order commitments from a certain range in the presence of a subcontracting option. This is concerned with how effective the level of internal capacity slack is as compared to the flexibility of QF contracts and the attractiveness of the subcontracting option. (*iii*) The third is how to adjust the manufacturer's production and inventory planning according to the buyers' order commitments with stochasticity in updates under QF contracts. This necessitates capturing the variation of commitment updates and the associated correlation structure. (*iv*) The fourth is related with how a particular, refined forecast update scheme differs from a more conventional time series approach in representing the underlying stochastic framework. As a consequence, all these issues require an integrated use of a model of the changes in the committed orders with an analytical model of the production and inventory planning. The integration has a system-view (holism rather than reductionism) intent, and the hope is that it can permit more effective management of the production and inventory system.

A key component in the dynamics of such an integrated use is the way in which the underlying stochastic framework is represented. Several approaches for modeling the uncertainty associated with buyers' orders exist. (*i*) Distribution-based models assume that the true mathematical form for the distribution of orders is known with estimated distribution parameters; see, e.g., Dvoretzky et al. (1952a). The most important case of the distribution-based models is Markov-modulated approach. In this approach, the demand process is driven by an underlying Markov chain. Chen and Song (2001) study multi-echelon inventory system with Markov-modulated demand where the demand distribution in each period is determined by the current state of an exogenous discrete-time Markov chain. Sethi and Cheng (1997) exemplify this kind of demand process for single location inventory literature. (*ii*) Bayesian models assume that we have some prior opinion about buyers' orders, and actual order realization would resolve some of the uncertainty as time passes and some kind of 'learning' takes place. Therein, the order process is assumed to be independent and identically distributed with an unknown probability distribution function. Dvoretzky et al. (1952b) first introduces Bayesian models in the inventory literature. Scarf (1959) and Azoury (1985) later study Bayesian updating mechanisms to learn about future demand from past history in forecasting/inventory models. (*iii*) Time-series models assume that the sequence of order realizations forms an autoregressive moving-average model. Graves (1999) and Lee et al. (2000) specifically study ARIMA(0, 1, 1) and AR(1) models of the demand process, respectively. Aviv (2003) provides a structural time-series framework for inventory management that is much more general than the specific cases of autoregressive time-series models.

All of the above approaches, however, are not adequate for taking into account the evolution of order commitments. The essence of evolution is interpreted from the buyer's perspective. Every buyer has a forecasting machinery that processes her observations on the market demand and generates some advance order information to be submitted to the manufacturer. The buyers commit for their orders a certain number of periods into the future, and the committed orders evolve over time. The order commitments are non-stationary by their very nature of periodic updates. The commitment updates are not insensitive to one another so that there exist correlations across buyers and through time. Furthermore, due to the flexibility terms of QF contracts, order commitments are updated in relation to the earlier commitments made.

As time passes, additional information in the form of updated order commitments becomes available to the manufacturer. Every time series thus is a series of updates. This necessitates a forecasting engine whereby the manufacturer generates and revises his forecasts on buyers' order information. The manufacturer, in turn, is concerned with forecasting buyers' future orders and with the management of discrepancies between his forecast and actual order realizations later. This practice requires the statistical knowledge of commitment updates to be an important consideration when applying inventory models. The resultant is an evolution-based inventory management. It includes a probabilistic model of how order commitments evolve into the future as an integral part of production and inventory planning. The evolution-based inventory management fits to today's supply chain environment more effectively as compared to conventional distribution-based inventory management. This is because the success is becoming more dependent on flexibility of the supply chain to react to evolving business conditions, which contributes to the improvement of proactiveness and responsiveness.

The above essential aspects should be accommodated in the stochastic framework, as performed in this research, intending to represent the underlying ordering behavior of the buyers adequately. To be specific, we assume that the manufacturer models the time series of the buyers' order commitments and realized orders as a *multiplicative forecast evolution process*. As a theoretical framework, we consider the martingale model of forecast evolution (MMFE) methodology due to Graves et al. (1986) and Heath and Jackson (1994). This structure can include non-stationary and correlated demands. It is quite general and accommodates judgmental forecasting as well as conventional time-series models (see Güllü 1993). We extend the evolution model to accommodate the revision limits stipulated in the QF contracts. Several other studies have adopted the MMFE to investigate production/inventory planning issues and its impact on inventory cost (Güllü 1996, 1997, Toktay and Wein 2001, Gallego and Özer 2001, Çakanyıldırım and Roundy 2002, Kaminsky and Swaminathan 2004, Kayhan et al. 2005, Iida and Zipkin 2006). Our research differs from these studies in several ways. These are (*i*) the use of multiplicative evolution model that is more effective than an additive form of evolution, (*ii*) the inclusion of contract revision limits in the evolution model, and (*iii*) the order-forecasting role of the evolution model (it is used in the manufacturer's forecasting while he has no information on the market demands faced by the buyers who actually perform forecasts).

Note that the inclusion of revision limits further complicates the probabilistic framework since the standard techniques of probability and statistics do not apply as it brings up cut-offs and hence lumpiness in probability mass. We attempt to resolve this complication by introducing a finite Markov chain approximation to the evolution model. Chapter 3 addresses these issues in more detail.

Although the stochastic multi-period and multi-dimensional nature of the problem, we develop an optimization-based approach rather than an experimental analysis through a simulation model of the production/inventory system. We formulate a finite-horizon stochastic production and inventory model minimizing the expected total cost of the manufacturer. The model assumes that buyers' order commitments are explicit component of the system state. We then consider dynamic programming to characterize the structure and properties of optimal replenishment policies of the manufacturer, as mentioned in Chapter 4.

As our solution procedure will have enumeration underlying in the stochastic recursions, we consider an efficient application of stochastic dynamic programming. We suggest an effective state-space reduction technique taking advantage of our forecast evolution structure. Chapter 5 discusses our way of reducing the state space requirements in more detail.

With this reduced version of the model, we shall perform an extensive computational investigation directed toward exploring the interplay of decisions. Chapter 6 is devoted to a detailed description of our computational approach. We also consider a comparative alternative to the inventory model under the forecast evolution. This alternative inventory control model builds upon the results from Graves (1999), and follows time-series approach for modeling the uncertainty associated with buyers' orders.

We present and discuss the numerical results in Chapter 7. Therein, we emphasize different perspectives on the manufacturer's decision problems so as to evaluate concerns of the manufacturer and his contract buyers. We propose a menu of various commitment and flexibility arrangements among which a particular buyer may select an appropriate contract. The buyers' preferences are evaluated in terms of the relationship between the cost performance of the manufacturer versus the extent of early order commitments. We also propose a menu of various capacity levels which differ in the cost differential between the in-house production and subcontracting and the backorder-to-holding cost ratio. The manufacturer's preferences are analyzed based on the cost improvements and the service level that can be attained by a particular choice of capacity level. Finally, we demonstrate that the manufacturer benefits significantly from using the forecast evolution framework in conjunction with production and inventory planning.

Finally, Chapter 8 contains a summary of conclusions and discusses various extensions and further research topics.

This research differs significantly from the related literature in the following aspects.

(1) We introduce a general approach for integrating a probabilistic model of the changes in the committed orders with an analytical model of the production and inventory planning under multi-period QF contracts. This integrated use of refined commitment update scheme enables the manufacturer to make an enhanced production/inventory planning that is informed of how the order commitments evolve from one period to another. We model the changes in the committed orders through the modified MMFE framework. The martingale evolution model that we develop is a sophisticated one, as we model the evolution as a multiplicative process (which is more complex but in general more useful in practice) and accommodate the revision limits stipulated in QF contracts. Although there have been other studies using multiplicative martingale process, our inclusion of revision limits into MMFE is novel. It is important to note that the associated evolution model has a prior estimation stage

(i.e., an MMFE fitting process for some historical data) ignored in our work.

(2) We consider QF contracts in a more general problem environment and study interfaces between information sharing and physical flow coordination at the operational level. We facilitate the determination of QF contract terms and conditions through an integrated analysis. The environment is characterized by capacitated manufacturer with an option of subcontracting, multi-period contract horizon, multiple contract buyers with varying service requirements, non-stationary stochastic buyer orders with commitment updates, correlations of commitment updates across buyers and through time, backlogs and positive replenishment lead-time. This problem scope permits an expanded analysis for the operational aspects of QF contracts. Its features allow us to shed some light on the supply chain flexibility issues that are becoming more important in improving responsiveness in today's supply chain environment.

(3) We introduce a finite Markov chain approximation to the martingale forecast evolution process having some revision limits. The problem of estimating the transition probabilities of this Markov chain is addressed by a general optimization model maximizing the goodness-of-fit to observations. The estimation process also imposes some regularity constraints to accommodate the revision limits and the correlations of commitment updates across buyers and through time. The whole process facilitates the probability modeling of the stochastic framework for our dynamic decision model under QF contracts. It essentially provides (i) a novel approach to discretization in stochastic dynamic programming, and (ii) an estimation scheme for non-stationary stochastic dynamics.

(4) Using this finite stochastic framework, we characterize the structure of optimal replenishment policy of the manufacturer under QF contracts. It is found to be a staircase, state-dependent order-up-to policy, as the manufacturer has two sources of supply and the order commitments from contract buyers are an explicit component of the system state. Our commitment evolution structure allows us to develop effective computational properties to be used in finding optimal order-up-to policies.

(5) We implement computational dynamic programming as a solution technique for our recurring stochastic decision model. This appears worthwhile as the computational considerations are often excluded in similar contexts. Or at least; the stochastic, multi-period and multi-dimensional nature of the problem is found to defy an optimization-based approach, and a shift to simulation-based research is most likely

observed. We also introduce an efficient approach for reducing the associated state-space dimensionality building upon our forecast evolution structure. The approach unifies the states into commitment clusters to facilitate stochastic dynamic recursions. States partitioned into the same commitment cluster yield identical replenishment decision. This makes the computation associated with the recurrence relations much less demanding. The corresponding state space can be searched more efficiently after the due merger of the relevant states.

(6) Finally, as an effort to benchmark, we make a comparison of the evolution-based inventory management with an ARIMA-based approach. The evolution-based inventory management that we suggest incorporates our martingale evolution model of correlated order commitments under QF contracts. The ARIMA-based approach, on the other hand, follows time-series approach for modeling the uncertainty associated with buyers' orders, and builds upon the results from Graves (1999). The inventory policy under the commitment evolution results in lower order-up-to positions and lower expected total costs.

CHAPTER 2

PROBLEM ENVIRONMENT

This chapter provides a detailed description of the problem environment that will be studied in this research. We first describe the supply chain structure and discuss the underlying assumptions. Then, an overview of the forecasting activities and the production planning and inventory management operations are given.

We consider a decentralized supply chain that consists of a single manufacturer and multiple buyers, as depicted in Figure 2.1. The manufacturer produces and sells a single item to the buyers, who in turn serve end markets with stochastic market demands. In this context, we study the interactions between the manufacturer and the buyers from the perspective of aggregate production/inventory planning.

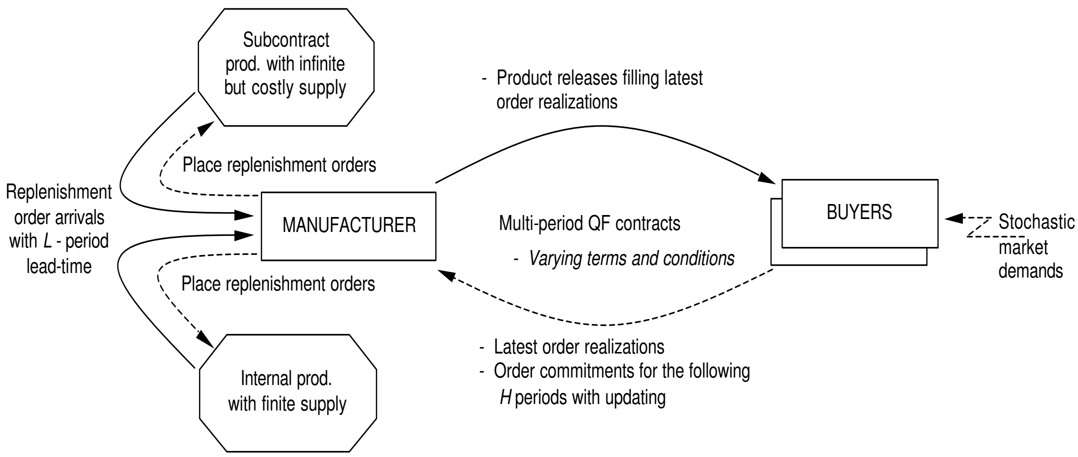


Figure 2.1: A schematic representation of the supply chain

The buyers are the immediate customers of the manufacturer. We assume that each buyer represents a different sales channel the manufacturer operates to service end market. Accordingly, they have different logistical and service requirements of the product, differentiating one buyer from another. The market demands observed by the buyers are not being directly pushed upstream to the manufacturer. Instead, the buyers undertake some demand forecasting activity, and place orders to the manufacturer one at a time. We do not require that the manufacturer has a complete knowledge of the buyers' forecasting machinery and order policy. The buyers only provide advance order information (AOI) on orders to be submitted on a range of future periods. They are allowed to be updated based on the observed market demands and in relation to the previous order patterns.

The AOI serves as a valuable input as it may assist the manufacturer in making better production and inventory decisions (in terms of production quantities fluctuating less from one period to another, for example). The effectiveness depends on the flexibility in periodical changes in the order information. Evolving order information, however, do not necessarily become more steady as they are updated successively; i.e., it may tend to vary often and widely. The manufacturer that relies on such AOI struggles to keep up with a basic problem when managing his operations. The problem is of *demand information distortion*, whereby much of the market demand variability is passed along to the manufacturer, especially when the buyers' updating the order information is made in an unrestricted way. As a result, operations at the manufacturer level do not avoid the familiar bullwhip effect, inevitably exacerbating the manufacturer's capacity adaptation and production/inventory planning.

The manufacturer gets into *contractual agreements* in mitigating this problem. He would like to implement those contracts that entice the buyers to commit their orders in advance. On the other hand, the buyers prefer the contracts that would allow them to adjust their orders when necessary. This is usually translated into *commitment* and *flexibility* stipulations in the contractual agreements. The commitment with flexibility ensures a reliable business volume to be committed by the buyers while providing them with an opportunity of adjusting to changes in market conditions. The buyers would like to have greater flexibility, allowing them to satisfy the uncertain market demand at a lower cost. The manufacturer, on the other hand, demands a lower flexibility with the aim of attaining smoother production schedules at high capacity utilization.

In particular, we study quantity flexibility (QF) contracts where the AOI is in the form of periodical order commitments on future order realizations, which will evolve eventually into realized orders one at a time. The *QF contract operation* can be described briefly as follows. We have two main stages in which the events takes place. (i) In the first stage, the relatively long-term decisions are made. QF contract terms and conditions are determined by the manufacturer, offered to the buyers, and agreed upon. The manufacturer may offer separate menu of multi-period QF contracts to each buyer. Contract terms and conditions are negotiated, and the rules that cover the actual implementation of the contract such as product pricing, order approval, and shipping are agreed upon. The negotiable parameters primarily include the length of the horizon over which the contracts will be valid, the length of commitment horizon for which AOI are available to the manufacturer, and flexibility limits that are allowed for the buyers' updating. In turn, the manufacturer builds his in-house capacity given the agreed-upon contracts and his belief about market condition. And, the buyers develop their forecasting machinery. (ii) Next, given the decisions made in the first stage, the agreed-upon contracts are executed. The buyers have limited downward and upward flexibility for updating the series of order commitments in every period. The manufacturer guarantees to satisfy every realized order up to a certain percentage above and below the its last commitment. So, the contract buyers are constrained by the series of their previous order commitments, and hence they are unlikely to arrive at unreasonable order amounts after a series of updates. The manufacturer makes replenishment ordering decisions in the face of these stochastic commitment updates in every period. But, we do not attempt to make an inventory planning of the buyers. Figure 2.2 helps to clarify the QF contract operation.

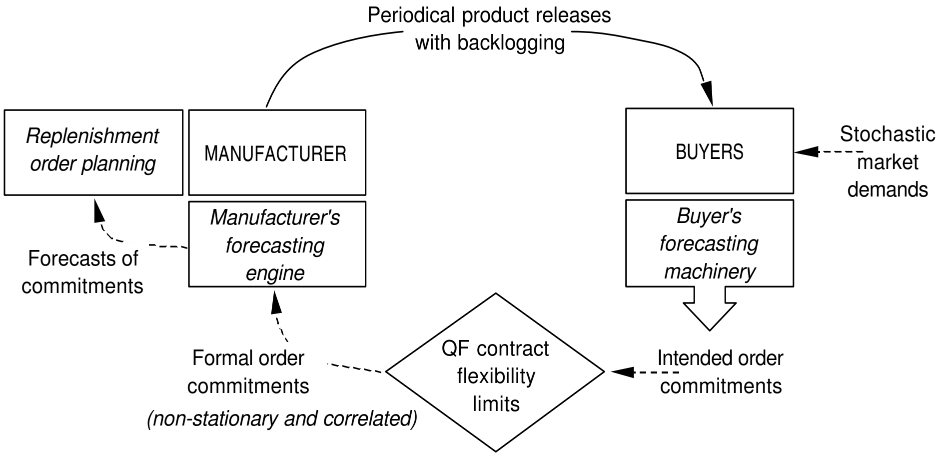


Figure 2.2: Information sharing and physical flow coordination under QF contracts

We assume that the buyers under QF contracts first determine their *intended order commitments* as their future self plans. These intended order commitments, however, are not transferred directly to the manufacturer [cf. Fig. 2.2]. If there were no contract flexibility limits then the intended order commitments would be directly transferred to the manufacturer. A commitment update decision intended in a particular period becomes the formal update decision of that period if it is already within the contract flexibility limits. The intended order commitments may hit one of the limits as well. In such a case, the limit value is submitted to the manufacturer.

It is important to see where the *non-stationarity* comes from in this setting. We leave market demands to be freely defined as we do not attempt to model the demand process confronted by the buyers. Rather, we use the order information (formal order commitments and eventual order realizations) submitted by the buyers and their evolution from one period to another. These are assumed to be non-stationary (by their very nature of periodic updates). It is the buyer's forecasting machinery that processes observations on the market demand and generates these advanced order information. The market demands however may come from a stable demand process.

The *buyers' forecasting machinery* generating these non-stationary order information is assumed to be a general one. It is based on several forecasting techniques including not only statistical forecasts but also expert judgment. In most industrial contexts especially with dynamic and volatile operating environments, market demand is often difficult to forecast based only on historical observations. Judgmental forecasting then becomes effective in the sense that it may reflect knowledge of events that have not been observed in the past but are expected in the future (e.g., planned price changes, trade promotions, product advertisements, etc.), and knowledge of recent events whose effects have not yet been observed in time series data. It can be expected, on the other hand, that the judgmental forecasting is much more prone to prediction errors and the underlying forecast volatility, characterized by the size and frequency of forecast errors, turns out to be high. This arises the need for gradual forecasting (evolving forecasts in time), and thus necessitates modeling a forecast updating process getting tuned over time as additional information is available. This point is also important to justify the presence of correlated commitments.

The order information submitted to the manufacturer are assumed to be correlated across buyers and through time periods. The following correlations are expected to occur. (*i*) Lower (resp., higher) than expected market demand in the current period may result in downward (resp., upward) modification to the order commitments for the subsequent period's order realizations. This exemplifies positively correlated order information from one period to the next. For example, a buyer may think that a price promotion will spread the word among her customers period after period. (*ii*) The buyer may be confident about the total market demand to be observed over a number of periods in the future, but is uncertain about the exact amounts to be observed in each period. So, if the market demand in the current period turns out to be lower (resp., higher) than expected, then the order commitments for the following periods' order realizations are modified upwards (resp., downwards). This illustrates negatively correlated order information through time periods. (*iii*) Sales support activities (e.g., promotions) lead to temporary demand lifts for the promoted buyer while depressing demand for the other buyers. This represents negatively correlated order information across buyers in the same period.

The manufacturer operates a *periodic-review inventory system* in the face of stochastic commitment updates. The manufacturing strategy is that products are processed and delivered into stock according to a forecast-driven production planning, and consequently buyers' order realizations are filled from the finished-goods inventory. The replenishment lead-time is L periods. Inventory is replenished in anticipation of buyer orders beginning L periods in advance of the fulfillment. Figure 2.3 illustrates the net inventory level and the inventory position (i.e., all outstanding replenishment orders less all realized buyer orders) over time from period s to $s+L+3$. At the beginning of period s, the replenishment order from L periods ago, q_{s-L}, has just been received [①]. The manufacturer then reviews the current inventory position and the latest order information from the buyers [②]. The inventory position in period s is restored by a replenishment order of q_s [③] to be received at the beginning of period $s+L$ [④]. Over the period s, both the inventory position and the net inventory level decrease as buyers' orders are realized, and the item goes into back-order [⑤] before the next replenishment order (q_{s-L+1}) is received. Notice that the amount ordered varies by period and the inventory is not restored to the same level after each replenishment. Although a replenishment order is placed in $s+1$ [⑥], no replenishment order is placed in period $s+2$ [⑦]. Hence the on-hand inventory receives no replenishment for period $s+L+2$ [⑧].

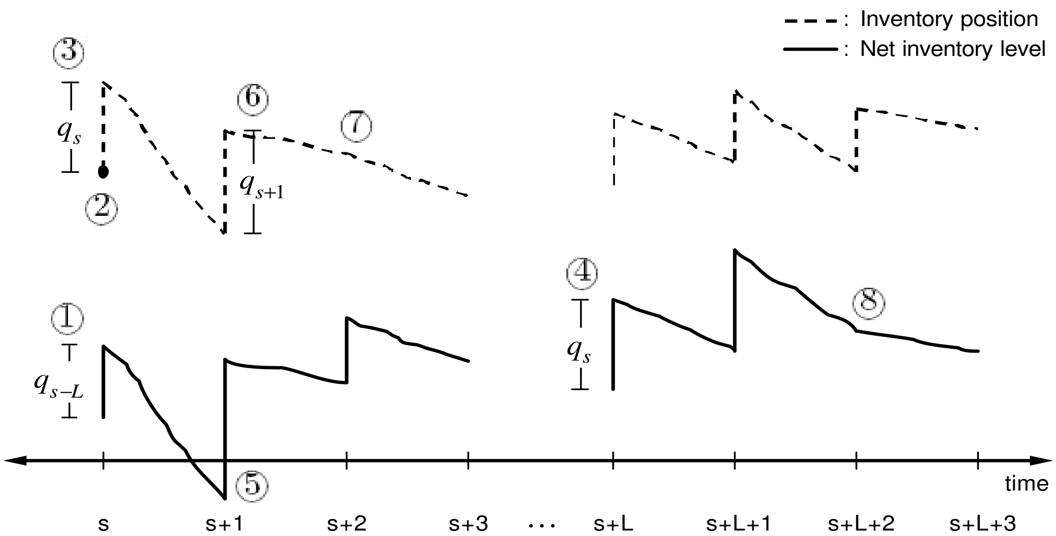


Figure 2.3: An illustration of inventory position and net inventory level over time

In this context, the manufacturer is concerned with forecasting buyers' orders to be realized on a range of future periods. He has a collection of historical order commitments and realized orders that acquired over a sufficiently long period of time. He models the time series of these order information as a *multiplicative forecast evolution process* through the martingale model of forecast evolution (MMFE) technique due to Heath and Jackson (1994). The evolution model accommodates flexibility ranges stipulated in the agreed-upon QF contracts. It is intended to come up with a better explanation of the buyers' ordering behavior in the light of the inherent correlation structure imbedded in the buyer's machinery. We should also note that the manufacturer's forecasting does not attempt improving accuracy of the order commitments. Rather, it is a technique for modeling the results of the buyers' ordering practices and takes into account the evolution of order information.

The manufacturer uses this martingale forecast evolution model as an integral part of his production and inventory planning. Martingale is a very basic yet general stochastic process with convenient properties for inventory modeling. The direction of a martingale process is completely unpredictable, and the only relevant information for forecasting the future is today's observation. For the particular construction we assume, a multiplicative version of the evolution model is expected to perform better because of the percent flexibility ranges of QF contracts and the non-stationarity of buyers' order commitments. Additive models would not be appropriate since commitment updates are scale dependent such that an update of 10% being independent of the commitment size is more reasonable. Putting another way, the size of arithmetic difference between successive order commitments depends on the size of the commitments. Multiplicative model describes commitment evolution mechanism by a multiplication operation so that it describes the commitment update made in period s as the ratio of successive order commitments generated in periods s and $s-1$ for a future order amount. Hausman (1969) and Hausman and Peterson (1972) exemplify the related literature on multiplicative forecasts. Hausman (1969) treats ratios of successive forecasts as independent lognormal variates for a recurring sequential decision problem with new improved forecasts before each decision stage. He illustrates the approach via a dynamic programming formulation with the current forecast being the state variable. Hausman and Peterson (1972) consider a stochastic forecast modification process for a capacitated, multi-item production scheduling problem. They assume that ratios of successive forecasts of total orders for a seasonal product are independent lognormal variates.

The manufacturer can replenish his inventory from two sources of supply; namely, the in-house production as well as an outside supplier. The replenishment cost does not include any fixed cost of ordering, rather the manufacturer incurs variable costs of the replenishment. This is mainly because ordering is rarely ignored, and within a replenishment order taking place in every period, fixed costs can be assumed to be included in the item cost. There exist some restrictions on the maximum amount that can be produced internally. The manufacturer somehow has an idea of the business volume that the buyers intend to develop in the future, and has built a certain level of in-house production capacity. The outside supplier is a subcontractor to whom the manufacturer is able to subcontract part of his business for capacity reasons. That is, the manufacturer is capable of producing the product, but he does not have all the production capacity required to produce all the order amount expected to realize. The subcontract replenishment amounts arrive only after a lead-time of L periods as well. Subcontracting arrangements are more costly than internal production but offers an infinite-supply opportunity. Subcontracting allows for short term capacity adjustments in the face of temporal demand variations, and hence improves the supply chain agility. It enables the capacitated manufacturer to make use of smoothing his releases when serving different buyers.

We do not assume any explicitly-set safety stock levels; instead, the safety margin is attained by the trade-off between the costs of inventory carrying and stock-outs. We assume that the manufacturer does not suffer lost sales when stock-outs occur so that the inventory level represents either on-hand inventory or backorders. The manufacturer has a tradeoff between utilizing his capacity in full, ordering large amounts to the subcontractor, hence carry inventory; versus restricting his supply flexibilities and experiencing occasional or usual stock-outs. This arises the question of how the supply-side flexibility through either maintaining excess capacity or resorts to subcontracting help to match supply with demand more effectively.

The manufacturer does not keep track of separate inventory pools for each buyer. It is assumed that he backlogs any order which is not satisfied immediately from his finished-goods inventory. He incurs a penalty for any shortfall of his delivery from the buyers' orders. Although the manufacturer does not make any inventory rationing (that is, he does not hold back any inventory for future periods), he differentiates between the buyers in order fulfillment process, allocating the finished-goods inventory by respecting a given precedence relationship among the buyers. Accordingly, total requirement of the highest-priority buyer is simply filled first. Then the next highest-priority buyer is satisfied, and so on. The corresponding shipments are assumed to arrive at the buyers immediately.

This problem environment has some similarities and dissimilarities to the environments of the past work in the literature. We mention the most closely related ones here. Heath and Jackson (1994) contribute for the development of the MMFE technique as a demand model and they provide motivation and detailed discussions. They introduce an MMFE for a multi-item production/distribution system with correlated demands across products and time periods. They adapt the MMFE to generate forecast modifications (in place of an existing time series-based forecasting procedure) and study the impact of forecast error on cost and customer service. They conduct a simulation study, and analyze the relationship between safety-stock levels and improvement in forecasts.

Toktay and Wein (2001) consider an MMFE demand process for discrete-time make-to-stock queues. They characterize effective policies under heavy-traffic assumptions for a capacitated single server. They define the planning horizon to be the effective period for which one utilizes the forecast information. They explore the impacts of dynamically evolving forecasts, demand correlation, and capacity utilization. They demonstrate that earlier information is more valuable under high capacity utilization levels, and the marginal value decreases with capacity utilization.

Kaminsky and Swaminathan (2004) model a forecasting process getting refined over time as new information becomes available. Differently from our research, they represent forecasts by a series of bands and introduce forecast-band evolution model based on these bands. Succeeding forecasts have a smaller band and are contained within the band defined by previous forecasts as time passes. They adapt the procedure to capacitated production planning environment and develop heuristics which utilize knowledge of demand forecast evolution. They perform an extensive computational study in which they employ simulation to explore the efficiency of heuristics under different settings. This is aimed at understanding the effect of forecast updates, seasonal fluctuation, and firm capacity.

Kayhan et al. (2005) introduce a general approach for integrating forecast evolutions and production inventory planning. They conduct a project for a food products company engaged in high volume production with a capital intensive operation with a highly skilled workforce. The operating environment is a multi-item, make-to-stock system with correlated stochastic demands without any formal contractual agreements, where production is set on a rolling horizon basis to restore inventory to a planned level. First, they test and model the time series of forecasts and demands as a multiplicative forecast evolution process through the MMFE. Therefore, they intend to capture interactions through time and across items, and come up with a better explanation of demand and forecast variability. Second, they integrate the forecast evolution model into the production-scheduling module of the company's enterprise resource planning system, through enhanced rules for determining safety stocks by taking the variance-covariance matrix of the forecast evolution model into account. They test the suggested framework by retrospectively simulating the company's production and inventory environment for 24 successive rolling horizons, each consisting of 12 planning periods. Their tests suggest that the model that they propose can bring considerable improvement in the inventory investment without a significant compromise in the realized service level.

Gallego and Özer (2001) consider a model of advance demand information for a single-stage periodic-review inventory system. They characterize the form of optimal replenishment policies, where the state of the inventory system reflects the knowledge of advance demand information. There are multiple customers providing different demand lead times, but they do not include the effect of limited capacity. They show the optimality of a state-dependent (s, S) and base-stock policies for systems with and without fixed ordering costs, respectively. The demand model that they use can be viewed as a special case of the MMFE. Their computational study demonstrates that advance demand information can lead to important cost reductions under the optimal replenishment policy.

Tsay et al. (1999) provide an extensive survey of model-based research on the supply chain contracting, and present a classification scheme defined by contract clauses (quantity flexibility is one of these clauses). They study how the design of contracts affects supply chain behavior and performance. Tsay and Lovejoy (1999) provide a detailed analysis of the QF contract in a complicated setting. They have multiple locations, multiple demand periods, lead times and demand forecast updates with a rolling production planning horizon; but excludes capacity restrictions and option of supplementary capacity. They propose a framework for performance analysis and design of QF supply chains. They provide insights as to where to position flexibility for the greatest benefit, and how much to pay for it, in particular by analyzing the buyer's willingness to pay for flexibility. They perform an extensive computational study, and evaluate the impact of demand variance and flexibility parameters on system performance. They demonstrate that the presence of flexibility can diminish the transmission of the variability up to the chain.

Sethi and Zhang (2004) develop a model to analyze a QF contract involving multiple periods, rolling horizon demand, one demand forecast update in each period and a spot market. The contract permits an initial order at the beginning of a period, a forecast revision in the middle of the period, and further purchases on contract and in the spot market before the demand is realized at the end of the period. The amount that can be purchased on contract is bounded by a given flexibility limit. They discuss the impact of the forecast quality and the level of flexibility on the optimal decisions.

Feng et al. (2006) present a periodic review inventory model with multiple delivery modes. They investigate why the base-stock policy is or is not optimal in different situations. They demonstrate that the optimality of a base-stock policy is closely related to the structure of the cost-to-go function. This is because the cost-to-go function for each period is separable in the inventory positions after ordering (but only for the first two delivery modes, and higher order delivery modes may not have a base-stock structure).

Graves (1999) considers an adaptive base-stock policy for a single-item inventory system with deterministic lead-time but subject to a stochastic non-stationary demand process. Similar to our research, he shows that the demand process for the upstream stage is not only non-stationary but also more variable than that for the downstream stage. He makes analytical analysis of the bullwhip effect for a single-stage and a multi-stage case. The demand model is an integrated moving average model of order (0, 1, 1). He uses an exponential smoothing procedure to estimate the periodic demand. The problem assumes complete information since given the current period demand, the distribution of the next-period demand is fully known through the assumed ARIMA model. He characterizes the results of his assumed policy while not asserting its optimality.

CHAPTER 3

STOCHASTIC FRAMEWORK FOR COMMITTED ORDERS

In this chapter, we present the stochastic framework for the manufacturer's decision problems that will be analyzed in this research. First, we model the time series of the buyers' order commitments and realized orders as a multiplicative forecast evolution process in §3.1. This is intended to provide an enhanced variability representation capturing the buyers' forecasting behaviors. Following Heath and Jackson (1994), we adopt the martingale model of forecast evolution (MMFE) methodology as a specific evolution modeling. Differently from previous studies, we develop a modified MMFE as it accommodates revision limits stipulated in the QF contracts. The inclusion of revision limits further complicates the probabilistic framework as an analytical expression for probability distribution do not apply, as discussed in §3.2. In the second part of this chapter, in §3.3 we attempt to resolve this complication by introducing a finite Markovian representation. The resulting Markov chain facilitates the numerical estimation of probability function of cumulative commitment updates to be made for future periods. In §3.4 we then introduce a general optimization model to suggest the transition probabilities of the Markov chain maximizing the goodness-of-fit to observations under certain regularity constraints. Finally, an example computational process is given for a specific industry case with the real data in §3.5.

3.1 Modeling Probabilistic Evolution of Committed Orders

In this section, we describe the probabilistic model that explains the evolution of the order commitments submitted to the manufacturer from one period to the next. Buyers commit themselves for the orders over a given number of periods. They let their earlier commitments evolve as new information becomes available over time. In order to adequately capture the dynamic nature of the order commitments and the underlying forecasting behavior of the buyers, the manufacturer describes this periodical modification activity by a probabilistic evolution model. The manufacturer forecasts buyer's commitments (which eventually convert to the realized order for the immediate period) using this probabilistic evolution model. The manufacturer's planning operations rely on these forecasts of commitments.

A good evolution model should have certain features for an adequate treatment of the buyers' ordering practices. (*i*) First, the order commitments that are estimated by the manufacturer's assumed evolution model should reflect the historical patterns in order commitments submitted by the buyer. (*ii*) Second, the forecast volatility, measured by the size and frequency of the commitment updates, should be captured. Order commitments do not necessarily become more accurate as they are successively updated from one period to the next. The forecast volatility may cause inefficiencies if one pays no attention to how the order commitments evolve. (*iii*) Finally, a good commitment evolution model should be able to accommodate the correlation structure inherent in the historical order commitments and reproduce them in the generated data. The environment that we study assumes correlations among the order commitments from a number of buyers as they cover a number of future periods.

We assume that *order commitments* on future realizations are available for a number of periods. In each period these commitments are updated on a rolling horizon basis. The order commitments for a particular period evolve eventually into the *realized order* in that period. The realized order is the last updated order commitment. Let $b \in \{1, 2, \ldots, B\}$ be index on different buyers, H be the length of the commitment horizon. Let $\mathbf{d_s} = [\, \mathbf{d_s^b},\ \forall\, b \in \{1, 2, \ldots, B\}\,]$ denote the vector of order commitments submitted to the manufacturer at the beginning of period s, $s = 1, 2, \ldots, N$ where

$$\mathbf{d_s^b} = [d^b_{s-1,s-1},\, d^b_{s-1,s},\, \ldots,\, d^b_{s-1,s+H-1},\, \mu_{D^b},\, \mu_{D^b},\, \ldots], \tag{3.1}$$

where $d^b_{s-1,s-1}$ is the realized order from buyer b in period $s-1$ (i.e., past period's realized order), and $d^b_{s-1,s+k-1}$ is the flexible order commitment for the amount to be ordered in period $s+k-1$, $k = 1, 2, \ldots, H$. We use the boldface letters for vector notations and denote $\mathbf{d^b_s}$ by $\mathbf{D^b_s}$ when it is unknown (i.e., $\mathbf{d^b_s}$ is a realization of random variable $\mathbf{D^b_s}$). Note that the mean realized order, μ_{D^b}, implies an implicit early order information for the periods beyond the commitment horizon H. Figure 3.1 helps to clarify the evolution from period $s-1$ to s.

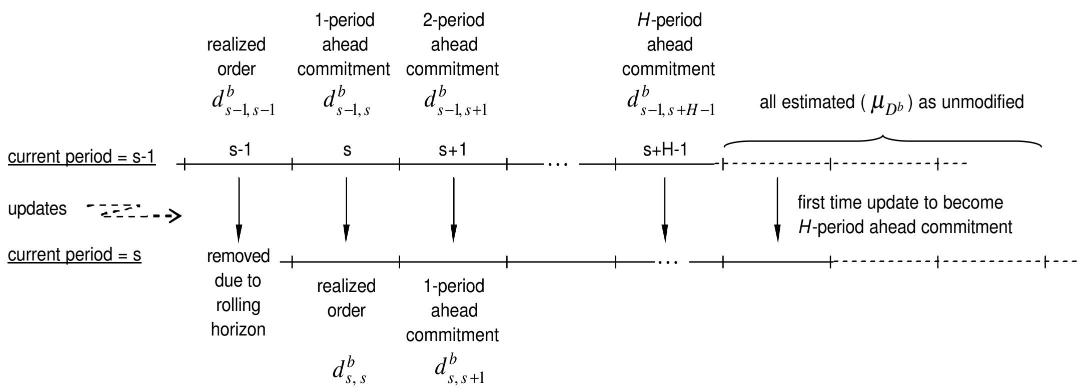


Figure 3.1: Commitment evolution from period $s-1$ to s

At the beginning of every period, an order realization is observed and all order commitments are updated with a new element getting appended to the end of the vector. From period s to period $s+1$ (before the commitment arrives) in particular, $\mathbf{d^b_s}$ will be replaced by $\mathbf{D^b_{s+1}}$ as the best prediction for the future realized orders. Commitments are updated in every period until they finally become a realized order. How these evolutions are modeled, however, depends on assumptions about whether or not the magnitude of the updates depends on the magnitude of the order commitments. For the particular problem environment that we study, the *multiplicative* evolution model is expected to perform better due to the non-stationary nature of order information and the revision limits defined in percentage of order commitments. The multiplicative model has some advantages over additive in representing the variation. These will be mentioned in §3.1.1.

3.1.1 Multiplicative commitment evolution model using the MMFE

Following Heath and Jackson (1994), we assume that the manufacturer models the time series of order commitments as a martingale model of forecast evolution (MMFE). The MMFE technique, among other approaches, has preferable properties for quite general forecasting environments, where forecasts are based on several forecasting techniques including not only analytical methods but also expert judgment. The MMFE technique imposes a certain structure to the evolution of forecasts, being consistent with the realities of the most real-life business cases. The assumptions that are required for the existence of a martingale forecast evolution model are simple however quite reasonable to assume. Hausman (1969), Graves et al. (1986) and Heath and Jackson (1994) are the main contributory studies for the development of the MMFE technique as a demand model and provide detailed discussion and motivation of the technique.

The MMFE technique follows two specific ways to describe how the available order commitments $\mathbf{d_s^b}$ in (3.1) evolve into the future: the additive model and the multiplicative model. We focus on the multiplicative evolution models, which are more complex but in general expected to be more useful in practice. The additive process would assume that the variance of commitment updates added to the mean realized order will remain the same regardless of the value of the mean realized order. The multiplicative process, on the other hand, assumes that the size of arithmetic differences between successive order commitments made for the same period might be related to the size of the commitments. For example, an update of 10% being independent of the commitment size is more reasonable in our problem. Thus, the variance of commitment updates becomes proportional to the mean realized order, resulting in a coefficient of variation not changing with a decrease/increase in the mean realized order. In general, the multiplicative models are expected to be more useful in practice (Hausman 1969). A multiplicative MMFE model can be converted to additive by a logarithmic transformation of the order commitments and realized orders, where all of the realized orders and hence also the order commitments are assumed to be strictly positive. Let $\varepsilon_{s,k}^b$ be the random variable denoting the multiplicative commitment update effective in period s for the amount to be committed for order in period $s+k-1$, $k = 1, 2, \ldots, H+1$. It is received at the end of period s, and given by

$$\varepsilon^{b}_{s,k} = \begin{cases} \ln D^{b}_{s,s-1+k} - \ln d^{b}_{s-1,s-1+k}, & \text{for } k = 1, 2, \ldots, H, \\ \ln D^{b}_{s,s-1+k} - \ln \mu_{D^{b}}, & \text{for } k = H+1. \end{cases} \tag{3.2}$$

Thus, let $\boldsymbol{\mathcal{E}_s} = [\boldsymbol{\mathcal{E}^b_s},\ \forall\ b \in \{1, 2, \ldots, B\}]$ be the random vector of multiplicative updates to be received at the end of period $s = 1, 2, \ldots, N$, where

$$\boldsymbol{\mathcal{E}^b_s} = \left[\varepsilon^{b}_{s,1}, \varepsilon^{b}_{s,2}, \ldots, \varepsilon^{b}_{s,H+1}, 0, 0, \ldots\right]. \tag{3.3}$$

We should underline that the random variable $\varepsilon^{b}_{s,H+1}$ corresponds to first time update made to μ_{D^b} to become H-period ahead commitment and $\varepsilon^{b}_{s,1}$ is the last update to be made to the most immediate commitment. This means there are $H+1$ updates in total. These specify the form of the evolution from one period to the next. In particular, (3.3) models the updates to the vector of order commitments going forward from period s to period $s+1$. The multiplicative model in this case defines the changes in the order commitments as the differences of logarithms. Alternatively, this multiplicative MMFE model can be represented by using $\mathbf{R_s} = [\, e^{\boldsymbol{\mathcal{E}^b_s}},\ \forall\ b \in \{1, 2, \ldots, B\}]$. The exponential is taken componentwise

$$\mathbf{R^b_s} = e^{\boldsymbol{\mathcal{E}^b_s}} = [R^{b}_{s,1}, R^{b}_{s,2}, \ldots, R^{b}_{s,H+1}, 1, 1, \ldots] \tag{3.4}$$

Then, a component of the vector $\mathbf{R_s}$ represents the ratio of order commitments for each period at or after period s submitted in two successive periods ($s-1$ and s).

As in Heath and Jackson (1994), the multiplicative evolution model is governed by the following structural assumptions.

Assumption 3.1 : The information available to make predictions in any period increases as time passes.

The rationale for this assumption is straightforward. Forecasting systems use available information to learn the true state space of the environment and make predictions for future random phenomena. Thus, this type of forecasting process can be modeled using expectations conditional on information sets. Formally, what is known at time s is represented in the form of an increasing family $\mathcal{F}_s \subset \mathcal{F}$ of σ-fields, generated by all observed events up to time s (like order commitments and realized orders). Hence, $\mathcal{F}_s$ is the smallest σ-field with respect to which the order vectors $\mathbf{D_1}$, $\mathbf{D_2}$, $\ldots$, $\mathbf{D_s}$ are measurable, and $\mathcal{F}_s \subseteq \mathcal{F}_{s+1}$.

Assumption 3.2 : The update vector $\boldsymbol{\mathcal{E}_s} = [\,\boldsymbol{\mathcal{E}_s^b},\ \forall\, b \in \{1,\, 2,\, \ldots,\, B\}\,]$ is uncorrelated with the information in set $\mathcal{F}_s$, and hence is uncorrelated with all combinations of vectors $\boldsymbol{\mathcal{E}_1},\, \boldsymbol{\mathcal{E}_2},\, \ldots,\, \boldsymbol{\mathcal{E}_{s-1}}$. Also, $\mathrm{E}[\boldsymbol{\mathcal{E}_s}] = \mathbf{0}$.

The reason for making this assumption is to specify an evolution model in which commitment updates are not predictable using the historical data. This rationale is meaningful, since if this assumption does not hold (i.e., when there exists some combination of the available information which is correlated with $\boldsymbol{\mathcal{E}_s}$), then this would result in some other nontrivial value of $\boldsymbol{\mathcal{E}_s}$, which is better than the existing one. Consequently, the existing order commitments can be improved with this new better value of $\boldsymbol{\mathcal{E}_s}$.

One important way of justifying this assumption is closely connected to the theory of martingales. Unfortunately, martingale theory requires some basic knowledge of abstract measure theory, and a formal treatment is thus outside the scope of this research (see Williams 1991 for a treatment of measure theory). An informal description of the martingale concept follows. If it is the case that the prediction of a future order commitment is its conditional expectation given the available information today, then $\{D^b_{s,t},\ s \leqslant t\}$ for a given t and buyer b; that is, successive order commitments for a particular future period, form a martingale.

Definition 3.1 : *A stochastic process $\{D^b_{s,t},\ s \leqslant t\}$ for a given t and buyer b is called a $\mathcal{F}_s$-martingale if the following conditions hold.*

(i) For any s, $D^b_{s,t}$ is adapted to $\mathcal{F}_s$.

(ii) For all s, $\mathrm{E}[\,|\,D^b_{s,t}\,|\,] < \infty$.

(iii) For all s and t with $s \leqslant s' \leqslant t$, $\mathrm{E}[D^b_{s',t}\,|\,\mathcal{F}_s] = D^b_{s,t}$.

The conditional expectation plays a central role in this informal definition. The first condition says that $D^b_{s,t}$ is $\mathcal{F}_s$-measurable for each s (the observed value $d^b_{s,t}$ is contained in $\mathcal{F}_s$), and the second condition is just a technical condition for the random variable $D^b_{s,t}$ being integrable. The important condition is the third one, which says that a martingale has an unpredictable direction (i.e., no systematic trend). Then, for a given value of order commitments submitted by the buyer b, $\mathbf{d^b_s}$, we have

$$\begin{aligned}\mathrm{E}[D^b_{s,s+k}\,|\,\mathcal{F}_s] &= \mathrm{E}[\mathrm{E}[D^b_{s+k,s+k}\,|\,\mathcal{F}_{s+1}]\,|\,\mathcal{F}_s] \\ &= \mathrm{E}[D^b_{s+k,s+k}\,|\,\mathcal{F}_s] \\ &= d^b_{s-1,s-1+k},\end{aligned}$$

for $k = 1, 2, \ldots, H$. This follows from the law of iterated expectations and applying statement (*iii*) of Definition 3.1. Thus, it can be shown, using properties of conditional expectations, that the values of any martingale difference series must be uncorrelated, and must have the mean equal to 0. Taking expectations of (3.2) conditioned on the information set $\mathcal{F}_s$ results in

$$\begin{aligned}
\mathrm{E}[\varepsilon^b_{s,k} \,|\, \mathcal{F}_s] &= \begin{cases} \mathrm{E}[\ln D^b_{s,s-1+k} \,|\, \mathcal{F}_s] - \mathrm{E}[\ln d^b_{s-1,s-1+k} \,|\, \mathcal{F}_s], & \text{for } k = 1, 2, \ldots, H, \\ \mathrm{E}[\ln D^b_{s,s-1+k} \,|\, \mathcal{F}_s] - \mathrm{E}[\ln \mu_{D^b} \,|\, \mathcal{F}_s], & \text{for } k = H+1 \end{cases} \\
&= \begin{cases} \ln d^b_{s-1,s-1+k} - \ln d^b_{s-1,s-1+k}, & \text{for } k = 1, 2, \ldots, H, \\ \ln \mu_{D^b} - \ln \mu_{D^b}, & \text{for } k = H+1 \end{cases} \\
&= 0,
\end{aligned}$$

so that $\mathrm{E}[\varepsilon^b_{s,k}] = 0$, indicating that the sequence of the logarithmic changes in order commitments forms a martingale difference.

Assumption 3.3 : The vector sequence $\{\boldsymbol{\mathcal{E}_s},\ s \geqslant 1\}$ form a stationary stochastic process.

The rationale for this assumption relates to predictability over time. If the $\boldsymbol{\mathcal{E}_s} = [\,\boldsymbol{\mathcal{E}^b_s},\ \forall\, b \in \{1, 2, \ldots, B\}\,]$ vectors form a non-stationary process then the degree of predictability would become relative to the time period considered and it would then become possible for some $t,\ t \geqslant s$ that $\mathrm{E}[\boldsymbol{\mathcal{E}_s} \,|\, \mathcal{F}_s] \neq \mathbf{0}$ while $\mathrm{E}[\boldsymbol{\mathcal{E}_s} \,|\, \mathcal{F}_s] = \mathbf{0}$ for $s > t$. Therefore, if the vector process $\boldsymbol{\mathcal{E}_s}$ for all s was non-stationary then the distributional parameters of the commitment updates would not be sufficient to capture all important characteristics of the evolution model. Hence $\boldsymbol{\mathcal{E}_s}$ has identical properties for all s.

Assumption 3.4 : The update vectors $\boldsymbol{\mathcal{E}_s} = \left[\,\boldsymbol{\mathcal{E}^b_s},\ \forall\, b \in \{1, 2, \ldots, B\}\,\right]$ at all s are multivariate Normal random vectors.

With these assumptions, the multiplicative form of the MMFE technique produces a multiplicative evolution model in which commitment update vectors through time (i.e., $\boldsymbol{\mathcal{E}_s}$ vectors for different s values) form a stationary and independent sequence of multivariate Normal random vectors with mean $\mathbf{0}$. If the MMFE assumptions hold for the stochastic process $\{\boldsymbol{\mathcal{E}^b_s},\ s \geqslant 1\}$ then it follows from the properties of Lognormal distribution (see Law and Kelton 2000) that the vector sequence $\{\mathbf{R^b_s},\ s \geqslant 1\}$ satisfies the MMFE assumptions as well. In this case, $\mathbf{R_s}$ vectors for different s values form a stationary and independent sequence of multivariate Lognormal random vectors with mean one.

Let $f_{\boldsymbol{\mathcal{E}_s}}(\cdot)$ be the *jointly* continuous probability density function of $\boldsymbol{\mathcal{E}_s}$, and its cumulative distribution function is denoted by $F_{\boldsymbol{\mathcal{E}_s}}(\cdot)$, which is a $B(H+1)$-fold integral. We define $F_{\varepsilon^b_k}(\cdot)$ to be the *marginal* distribution of $\varepsilon^b_{s,k}$, and $f_{\varepsilon^b_k}(\cdot)$ its marginal density. We should note that $F_{\varepsilon^b_k}(\cdot) = F_{\varepsilon^b_{s,k}}(\cdot)$ for all s, k since the vector sequence $\{\boldsymbol{\mathcal{E}_s},\ s \geqslant 1\}$ form a stationary stochastic process due to Assumption 3.3 given above. The expected value of the kth component of $\boldsymbol{\mathcal{E}^b_s}$ is denoted by $\mu_{\varepsilon^b_k} = \mathrm{E}[\varepsilon^b_{s,k}]$ such that $\mu_{\boldsymbol{\mathcal{E}^b}} = [\,\mu_{\varepsilon^b_k} = 0,\ \forall\ k \in \{1,\, 2,\, \ldots,\, H+1\}\,]$ is the mean vector of $\boldsymbol{\mathcal{E}^b_s}$. The collection of these subvectors $\mu_{\boldsymbol{\mathcal{E}^b}}$ completely determines the partitioned mean vector $\mu_{\boldsymbol{\mathcal{E}}}$ of commitment update vector $\boldsymbol{\mathcal{E}_s}$

$$\mu_{\boldsymbol{\mathcal{E}}} = [\,\mu_{\boldsymbol{\mathcal{E}^1}} \,|\, \mu_{\boldsymbol{\mathcal{E}^2}} \,|\, \ldots \,|\, \mu_{\boldsymbol{\mathcal{E}^B}}\,]\,. \tag{3.5}$$

The variance of the kth component of $\boldsymbol{\mathcal{E}^b_s}$ and the covariance between the kth component of $\boldsymbol{\mathcal{E}^b_s}$ and the lth component of $\boldsymbol{\mathcal{E}^r_s}$ are denoted by $\sigma^2_{\varepsilon^b_k} = \mathrm{Var}(\varepsilon^b_{s,k})$ and $\sigma_{\varepsilon^b_k,\varepsilon^r_l} = \mathrm{Cov}(\varepsilon^b_{s,k},\, \varepsilon^r_{s,l})$, respectively. The variance-covariance matrix $\Sigma_{\boldsymbol{\mathcal{E}^b}}$ of $\boldsymbol{\mathcal{E}^b_s}$ and the diagonal covariance matrix $\Sigma_{\boldsymbol{\mathcal{E}^b},\boldsymbol{\mathcal{E}^r}}$ between $\boldsymbol{\mathcal{E}^b_s}$ and $\boldsymbol{\mathcal{E}^r_s}$ are

$$\Sigma_{\boldsymbol{\mathcal{E}^b}} = \begin{bmatrix} \sigma^2_{\varepsilon^1_1} & \sigma_{\varepsilon^1_1,\varepsilon^1_2} & \cdots & \sigma_{\varepsilon^1_1,\varepsilon^1_{H+1}} \\ \sigma_{\varepsilon^1_2,\varepsilon^1_1} & \sigma^2_{\varepsilon^1_2} & \cdots & \sigma_{\varepsilon^1_2,\varepsilon^1_{H+1}} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{\varepsilon^1_{H+1},\varepsilon^1_1} & \sigma_{\varepsilon^1_{H+1},\varepsilon^1_2} & \cdots & \sigma^2_{\varepsilon^1_{H+1}} \end{bmatrix} \qquad \Sigma_{\boldsymbol{\mathcal{E}^b},\boldsymbol{\mathcal{E}^r}} = \begin{bmatrix} \sigma^2_{\varepsilon^b_1,\varepsilon^r_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{\varepsilon^b_2,\varepsilon^r_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma^2_{\varepsilon^b_{H+1},\varepsilon^r_{H+1}} \end{bmatrix}.$$

Hence we have the following partitioned $B(H+1) \times B(H+1)$ variance-covariance matrix of the random vector $\boldsymbol{\mathcal{E}_s}$

$$\Sigma_{\boldsymbol{\mathcal{E}}} = \begin{bmatrix} \Sigma_{\boldsymbol{\mathcal{E}^1}} & \Sigma_{\boldsymbol{\mathcal{E}^1},\boldsymbol{\mathcal{E}^2}} & \cdots & \Sigma_{\boldsymbol{\mathcal{E}^1},\boldsymbol{\mathcal{E}^B}} \\ \Sigma_{\boldsymbol{\mathcal{E}^2},\boldsymbol{\mathcal{E}^1}} & \Sigma_{\boldsymbol{\mathcal{E}^2}} & \cdots & \Sigma_{\boldsymbol{\mathcal{E}^2},\boldsymbol{\mathcal{E}^B}} \\ \vdots & \vdots & \vdots & \vdots \\ \Sigma_{\boldsymbol{\mathcal{E}^B},\boldsymbol{\mathcal{E}^1}} & \Sigma_{\boldsymbol{\mathcal{E}^B},\boldsymbol{\mathcal{E}^2}} & \cdots & \Sigma_{\boldsymbol{\mathcal{E}^B}} \end{bmatrix}. \tag{3.6}$$

It should be noted that we allow correlations among the components of $\boldsymbol{\mathcal{E}_s}$ for a fixed period s by the multivariate normal distribution assumption; but not the components across different s values. More specifically, two types of correlation are allowed; (i) correlations among the components of $\boldsymbol{\mathcal{E}_s^b}$ for a particular buyer b, and (ii) correlations between buyers in the same k, $k = 1, 2, \ldots, H+1$. This implies that using the expression (3.2) we can specify any demand correlation structure inherent in the system across buyers and in time given all other assumptions (realized demand is eventually formed by the multiplicative updates). Consequently, all the important characteristics of the evolution model (3.2) can be captured from the variance-covariance matrix $\Sigma_{\boldsymbol{\mathcal{E}}}$ of the update vectors $\boldsymbol{\mathcal{E}_s}$.

Consequently, we require only the variance-covariance matrix $\Sigma_{\boldsymbol{\mathcal{E}}}$ of the commitment updates and the initial state of the sequence $\{\mathbf{D_s},\ s \geqslant 1\}$ ($\mathbf{D_s}$ designated as the *forecast of commitments* by the manufacturer in period s) to model the forecasting behavior of the buyers. This variance-covariance matrix will be estimated as a result of the MMFE fitting process on the time series of historical order commitments eventuating as the realized orders. It is important to note that the efficacy of the MMFE technique depends on the accuracy of the estimated variance-covariance matrix $\Sigma_{\boldsymbol{\mathcal{E}}}$ of the update vectors $\boldsymbol{\mathcal{E}_s}$. The more accurate the estimated variance-covariance matrix $\Sigma_{\boldsymbol{\mathcal{E}}}$, the more reliable the computational work that will be made in this research.

3.1.2 Incorporating revision limits into the multiplicative MMFE model

We now turn to the consideration of certain revision limits that restrict the evolution of order commitments from one period to the next in the multiplicative evolution model. Although the revision limits could be put to use in a variety of ways, we chose, in this research, to exercise them via a supply contract with quantity flexibility. Attention is then given to how the flexibility characteristics of the system under revision limits impact the evolution of order commitments.

Quantity flexibility (QF) contract is an arrangement where parties agree upon the rules that impose certain limits on the range of allowable volumes for their future businesses. In this research, we consider a multi-period setting with non-stationary buyer orders, where the manufacturer's periodical replenishment decisions occur after the receival of the buyer's order commitments but before the buyer places her realized order. We thus assume that the manufacturer has the rolling-horizon QF contracts which allow order information updates. The contract is based on setting upper and lower limits on how much flexibility the buyer has in updating $\mathbf{d_s^b}$ going forward in time. So, each buyer is constrained by her previous order commitments, and hence she is unlikely to provide unreasonable prior order commitments. As a result, the buyer has a limited downward flexibility for updating order commitments while the manufacturer guarantees to satisfy up to a certain percentage above them (see Tsay and Lovejoy 1999, Tsay 1999, Anupindi and Bassok 1998).

Let ω_k^b and α_k^b denote the lower and upper flexibility limit to the modification of k-step order commitment from buyer b. Hence buyer b is entitled to the set of downward contract flexibility limits

$$\boldsymbol{\mathcal{W}^b} = \{\, \omega_k^b, \quad k = 1,\, 2,\, \ldots,\, H \,\}, \tag{3.7}$$

for $\omega_k^b \in [0,\, 1]$, and the set of upward contract flexibility limits

$$\boldsymbol{\mathcal{A}^b} = \{\, \alpha_k^b, \quad k = 1,\, 2,\, \ldots,\, H \,\}, \tag{3.8}$$

for $\alpha_k^b \geqslant 0$. Then, buyer b can not revise her k-step order commitment downward by a fraction of more than ω_k^b or upward by more than α_k^b each time. Specifically, given the latest order commitments $\mathbf{d_s^b}$, the values of the next order commitments $\mathbf{D_{s+1}^b}$ to be published from buyer b at the end of period s, $s = 1,\, 2,\, \ldots,\, N$ are known to lie within the range

$$(1-\omega_k^b)\, d_{s-1,s+k-1}^b \;\leqslant\; D_{s,s+k-1}^b \;\leqslant\; (1+\alpha_k^b)\, d_{s-1,s+k-1}^b, \tag{3.9}$$

for $k = 1,\, 2,\, \ldots,\, H$. By exercising the flexibility limits on successive commitment updates, buyer b is required to restrict her realized order quantities to be within the range

$$\prod_{j=1}^{k}(1-\omega_j^b)\, d_{s-1,s+k-1}^b \;\leqslant\; D_{s+k-1,s+k-1}^b \;\leqslant\; \prod_{j=1}^{k}(1+\alpha_j^b)\, d_{s-1,s+k-1}^b, \tag{3.10}$$

for $k = 1,\, 2,\, \ldots,\, H$. This also implies that the commitment updates for the multiplicative evolution model occur within a certain range defined by the contract flexibility limits,

$$\ln(1-\omega_k^b) \;\leqslant\; \varepsilon_{s,k}^b \;\leqslant\; \ln(1+\alpha_k^b), \tag{3.11}$$

for $k = 1,\, 2,\, \ldots,\, H$, and cumulatively,

$$\sum_{j=1}^{k}\ln(1-\omega_j^b) \;\leqslant\; \sum_{j=1}^{k}\varepsilon_{s+k-j,j}^b \;\leqslant\; \sum_{j=1}^{k}\ln(1+\alpha_j^b). \tag{3.12}$$

Consequently, the multiplicative evolution model that we discussed in the previous section needs to be modified under QF contracts.

3.1.3 Censored distributions as a way of incorporating revision limits

In order to incorporate the contract flexibility limits (3.7) and (3.8) into the multiplicative evolution model, we must first address how the statistical analysis of order commitments should incorporate flexibility limits. The question to be raised at this point is how we account for the tail probabilities of the distributions of ε_k^b's associated with these limits. We will assume that the existence of the revision limits implies *interval censored distributions* (see Greene 2000) for commitment updates with censoring points defined by those revision limits. For our purposes, a censored distribution is the part of an uncensored distribution that is above and/or below some specified censoring points with spikes. It integrates to one over the allowable range of commitment updates. Hence, a useful way to view censoring is in terms of the probability that $\varepsilon_{s,k}^b$ is less than $\ln(1-\omega_k^b)$ and larger than $\ln(1+\alpha_k^b)$, which we shall call *the degree of censoring*. This is an increasing function of $\ln(1-\omega_k^b)$ and $\ln(1+\alpha_k^b)$. As this probability increases, a greater proportion of the distribution is being transformed to censoring points. The censored Normal distribution, with $\mu = 0$ and $\sigma = 0.2$, is illustrated for $\omega_k^b = 0.3$ and $\alpha_k^b = 0.4$ in Figure 3.2.

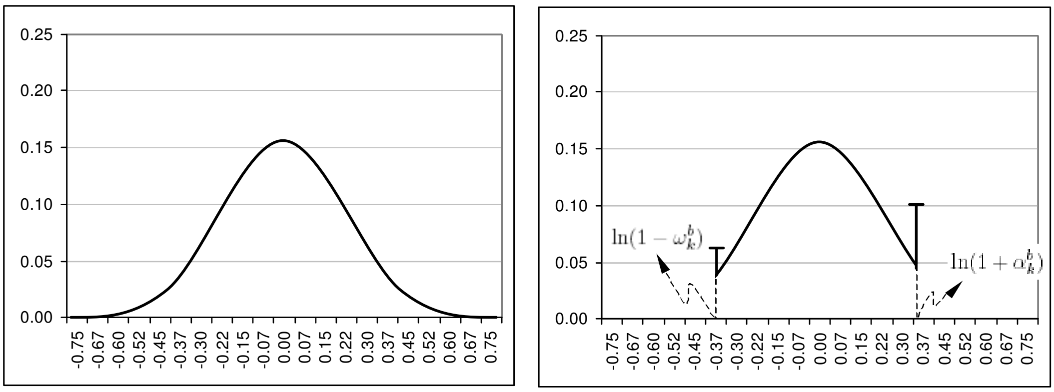


Figure 3.2: Plot of the pdf of a Normal distribution with $\mu = 0$, $\sigma = 0.2$, $\omega_k^b = 0.3$, $\alpha_k^b = 0.4$

The following assumption is required in order to achieve a meaningful interpretation of this.

Assumption 3.5 : The buyer's forecasting machinery in generating the order commitments is not influenced by the level of revision limits asked from the manufacturer (i.e., censoring does not impact forecasts to commit later. The buyer naively forecasts first, submits commitments next). So, commitment update decisions in different periods are evaluated independently, and hence previous restrictions on the values of order commitments and future expectations do not influence commitments submitted.

This assumption is crucial for the theory and methods of martingale inference, namely for the preservation of the martingale property under the existence of contract flexibility limits. That is why we have considered the problem setting where each buyer under QF contract first determines her *intended order commitments* in each period. These intended order commitments, however, are not transferred directly to the manufacturer. Rather, they can be thought as intended future self plans, being also revised on a rolling-horizon basis. The contract flexibility limits asked by the manufacturer are then applied to these self plans. If there were no contract flexibility limits then the intended order commitments would be directly transferred to the manufacturer. A commitment update decision intended in a particular period becomes the formal update decision in that period if it is already within the contract flexibility limits. The intended order commitments may hit one of the limits as well. In such a case, the limit value is submitted to the manufacturer. The difference between the transferred (formal) quantity and the intended order quantity does not influence the next period's commitment decision. Thus, we do not, for instance, accommodate the cases where a buyer compensates an intended higher order than the upper limit. Note that this is an approximation. Intended quantities usually may contain buffers and hence this assumption partially is fulfilled.

We recall from Definition 3.1 in the previous section that the conditional expectation plays a central role in the martingale property (Williams 1991). In a similar manner, the expectations under contract flexibility limits will be a function of the degree of censoring both from below and from above. If the degree of censoring from either side is equal (symmetric flexibility), the time series of successive commitment updates for a fixed k continue to be a martingale difference as we have $\mathrm{E}[D^b_{s,s+k} \,|\, \mathcal{F}_s,\, \boldsymbol{\mathcal{W}^b},\, \boldsymbol{\mathcal{A}^b}] = d^b_{s-1,s+k}$. However, if the degree of censoring from below is higher than that from above then the stochastic process becomes a sub-martingale since $d^b_{s-1,s+k} < \mathrm{E}[D^b_{s,s+k} \,|\, \mathcal{F}_s,\, \boldsymbol{\mathcal{W}^b},\, \boldsymbol{\mathcal{A}^b}]$; otherwise, it is a super-martingale since

$\mathrm{E}[D^b_{s,s+k} \,|\, \mathcal{F}_s,\, \boldsymbol{\mathcal{W}^b},\, \boldsymbol{\mathcal{A}^b}] < d^b_{s-1,s+k}$.

Definition 3.2 : *A stochastic process $\{D^b_{s,k},\ s \leqslant k\}$ for a buyer b and a given k satisfying, for all s and t with $s \geqslant t$, the inequality $\mathrm{E}[D^b_{t,k} \,|\, \mathcal{F}_s] \leqslant D^b_{s,k}$ is called a super-martingale, and satisfying the inequality $\mathrm{E}[D^b_{t,k} \,|\, \mathcal{F}_s] \geqslant D^b_{s,k}$ is called a sub-martingale, in addition to the first two conditions given in Definition 3.1.*

Consequently, with Assumption 3.5 we have ensured that the martingale property is preserved and hence the MMFE technique will be valid (for stationarity of $\{\boldsymbol{\mathcal{E}^b_s},\, s \geqslant 1\}$) under contract flexibility limits. The existence of contract flexibility limits is translated into censored distributions for commitment updates $\varepsilon^b_{s,k}$ in the multiplicative evolution model. We will henceforth specify the uncensored counterpart of a variable by a tilde ($\tilde{}$) above that variable, as need arises. That is, $\tilde{\varepsilon}^b_{s,k}$ will denote the k-step intended order commitment update of buyer b in period s. Thus, $\varepsilon^b_{s,k}$ should actually be represented by

$$\varepsilon^b_{s,k} = \begin{cases} \ln(1-\omega^b_k), & \text{for } \ \tilde{\varepsilon}^b_{s,k} < \ln(1-\omega^b_k) \\ \varepsilon^b_{s,k}, & \text{for } \ \ln(1-\omega^b_k) \leqslant \tilde{\varepsilon}^b_{s,k} \leqslant \ln(1+\alpha^b_k) \\ \ln(1+\alpha^b_k), & \text{for } \ \tilde{\varepsilon}^b_{s,k} > \ln(1+\alpha^b_k). \end{cases} \tag{3.13}$$

$\varepsilon^b_{s,k}$ has a mixture of discrete and continuous distributions. Continuous part is the original distribution; that is, $F_{\tilde{\varepsilon}^b_k}(\cdot)$. We assume in our case that any order commitment greater than the associated upward flexibility limit (resp., lower than the downward flexibility limit) are censored to the associated upper flexibility limit (resp., to the lower flexibility limit). Hence each discrete part is represented by a Bernoulli distribution with a probability equivalent to upper or lower-tail mass of the original distribution $F_{\tilde{\varepsilon}^b_k}(\cdot)$ at the associated censoring points. The order commitments within the flexibility limits, however, remain unchanged. Censoring leads to distributions conditional on the range defined by the contract flexibility limits. So, let $f_{\boldsymbol{\mathcal{E}_s}}(\cdot \,|\, \boldsymbol{\mathcal{W}^b},\, \boldsymbol{\mathcal{A}^b})$ be the *jointly* continuous probability density function of $\boldsymbol{\mathcal{E}_s}$ conditioned on the value of flexibility limits ($\boldsymbol{\mathcal{W}^b},\, \boldsymbol{\mathcal{A}^b}$), and its conditional cumulative distribution function is denoted by $F_{\boldsymbol{\mathcal{E}_s}}(\cdot \,|\, \boldsymbol{\mathcal{W}^b},\, \boldsymbol{\mathcal{A}^b})$, which is a $B(H+1)$-fold integral. We define $F_{\varepsilon^b_k}(\cdot \,|\, \omega^b_k,\, \alpha^b_k)$ to be the marginal distribution of $\varepsilon^b_{s,k}$, conditioned on the value of lower and upper flexibility limits $(\omega^b_k,\, \alpha^b_k)$, and $f_{\varepsilon^b_k}(\cdot \,|\, \omega^b_k,\, \alpha^b_k)$ its marginal density. Thus, we have the following interval censored Normal distribution that applies to $\varepsilon^b_{s,k}$, $k = 1,\, 2,\, \ldots,\, H$.

$$
\begin{aligned}
&Pr\{\,\varepsilon^b_{s,k} = \ln(1-\omega^b_k)\,\} \;=\; Pr\{\,\tilde{\varepsilon}^b_{s,k} \;<\; \ln(1-\omega^b_k)\,\} \;=\; F_{\tilde{\varepsilon}^b_k}(\ln(1-\omega^b_k)),\\
&Pr\{\,\varepsilon^b_{s,k} = \ln(1+\alpha^b_k)\,\} \;=\; Pr\{\,\tilde{\varepsilon}^b_{s,k} \;>\; \ln(1+\alpha^b_k)\,\} \;=\; 1-F_{\tilde{\varepsilon}^b_k}(\ln(1+\alpha^b_k)),\\
&\tilde{\varepsilon}^b_{s,k} \text{ has the same density with } \varepsilon^b_{s,k} \text{ for } \ln(1-\omega^b_k) \;\leqslant\; \tilde{\varepsilon}^b_{s,k} \;\leqslant\; \ln(1+\alpha^b_k).
\end{aligned}
\tag{3.14}
$$

Note that the contract flexibility limits do not apply to the furthest period where $k = H+1$ (since this is not an update over a previous commitment), and hence $\tilde{\varepsilon}^b_{s,H+1}$ has exactly the same density with $\varepsilon^b_{s,H+1}$ over its entire allowable range.

3.1.4 Parameters of the censored distributions

We now need to specify the mean and the variance-covariance matrix of the interval censored update vectors, $\boldsymbol{\mathcal{E}_s}$ [1]. Henceforth, we shall use the terms *censored mean* and *censored variance-covariance* to refer to the mean and variance-covariance in the censored distribution. With both upward and downward contract flexibility limits ($\boldsymbol{\mathcal{W}^b}$, $\boldsymbol{\mathcal{A}^b}$), each censored distributional parameter is a function of the degree of censoring both from below and from above. Following Rose (1994), the explicit solution to the first moment of an interval censored Normal distribution of $\varepsilon^b_{s,k}$ is given by

$$
\begin{aligned}
\mathrm{E}[\varepsilon^b_{s,k}\,|\,\omega^b_k,\,\alpha^b_k] \;=\;& \ln(1-\omega^b_k)\,F_{\tilde{\varepsilon}^b_k}(\ln(1-\omega^b_k)) + \ln(1+\alpha^b_k)\,[1-F_{\tilde{\varepsilon}^b_k}(\ln(1+\alpha^b_k))]\\
&+\;\mu_{\tilde{\varepsilon}^b_k}\,[F_{\tilde{\varepsilon}^b_k}(\ln(1+\alpha^b_k)) - F_{\tilde{\varepsilon}^b_k}(\ln(1-\omega^b_k))]\\
&-\;\sigma^2_{\tilde{\varepsilon}^b_k}\,[f_{\tilde{\varepsilon}^b_k}(\ln(1+\alpha^b_k)) - f_{\tilde{\varepsilon}^b_k}(\ln(1-\omega^b_k))].
\end{aligned}
\tag{3.15}
$$

We derive the variance of an interval censored Normal distribution of $\varepsilon^b_{s,k}$ to be

$$
\begin{aligned}
&\mathrm{Var}(\varepsilon^b_{s,k}\,|\,\omega^b_k,\,\alpha^b_k) \;=\\
&\quad(\ln(1-\omega^b_k))^2\,F_{\tilde{\varepsilon}^b_k}(\ln(1-\omega^b_k)) + (\ln(1+\alpha^b_k))^2\,[1-F_{\tilde{\varepsilon}^b_k}(\ln(1+\alpha^b_k))]\\
&\quad+\;(\mu^2_{\tilde{\varepsilon}^b_k}+\sigma^2_{\tilde{\varepsilon}^b_k})\,[\,F_{\tilde{\varepsilon}^b_k}(\ln(1+\alpha^b_k)) - F_{\tilde{\varepsilon}^b_k}(\ln(1-\omega^b_k))\,]\\
&\quad-\;\sigma^2_{\tilde{\varepsilon}^b_k}\,[\,(\mu_{\tilde{\varepsilon}^b_k}+\ln(1+\alpha^b_k))\,f_{\tilde{\varepsilon}^b_k}(\ln(1+\alpha^b_k))\\
&\qquad-\;(\mu_{\tilde{\varepsilon}^b_k}+\ln(1-\omega^b_k))\,f_{\tilde{\varepsilon}^b_k}(\ln(1-\omega^b_k))\,] - \mathrm{E}^2[\varepsilon^b_{s,k}\,|\,\omega^b_k,\,\alpha^b_k].
\end{aligned}
\tag{3.16}
$$

We find the following semi-explicit solution to the covariance between $\varepsilon^b_{s,k}$ and $\varepsilon^r_{s,l}$

[1] These two suffice since the only model parameters of the MMFE structure are the variance-covarince matrix for the distribution of each update vector and the initial forecast state.

$$
\begin{aligned}
&\mathrm{Cov}(\varepsilon^b_{s,k},\, \varepsilon^r_{s,l} \,|\, \omega^b_k,\, \alpha^b_k,\, \omega^r_l,\, \alpha^r_l) \;\; = \\
&\quad \ln(1-\omega^b_k) \int\limits_{-\infty}^{\ln(1-\omega^b_k)} [\int\limits_{-\infty}^{\ln(1-\omega^r_l)} \ln(1-\omega^r_l)\mathrm{d}f(\tilde{\varepsilon}^b_{s,k}, \tilde{\varepsilon}^r_{s,l}) + \int\limits_{\ln(1+\alpha^r_l)}^{\infty} \ln(1+\alpha^r_l)\mathrm{d}f(\tilde{\varepsilon}^b_{s,k}, \tilde{\varepsilon}^r_{s,l})\,] \\
&+ \;\ln(1+\alpha^b_k) \int\limits_{\ln(1+\alpha^b_k)}^{\infty} [\int\limits_{\ln(1+\alpha^r_l)}^{\infty} \ln(1+\alpha^r_l)\mathrm{d}f(\tilde{\varepsilon}^b_{s,k}, \tilde{\varepsilon}^r_{s,l}) + \int\limits_{-\infty}^{\ln(1-\omega^r_l)} \ln(1-\omega^r_l)\mathrm{d}f(\tilde{\varepsilon}^b_{s,k}, \tilde{\varepsilon}^r_{s,l})\,] \\
&+ \int\limits_{\ln(1-\omega^b_k)}^{\ln(1+\alpha^b_k)} \int\limits_{\ln(1-\omega^r_l)}^{\ln(1+\alpha^r_l)} (\mu_{\tilde{\varepsilon}^b_k} + \sigma_{\tilde{\varepsilon}^b_k}\tilde{\varepsilon}^b_{s,k})(\mu_{\tilde{\varepsilon}^r_l} + \sigma_{\tilde{\varepsilon}^r_l}\tilde{\varepsilon}^r_{s,l})\,\mathrm{d}f(\tilde{\varepsilon}^b_{s,k}, \tilde{\varepsilon}^r_{s,l}) \\
&- \;\mathrm{E}[\varepsilon^b_{s,k} \,|\, \omega^b_k,\, \alpha^b_k]\; \mathrm{E}[\varepsilon^r_{s,l} \,|\, \omega^r_l,\, \alpha^r_l]. \qquad (3.17)
\end{aligned}
$$

A numerical valuation may help to see how the censoring affects the distributional parameters. Consider a simple 1-buyer case where $k = s+1$ and $l = s+2$. Suppose we have $\mathrm{E}[\tilde{\varepsilon}^1_{s,s+1}] = \mathrm{E}[\tilde{\varepsilon}^1_{s,s+2}] = 0$, $\mathrm{Var}(\tilde{\varepsilon}^1_{s,s+1}) = \mathrm{Var}(\tilde{\varepsilon}^1_{s,s+2}) = 4$, and $\mathrm{Cov}(\tilde{\varepsilon}^1_{s,s+1},\, \tilde{\varepsilon}^1_{s,s+2}) = 3.2$ corresponding to correlation coefficient of 0.8. When the contract flexibility limits are taken to be $\omega^1_1 = \alpha^1_1 = 0.2$ and $\omega^1_2 = \alpha^1_2 = 0.4$ (i.e., symmetric flexibility up and down), the censored distributional parameters become $\mathrm{E}[\varepsilon^1_{s,s+1} \,|\, \omega^1_1,\, \alpha^1_1] = -0.0188$, $\mathrm{E}[\varepsilon^1_{s,s+2} \,|\, \omega^1_2,\, \alpha^1_2] = -0.0726$, $\mathrm{Var}(\varepsilon^1_{s,s+1} \,|\, \omega^1_1,\, \alpha^1_1) = 0.6791$, $\mathrm{Var}(\varepsilon^1_{s,s+2} \,|\, \omega^1_2,\, \alpha^1_2) = 1.446$, and $\mathrm{Cov}(\varepsilon^1_{s,s+1},\, \varepsilon^1_{s,s+2} \,|\; \omega^1_1,\, \alpha^1_1,\, \omega^1_2,\, \alpha^1_2) = 0.0479$. For each update variable, the degree of censoring from above is greater than that of from below because of the logarithmic transformation [cf. Eq. 3.13]. Hence, censored mean turns out to be smaller than the mean of the original one. This is also caused by the existence of positive correlation of 0.8 between $\tilde{\varepsilon}^1_{s,s+1}$ and $\tilde{\varepsilon}^1_{s,s+2}$. The censoring of $\tilde{\varepsilon}^1_{s,s+1}$ pushes the distribution of $\tilde{\varepsilon}^1_{s,s+2}$ to the left. In general, the censored mean is pushed in the opposite direction of the correlation if the degree of censoring from above is greater than that of from below, and in the opposite direction if otherwise. Finally, censoring reduces the variance compared with the variance in the uncensored distribution.

Consequently, we have characterized the implications of the contract flexibility limits ($\boldsymbol{\mathcal{W}^b}$, $\boldsymbol{\mathcal{A}^b}$) on the multiplicative MMFE model. We represented the censored mean and the censored variance in terms of the first two moments for the distribution of intended commitment updates and the flexibility limits ($\boldsymbol{\mathcal{W}^b}$, $\boldsymbol{\mathcal{A}^b}$). These parameters will be used as regularity measures in probability modeling of the stochastic framework under QF contracts. The following sections will discuss how we calculate the joint probabilities of unknown distributional form.

3.2 Probability Modeling of the Stochastic Framework under QF Contracts

Having modeled the probabilistic evolution of order commitments under QF contracts, we now need to determine the probabilities of possible values that the order commitments and realized orders may assume in the future. These will be necessary to calculate performance and risk estimates with satisfactory accuracy for the problem at hand.

The evolution model describes how the available order commitments $d^b_{s,s+k}$ evolve as new information becomes available in time. This implies a conditional relationship as a necessary result. This conditional relationship says that the manufacturer can forecast any order quantity to be received in the future in terms of its latest estimate and successive random commitment updates (to be made for that particular quantity). More specifically, given the latest order commitment $d^b_{s,s+k}$, order realization to be received from buyer b in period $s+k$, $D^b_{s+k,s+k}$, can be forecasted by the multiplicative evolution model as

$$D^b_{s+k,s+k} \quad = \quad d^b_{s,s+k}\; e^{(\varepsilon^b_{s+1,k}+\varepsilon^b_{s+2,k-1}+\,\cdots\,+\varepsilon^b_{s+k,1})}, \tag{3.18}$$

for $k = 1, 2, \ldots, H$. In (3.18) we call the sum of k successive random commitment updates, given in the term $e^{(\cdot)}$, the *cumulative commitment update* over k periods. Note that H represents the length of the commitment horizon for which nontrivial order information are available to the manufacturer. On the other hand, for the periods beyond the commitment horizon H the mean realized order size serves as an implicit early order information. Thus, more generally, we can characterize the order realization in period $s+k$ for $k > H$ as

$$D^b_{s+k,s+k} \quad = \quad \mu_{D^b}\; e^{(\varepsilon^b_{s+k-H,H+1}+\varepsilon^b_{s+k-H+1,H}+\,\cdots\,+\varepsilon^b_{s+k,1})}, \tag{3.19}$$

where μ_{D^b} denotes the expected value of the realized orders from buyer b. This representation involves random cumulative commitment update over the $(H+1)$ periods, independent of k. The conditional probabilities play a central role in these relationships. Possible values of the random variable $D^b_{s+k,s+k}$ are conditioned on the latest order commitment $d^b_{s,s+k}$ in (3.18) since a buyer operating under a QF contract is ultimately required to restrict her orders to be within the range defined by contract flexibility limits.

The trouble with (3.19) is to evaluate the probabilities of $D^b_{s+k,s+k}$ since it is a function of $H+1$ random variables each interval censored and jointly from a multivariate Normal distribution, namely $F_{\boldsymbol{\varepsilon}_s}(\cdot\,|\,\boldsymbol{\mathcal{W}^b},\,\boldsymbol{\mathcal{A}^b})$. Figure 3.3 helps to see the correlation structure inherent in buyer orders through time. It shows how $D^b_{s,s}$ and $D^b_{s+1,s+1}$ are correlated as they include commitment updates from the same update vector (for instance, $\varepsilon^b_{s,1}$ and $\varepsilon^b_{s,2}$ are from the vector $\boldsymbol{\mathcal{E}^b_s}$ the components of which are known to be correlated). The further out the orders from each other, the smaller the number of commitment updates from the same update vector (since update coefficients of further periods occur in less of earlier commitment terms). The commonality vanishes at lags greater than H. Note that orders are correlated across buyers as well, as commitment updates submitted from different buyers at the same period are not independent.

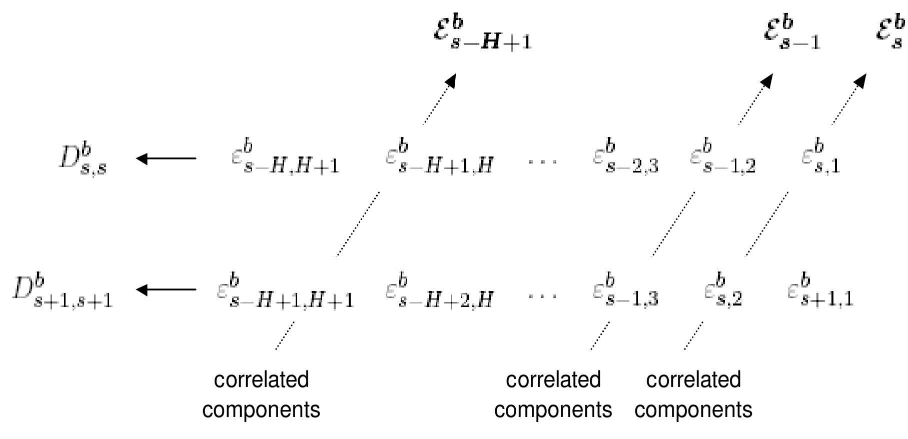


Figure 3.3: Illustration of the correlation structure in buyer orders

It is not straightforward to establish such a multivariate distribution, particularly for high dimensional problems. A true mathematical form for the multivariate distribution $F_{\boldsymbol{\mathcal{E}}_s}(\cdot\,|\,\boldsymbol{\mathcal{W}^b},\,\boldsymbol{\mathcal{A}^b})$ is complicated, since it is not clear how censored commitment updates should be included jointly in a distributional form. The main difficulty is due to the fact that the additive reproductive property of Normal distribution is not preserved under censoring, implying that a sum of censored normals does not lead to a known distributional form.

One immediate but not necessarily accurate solution would be to follow the central limit theorem for sums of random commitment updates. The central limit theorem in our case simply would suggest that the realized order $D^b_{s+k,s+k}$ has an interval censored Lognormal distribution with censoring points being equal to cumulative lower and upper limits, namely $\prod_{j=1}^{H}(1-\omega^b_j)$ and $\prod_{j=1}^{H}(1-\alpha^b_j)$. The central limit theorem has been studied intensively in probability theory. In the literature some techniques are available for establishing the accuracy of approximation in the central limit theorem (Greene 2000). They are out of the scope of this research. There, however, are several observations that should be made. The use of the central limit theorem provides a close approximation only under certain conditions in our case. (i) The first condition states that one would achieve a close approximation only if the degree of censoring for the associated distributions are small. This corresponds to relatively loose flexibility limits, which is not always the case in practice. (ii) The second condition requires that the random commitment updates, taken individually, contribute a nearly equal amount to the variance of the sum. This is closely related to how total commitment variability resolves as the system evolves from one period to the next. If most uncertainty is not resolved until a few of periods in advance, then it is likely that a single commitment update just before the period of order realization makes a large contribution to the eventual sum. Thus one would not expect to produce adequate accuracy using the Normal distribution to approximate the distribution of cumulative commitment updates.

This motivates the interest in empirical methods to get reasonably accurate approximations to the distribution of cumulative commitment updates. Rather than fitting a smooth functional distribution form, one can use an empirical distribution. This may represent a finite set of possible values that a random variable may assume in the future. The issue, however, is to develop a reasonably accurate approximation for the unknown distributional form. A good approximation in our case should accommodate three main aspects of the problem that we study. (i) The probabilities that are generated by the approximation method should reflect the existence of the contract flexibility limits. That is, it should preserve the censored nature of the random variables. (*ii*) Since underlying structures for the first and second moments are known in advance, the approximation method should suggest those probabilities which best fit this structure. (*iii*) Finally, a good approximation method should be able to accommodate the correlation structure inherent in the historical order commitments across buyers and in time. This requires sufficiently large volume of sample data. If this had not been the case, then estimating so many probabilities is bound to lead to inaccuracies, and all conclusions concerning the problem at hand would become less reliable.

In order to develop an empirical method to approximate the distribution of cumulative commitment updates we suggest a two-step approach. As a first step, we will develop a new modeling framework for cumulative commitment updates. We consider Markovian structure with discretized state-space model, which substantially helps in understanding the problem of interest. Markovian structure in general provides a well-understood conceptual framework in which many complex stochastic problems fit. The main attractiveness of Markov models, besides flexibility, lies in analytical tractability. They are simple enough to allow mathematical analysis, complex enough to adequately mirror the behavior of the underlying process. Markovian structure is not only capable of capturing the correlation structure in the evolution, but it is also scalable to allow analysis using well-known techniques.

In the second step, we will represent the problem of estimating the transition probabilities of this Markovian structure as an optimization problem. This optimization problem suggests the transition probabilities of the Markovian structure, which maximizes the goodness-of-fit to observations under certain regularity conditions. As a result, these conditional probabilities will yield the probability function of cumulative commitment updates.

3.3 Markov Chain Representation of Cumulative Update Process

In this section we take the first step in developing our empirical approach for approximating the probabilities of cumulative commitment updates, when they are restricted to be within the range defined by the contract flexibility limits. Attention is here given to the MMFE representation of realized orders as in (3.18), which allows use of random cumulative commitment updates in describing the uncertainty in the order realizations. Specifically, we are interested in the chance of realizing particular values for $(\varepsilon^b_{s+k-H,H+1} + \varepsilon^b_{s+k-H+1,H} + \cdots + \varepsilon^b_{s+k,1})$ in order to approximate the *discrete probabilities* for specific $D^b_{s+k,s+k}$ occurrences.

This modeling contribution addresses a stochastic process which accumulates random commitment updates to be made successively for a given future period. Note that the manufacturer serves B distinct contract buyers and all buyers make their commitment updates simultaneously in any period. Thus, we model B dependent commitment update processes occurring simultaneously in the environment as a multivariate stochastic process. The resulting process is called the *multivariate cumulative update process.*

Suppose we need a cumulative commitment update over k periods into the future, $k = 1, 2, \ldots, H + 1$. We observe the state of the associated multivariate cumulative update process at discrete points in time labeled $j = 1, 2, \ldots, k$. Let $(U^1_j, U^2_j, \ldots, U^B_j)$ denote the state of the process at time j [2]. More precisely, for period s taken as the present, it accumulates the random variables $\varepsilon^b_{s+1,k}$, $\varepsilon^b_{s+2,k-1}$, $\ldots$, $\varepsilon^b_{s+k,1}$ moving backward over the time interval $[s + 1, s + k]$,

$$\begin{aligned} U^b_1 &= \varepsilon^b_{s+k,1} \quad \text{and} \\ U^b_j &= \varepsilon^b_{s+k,1} + \varepsilon^b_{s+k-1,2} + \cdots + \varepsilon^b_{s+k-j+1,j} \\ &= U^b_{j-1} + \varepsilon^b_{s+k-j+1,j}, \end{aligned} \tag{3.20}$$

for $j = 2, 3, \ldots, k$. Figure 3.4 helps in understanding the accumulation of commitment updates to form U^b_k.

[2] Note that U^b_j does not depend on s, hence we drop the subscript in what follows, since the vector sequence $\{\boldsymbol{\mathcal{E}}_s,\ s \geqslant 1\}$ forms a stationary stochastic process due to Assumption 3.3. This also constitutes the reason for only one multivariate cumulative update process being sufficient to approximate the probabilities of $D^b_{s+k,s+k}$.

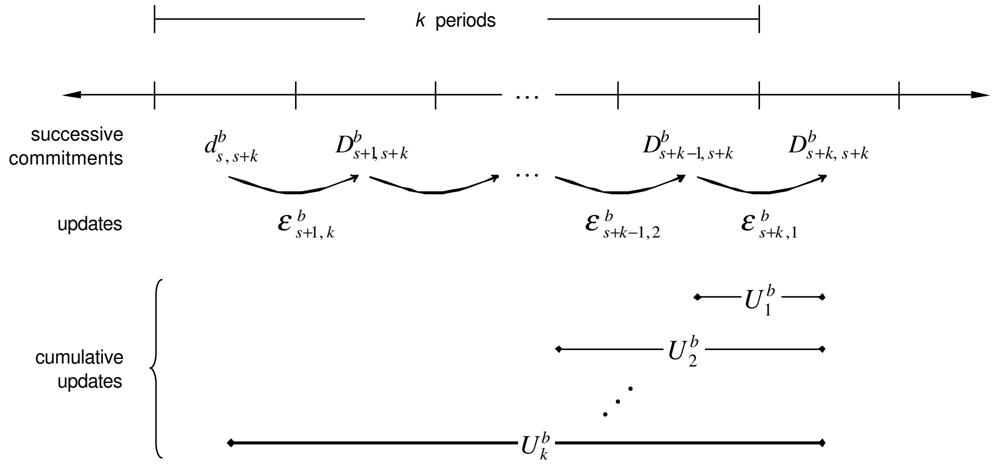


Figure 3.4: Accumulation of commitment updates over a k-period horizon

The equations (3.18) and (3.19) can now be given by $D^b_{s+k,s+k} = d^b_{s,s+k}\, e^{U^b_k}$ and $D^b_{s+k,s+k} = \mu_{D^b}\, e^{U^b_{H+1}}$, respectively. The random variable U^b_k represents the sum of logarithmic changes in successive order commitments submitted by buyer b over the following k-period time interval. The allowable range of U^b_k values, for all $k = 1, 2, \ldots, H$, is given by [3]

$$U^b_k \in \left[\sum_{j=1}^{k} \ln(1-\omega^b_j),\ \sum_{j=1}^{k} \ln(1+\alpha^b_j) \right], \tag{3.21}$$

where the interval bounds are defined by the contract flexibility limits and become looser as k moves from 1 to H.

The multivariate cumulative update process $\left\{(U^1_k, \ldots, U^B_k),\ 1 \leqslant k \leqslant H+1\right\}$ has the characteristics required by the Markovian property. A Markov process is a state-space model which allows the next progression to be determined only by the current state and not by previous states. Assumption 3.3 given in §3.1 implies that the process $\left\{(U^1_k, \ldots, U^B_k),\ 1 \leqslant k \leqslant H+1\right\}$ has the property that its future evolution is conditionally independent of its past provided that the present is known. However the process itself consists of states related with correlated transitions. The probability distribution of the states at successive epochs is dependent on the preceding steps in the Markovian sense. The multivariate cumulative update process

[3] For $k = H+1$ we do not have any flexibility limits stipulated in the QF contract since this is not an update over a previous commitment.

$\left\{(U_k^1, \ldots, U_k^B),\ 1 \leqslant k \leqslant H+1\right\}$ can be represented as a discrete-time multivariate Markov process with a joint transition probability matrix.

Every U_k^b is a continuous random variable taking values in the interval (3.21). However, the multivariate Markov process $\left\{(U_k^1, \ldots, U_k^B),\ 1 \leqslant k \leqslant H+1\right\}$ is assumed to be a discrete state-space model. Discrete state-space representation is expected to reduce the amount of data while predictive accuracy is at the analyst's discretion. The associated probability distribution functions are then defined over a finite number of possible discretized states. We will approximate every continuous random variable U_k^b, $k = 1, 2, \ldots, H+1$ by discretization. In the overall, we are interested in the process of discretizing $B(H+1)$ jointly continuous variables. Determining the method of discretization, however, involves a trade-off between speed and accuracy. We thus make univariate discretization for the sake of efficiency, whereby we discretize one continuous variable at a time, although it would be more accurate to consider multiple variables simultaneously.

Suppose this is performed by dividing the range of every U_k^b independently into a specified number of disjoint intervals. We adapt equal-width method where the continuous range of the variable is evenly divided into disjoint equal-width sub-intervals. Although equal-width method is simple and easy to implement, its accuracy would worsen as the distribution of the continuous variable deviates from the uniformity. Fortunately, since the distributions of U_k^b's are interval censored, we do not observe the adverse effects of the tail-mass or even the outliers on the resultant accuracy.

The number of disjoint sub-intervals in discretization is arbitrarily specified to be M since we usually do not know what a proper value M is. The width of these sub-intervals is a function of M and the range of the continuous variable of interest. As the discretization becomes finer; that is, as M increases, the evaluations of the discretized problem converge to those of the continuous problem, but comes with a serious computational burden. A very low M, on the other hand, may affect predictive accuracy negatively. Then, the ultimate goal is to identify a finer grid of points that provides adequate predictive accuracy not causing impractical computational requirements. Let $\aleph_k^b$ denote the set of discrete states corresponding to U_k^b, labeled $1, 2, \ldots, M$. The state space representation is complicated as the range of continuous variable U_k^b becomes wider as k increases from 1 to $H+1$. States represent continuous values in wider sub-intervals as k gets larger. Suppose $\aleph_{k,m}^b$ is the midpoint value of the sub-interval for the state m, with the exception for the first and final states which represent the lower and upper censoring points, respectively. Figure 3.5 helps in understanding our discretization.

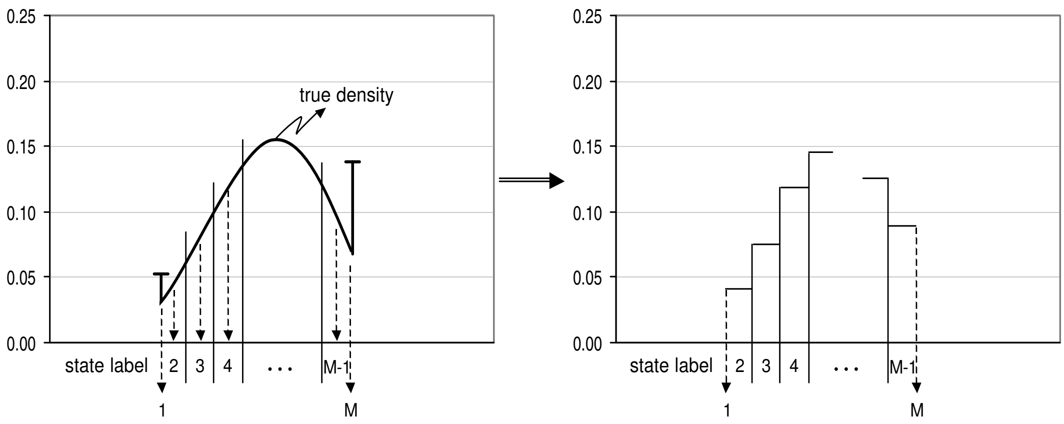


Figure 3.5: Illustration for the discretization of an interval censored random variable

Suppose that at time instant j the multivariate Markov chain $\{(U_k^1, \ldots, U_k^B), 1 \leqslant k \leqslant H+1\}$ is in some state $(m_1, \ldots, m_B)$, $m_b \in \aleph_j^b$. At the next time instant $j+1$, there occurs a transition to another state $(n_1, \ldots, n_B)$, $n_b \in \aleph_{j+1}^b$ with the probability

$$Pr\{\, U_{j+1}^1 = n_1, \ldots, U_{j+1}^B = n_B \;\mid\; U_j^1 = m_1, \ldots, U_j^B = m_B \,\},$$

which is the one-step (joint) transition probability for the multivariate Markov chain. Although the possible transitions are clear, the probability law relating the next period's state to the current state does not remain stationary over time, leading to the transition probabilities dependent on k. This is due to the following: (1) commitment update correlations through time instant $k = 1, 2, \ldots, H+1$ exist in the environment (i.e., correlated transitions among states); and (2) M discrete states in state space $\aleph_j^b$ correspond to different state values as j changes. Thus the joint transition probabilities for the multivariate Markov chain should be *time-nonhomogeneous*. More generally, we define $\mathbf{JTP_{j,k}}$ to be the $(k-j)$-step joint transition probability matrix. We can use Chapman-Kolmogorov relations to show that the $(k-j)$-step transition probabilities are obtained by matrix multiplication. When $B = 2$ one can represent $\mathbf{JTP_{j,k}}$ as follows.

$$\mathbf{JTP_{j,k}} = [\, Pr\{\, U_k^1 = n_1,\, U_k^2 = n_2 \mid U_j^1 = m_1,\, U_j^2 = m_2 \,\}\,]_{m_b \in \aleph_j^b,\, n_b \in \aleph_k^b}$$

$$= \begin{array}{c} \\ (\aleph_{j,1}^1, \aleph_{j,1}^2) \\ \vdots \\ (\aleph_{j,1}^1, \aleph_{j,M}^2) \\ (\aleph_{j,2}^1, \aleph_{j,1}^2) \\ \vdots \\ \vdots \\ (\aleph_{j,M}^1, \aleph_{j,M}^2) \end{array} \begin{array}{c} (\aleph_{k,1}^1, \aleph_{k,1}^2) \;\cdots\; (\aleph_{k,1}^1, \aleph_{k,M}^2) \;\; (\aleph_{j,2}^1, \aleph_{j,1}^2) \;\cdots\;\cdots\; (\aleph_{k,M}^1, \aleph_{k,M}^2) \\ \left[\begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \end{array} \right] \end{array}.$$

The size of the joint transition probability matrix is given by $M^B \times M^B$, in general. The model framework of the multivariate Markov chain also requires us to define $Pr\{\, U_1^1 = m_1, \ldots, U_1^B = m_B \,\}$ to be the probability that the chain is in state $(m_1, \ldots, m_B)$ at time $k = 1$. We call the $1 \times M^B$ vector $\mathbf{JP_1}$ the initial probability distribution of the multivariate Markov chain

$$\mathbf{JP_1} = [\;Pr\{\, U_1^1 = m_1, \ldots, U_1^B = m_B \,\}\;]_{m_b \in \aleph_1^b}.$$

If we know the initial distribution $\mathbf{JP_1}$ and the conditional probabilities $\mathbf{JTP_{j,k}}$ for all $j < k$, the other joint probabilities $\mathbf{JP_k}$ is computable in terms of $\mathbf{JP_1}$ and $\mathbf{JTP_{j,k}}$. This is sufficient to describe the multivariate Markov chain model of cumulative commitment updates. These can be used to derive all the important characteristics of this Markov chain. In principle, we do not need the limiting state probabilities. The target is simply an $(k-1)$-step joint transition matrix. Hence a trivial matrix multiplication yields the desired outcome. That is, we are interested in the the sequence of joint transition probability matrices $\mathbf{JTP_{1,2}}$, $\mathbf{JTP_{2,3}}$, $\ldots$, $\mathbf{JTP_{k-1,k}}$ and the final joint probabilities sought are given by

$$\begin{aligned} \mathbf{JP_k} &= \mathbf{JP_1}\, \mathbf{JTP_{1,2}}\, \mathbf{JTP_{2,3}} \ldots \mathbf{JTP_{k-1,k}} \\ &= \mathbf{JP_1}\, \mathbf{JTP_{1,k}}. \end{aligned}$$

A typical problem in the application of Markov models is the estimation of the transition probabilities. The transition matrices and final joint probabilities for our multivariate Markov chain are the unknowns. This poses the problem of accommodating the distributional characteristics of the problem, since it is not clear how information contained in the censored values of the variables should be utilized and how the underlying correlation structure (across buyers and in time) can be imposed in the estimation. In the next section we shall deal with this problem. This will be attacked by a nonlinear optimization problem that models the transition probabilities of the multivariate Markov chain to maximize the goodness-of-fit to observations under certain regularity constraints.

3.4 Estimating the Probabilities of Cumulative Update Process

Having discussed the stochastic process for the cumulative commitment updates and represented its stochastic behavior as a discrete-time multivariate Markov chain, we are now ready to estimate parameters of this Markovian representation. This amounts to determining reasonably accurate approximations to the transition probabilities as it operates under QF contracts.

A typical problem in the application of Markov models is the estimation of the transition probabilities between the states. Various methods have been proposed in the literature for dealing with this estimation problem. A common approach in the literature is to estimate the transition probabilities from a sampled data while assuming that this sample of observations follows a known probability distribution. Its performance depends on how well the presumed distributional form predicts the population. The estimation performance is often poor, particularly for high dimensional problems with limited sampling. The problem becomes even more important for our Markovian representation. This is because it is not clear how to specify an appropriate mathematical form to be assumed for the multivariate distribution, as the censored nature of distributions complicates the multivariate structure. Instead, this section will introduce an optimization procedure to estimate the transition probabilities between the states of the multivariate Markov chain $\{(U_k^1, \ldots, U_k^B),\ 1 \leqslant k \leqslant H+1\}$.

3.4.1 Optimization procedure for estimating transition probabilities

The optimization procedure that we will introduce is a general method of inference about an unknown probability distribution when there exists a prior sample estimate of the density and information on some distributional parameters. More specifically, we formulate our problem of estimating the transition probabilities between the states of the multivariate Markov chain $\{(U_k^1, \ldots, U_k^B),\ 1 \leqslant k \leqslant H+1\}$ as a problem of determining the unknown transition probabilities which deviates to a minimal degree (in the statistically acceptable sense) from the sample frequencies observed in the problem environment. In addition these will be certain problem-specific constraints derived from some other sample statistics.

The observations made in the problem environment are the periodical order commitments and the eventual realized orders from the buyers. Suppose some historical data for the order commitments, collected over a sufficiently long period of time, is made available to the manufacturer. A sample of observations on the transitions of the process $\{(U_k^1, \ldots, U_k^B),\ 1 \leqslant k \leqslant H+1\}$ can then be recovered from this data, by applying the MMFE structure. We have restricted attention to the cases where the state space of the process $\{(U_k^1, \ldots, U_k^B),\ 1 \leqslant k \leqslant H+1\}$ is discrete. Discretizing continuous ranges of commitment updates directly is critical for the decision problems that will be analyzed in this research. The reason for discretization is that our solution procedure will be enumerative. The probabilistic framework of order commitments can benefit from discretization as the amount of data is reduced through discretization without sacrificing much of the predictive accuracy. We have introduced a simple discretization already in the previous section. Therein, we divided the continuous range of every U_k^b into M categories. Each observed sample of commitment updates is then classified into one of the M categories, and the respective observed frequencies of these M categories are calculated.

We are now in a situation where we have the observed frequencies of various categories but the systematic frequencies from a distributional model yielding those categories are not known. The quantifications (decisions) to be made in our optimization procedure to estimate the transition probabilities of the multivariate Markov chain $\{(U_k^1, \ldots, U_k^B),\ 1 \leqslant k \leqslant H+1\}$ will suggest (indirectly) these unknown frequencies. We define the feasible region to be the set of unknown resulting frequencies (in M

categories) that satisfy certain regularity constraints. Thus, we must infer whether the observed frequencies of sampled data in each category differ significantly from the estimated frequencies. This requires us to pose a goodness-of-fit for any candidate feasible solution as a measure of performance.

An optimal solution will be a point in the feasible region with the largest goodness-of-fit. Consequently, there is a need to specify an appropriate measure of goodness-of-fit statistic to summarize how well the estimated frequencies of M categories in the feasible region fit the sample values observed in the problem environment. We consider a log-likelihood-ratio type statistic, namely the G-statistic, to compare the sample and estimates for the unknown model distributions (see Kendall and Stuart 1979). The G-statistic is closely related to the logarithmic-based information theory and entropy measures, and the Chi-square statistics provide adequate approximations of the log-likelihood ratio. The formula for computing the G-statistic value is

$$G = 2 \sum_{m}^{M} O_m \ln(O_m / E_m), \tag{3.22}$$

where $m = \{1, 2, \ldots, M\}$ is index on the number of discrete categories, O_m is sample frequency observed in category m, and E_m is estimate for frequency from an unknown distribution for the same category. The G-statistic will indicate the probability that the observed frequencies in the problem environment result from random sampling drawn from a distribution with the estimated frequencies. The higher the value, the lower the probability that the observed and estimated frequencies come from the same population.

In our optimization procedure, the overall problem is decomposed into smaller problems, each dealing with numerical estimation of an individual one-period joint transition matrix. This is because the conditional probabilities of making a transition to the next period's state are dependent on the current period, leading to time-nonhomogeneous conditional probabilities. Thus we will have a sequence of H optimization programs to be solved successively for $\mathbf{JTP_{1,2}}$, $\mathbf{JTP_{2,3}}$, ..., and $\mathbf{JTP_{H,H+1}}$ (each is referred to as GOF_{k-1} for a particular $k = 2, 3, \ldots, H+1$, given an initial joint probability function $\mathbf{JP_{k-1}}$). Figure 3.6 illustrates the successive nature of our optimization procedure when $B = 1$.

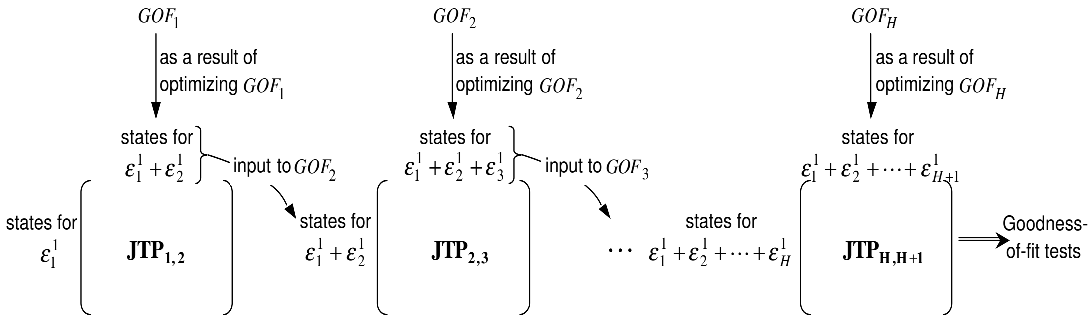


Figure 3.6: Optimization procedure

Initially we have M^B discrete states of $(\varepsilon_1^1, \ldots, \varepsilon_1^B)$ and their joint probabilities in the form of $\mathbf{JP_1}$ row vector. Each element corresponds to a combination of ε_1^b's, $b = 1, 2 \ldots, B$. We solve GOF_1 and for each choice of $(\varepsilon_1^1, \ldots, \varepsilon_1^B)$ find discrete states and probabilities of $(U_2^1, \ldots, U_2^B)$ where $U_2^b = \varepsilon_1^b + \varepsilon_2^b$, $\forall\ b$. Then we solve GOF_2 and for each choice of $(U_2^1, \ldots, U_2^B)$ (being obtained as a result of optimizing GOF_1) find discrete states and probabilities of $(U_3^1, \ldots, U_3^B)$, and so on. This procedure is repeated when changing the flexibility limits in the experimental analysis that will be mentioned in Chapter 7. The generic optimization model GOF_{k-1}, for a particular transition from $k-1$ to k, $k = 2, 3, \ldots, H+1$, can then be constructed as follows:

Optimization model GOF_{k-1}**:**

The input parameters of the model GOF_{k-1} can be defined as:

- $(\boldsymbol{\mathcal{W}^b}, \boldsymbol{\mathcal{A}^b})$, the set of lower and upper flexibility limits for each buyer b, given by $\{(\omega_k^b, \alpha_k^b),\ \ k = 1, 2, \ldots, H\}$,
- M, the number of distinct categories assumed in discretizing the continuous range of variables such that $m \in \{1, 2, \ldots, M\}$,
- $\aleph_k^b$, the set of compacted discrete states that the cumulative commitment update U_k^b may take, labeled $m \in \{1, 2, \ldots, M\}$. The set bounds are defined in terms of the contract flexibility limits $(\boldsymbol{\mathcal{W}^b}, \boldsymbol{\mathcal{A}^b})$,
- $\mathbf{JP_{k-1}}$, the joint probability mass function of $(U_{k-1}^1, \ldots, U_{k-1}^B)$ in each of M^B

categories. When $k-1=1$, it represents the initial state of the Markov chain. Otherwise, it is the output of previous optimization GOF_{k-2},

- T, the number of periods over which historical observations on time-series of order commitments $\mathbf{d_s}$ are available for use in quantifying the observed frequencies,

- $\mathbf{JP_k^{obs}}$, the vector of joint frequencies observed in each of M^B categories from a sample of size T,

- $\mu_{\boldsymbol{\mathcal{E}}}^{obs}$, the $1 \times B(H+1)$ censored mean vector of $\boldsymbol{\mathcal{E}_s}$, conditioned on the value of $(\boldsymbol{\mathcal{W}^b}, \boldsymbol{\mathcal{A}^b})$ and estimated from a sample of size T, [cf. §3.1.4]

- $\Sigma_{\boldsymbol{\mathcal{E}}}^{obs}$, the $B(H+1) \times B(H+1)$ censored variance-covariance matrix of $\boldsymbol{\mathcal{E}_s}$, conditioned on the value of $(\boldsymbol{\mathcal{W}^b}, \boldsymbol{\mathcal{A}^b})$ and estimated from a sample of size T, [cf. §3.1.4]

- $(\mathcal{T}_\mu, \mathcal{T}_\sigma)$, the set of error tolerances allowed in setting the mean and covariance values.

The decision variables of the optimization model GOF_{k-1} can be defined as:

- $\mathbf{JTP_{k-1,k}}$, the one-step joint transition probability matrix from state $(U_{k-1}^1, \ldots, U_{k-1}^B)$ to state $(U_k^1, \ldots U_k^B)$.

The resultant variables of the optimization model GOF_{k-1} are

- $\mathbf{JP_k}$, the vector for joint probability mass function of $(U_k^1, \ldots, U_k^B)$, obtained by vector-matrix product $\mathbf{JP_{k-1}JTP_{k-1,k}}$,

- $f_{U_k^b}(\cdot \mid \boldsymbol{\mathcal{W}^b}, \boldsymbol{\mathcal{A}^b})$, which is the probabilities of U_k^b in each of M categories, given in terms of $\mathbf{JP_k}$,

- $\mu_{\boldsymbol{\mathcal{E}}}$, which is the $1 \times B(H+1)$ censored mean vector of $\boldsymbol{\mathcal{E}_s}$ that is calculated in terms of model parameters and decision variables within the specified error tolerance $\mathcal{T}_\mu$,

- $\Sigma_{\boldsymbol{\mathcal{E}}}$, which is the $B(H+1) \times B(H+1)$ censored variance-covariance matrix of $\boldsymbol{\mathcal{E}_s}$ that is calculated in terms of model parameters and decision variables within the specified error tolerance $\mathcal{T}_\sigma$.

The model GOF_{k-1} (for a particular transition from $k-1$ to k, $k = 2, 3, \ldots, H+1$) with some constraints expressed in definitional form is defined as follows:

$$\text{Minimize} \quad \sum_{m_1 \in \aleph_k^1} \cdots \sum_{m_B \in \aleph_k^B} \mathbf{JP_k^{obs}}(m_1, \ldots, m_B) \; \ln\left(\frac{\mathbf{JP_k^{obs}}(m_1,\ldots,m_B)}{\mathbf{JP_k}(m_1,\ldots,m_B)}\right) \tag{3.23}$$

subject to

$$\text{row sum of } \mathbf{JTP_{k-1,k}} = 1 \qquad \forall\, b \text{ and } m_b \in \aleph_{k-1}^b \tag{3.24}$$

$$\mathbf{JP_k} = \mathbf{JP_{k-1}}\ \mathbf{JTP_{k-1,k}} \qquad \forall\, b \text{ and } m_b \in \aleph_k^b \tag{3.25}$$

$$\sum_{m_1 \in \aleph_k^1} \cdots \sum_{m_B \in \aleph_k^B} \mathbf{JP_k}(m_1, \ldots, m_B) = 1 \tag{3.26}$$

$$f_{U_k^b}(n_b \,|\, \boldsymbol{\mathcal{W}^b}, \boldsymbol{\mathcal{A}^b}) = \sum_{m_1 \in \aleph_k^1} \cdots \sum_{m_B \in \aleph_k^B} \mathbf{JP_k}(m_1, \ldots, n_b, \ldots, m_B) \qquad \forall\, b \text{ and } n_b \in \aleph_k^b \tag{3.27}$$

$$\mathbf{JTP_{k-1,k}} \text{ and } \mathbf{JP_k} \geqslant \mathbf{0} \tag{3.28}$$

$$\mu_{\varepsilon_k^b} = \sum_{m \in \aleph_k^b} m\, f_{\varepsilon_k^b}(m \,|\, \boldsymbol{\mathcal{W}^b}, \boldsymbol{\mathcal{A}^b}) \qquad \forall\, b \tag{3.29}$$

$$\sigma_{\varepsilon_k^b, \varepsilon_l^r} = \text{ a function of } \mathbf{JP_k} \qquad \forall\, b \text{ and } r \tag{3.30}$$

$$\Sigma_{\boldsymbol{\mathcal{E}}} \text{ forms a positive semi-definite matrix} \tag{3.31}$$

$$|\ \mu^{obs}_{\varepsilon_k^b} - \mu_{\varepsilon_k^b}\ | \ \leq\ \mathcal{T}_\mu \qquad \forall\, b \tag{3.32}$$

$$|\ \sigma^{obs}_{\varepsilon_k^b, \varepsilon_l^r} - \sigma_{\varepsilon_k^b, \varepsilon_l^r}\ | \ \leq\ \mathcal{T}_\sigma \qquad \forall\, b \text{ and } r \tag{3.33}$$

The objective function (3.23) minimizes the G-statistic and thus strengthens the significance of the statistical conformance. Such a model fit to observed data inevitably faces a very ample degrees of freedom. A natural question to be raised at this point is how we ensure a feasible solution in this model. The model incorporates two main leverages allowing us to ensure "tighter" feasibility: (i) the specified error tolerances used in calculating the distributional parameters and (ii) the regularity constraints imposed on the parameters. The error tolerances are small being specified as $\mathcal{T}_\mu = 0.01$ and $\mathcal{T}_\sigma = 0.001$. The regularity constraints in the numerical estimation process allow us to attain reasonable freedom of adjustment. Especially, we check whether the correlation structure inherent in the data set is appropriately modeled by any feasible solution.

Constrain set (3.24), involving M^B individual constraints, ensure that every row-wise sum of joint transition probability matrix $\mathbf{JTP_{k-1,k}}$ will be unity. M^B definitional constraints in (3.25) compute the joint probability mass function of $(U_k^1, \ldots, U_k^B)$ in terms of decision variable $\mathbf{JTP_{k-1,k}}$ and input parameter $\mathbf{JP_{k-1}}$. A single constraint (3.26) ensures that the sum of probabilities in joint probability mass function $\mathbf{JP_k}$ is made equal to 1. Each of $M^B k$ constraints (3.27) compute the marginal probability mass function of U_k^b by summing $\mathbf{JP_k}$ out the other buyers for each of M categories. Constraints (3.28) are for the nonnegativity of probabilities. B definitional constraints (3.29) compute the censored means of commitment update vectors in terms of the resultant probabilities $f_{\varepsilon_k^b}(\cdot \mid \boldsymbol{\mathcal{W}^b}, \boldsymbol{\mathcal{A}^b})$. Nonlinear constraints (3.30) compute the censored variance-covariance matrix of commitment update vectors. Nonlinear constraints (3.31) ensure the positive semi-definiteness of the computed variance-covariance matrix. Constraints (3.32) to (3.33) ensure that the absolute deviations in calculating the means and covariances are within the specified error tolerances [4].

3.4.2 Testing goodness-of-fit of the optimal solutions

We use goodness-of-fit tests to summarize how well the optimal solutions to our nonlinear optimization models fit the relevant observations. A test of goodness-of-fit establishes whether or not an observed frequency distribution differs from a postulated distribution. We consider a log-likelihood-ratio type test, namely the G-test, for testing the hypothesis

H_0 : *The observed frequencies in the problem environment result from random sampling from a distribution with the estimated frequencies obtained as a result of optimizing* GOF_{k-1}.

G-test is chosen over the more traditional Chi-square test due to several reasons [5]. First, although the Chi-square test is the most widely used of the goodness-of-fit tests which may be used with discrete data, its use with small sample size is

[4] The objective function may also minimize the maximum absolute deviations between the estimated and observed mean and covariance values.

[5] There are also other goodness-of-fit tests designed for specific discrete distributions, namely the multinomial, the discrete Kolmogorov-Smirnov, and Anderson-Darling tests (Law and Kelton 2000).

disputable. The main problem that we face in using the Chi-square test is the choice of the value M and the size of the continuous sub-intervals these M categories represent. This is because the Chi-square test will not be acceptable if the estimated frequencies are too low (i.e., when more than 10% of the categories have estimated frequencies below 5) [6]. Second, the approximation to the Chi-square distribution for the G-statistic would be better than for the Chi-square statistic in cases where the deviation of observed frequencies against estimated frequencies is greater than the value of estimated frequency in any category m. The G-test statistic is approximately Chi-square distributed, with the same number of degrees of freedom as in the corresponding Chi-square test. It has the utility of being fairly close to Chi-square distributed even at quite restricted sample sizes. In practice, they are only asymptotically equivalent. Algebraically, the Chi-square statistic is a second order Taylor approximation of the G-statistic (see Loukas and Kemp 1986 and Greene 2000).

Since the test is based on the deviation of the estimated model frequencies from the observed frequencies, it rejects the hypothesis H_0 for large values of the objective function. The objective function value is an approximate statistic value anyway, and it is the general order of magnitude one should be concerned with, and not a rigorous test to see if the statistic passes a critical threshold for a certain significance level. If the goodness-of-fit turns out to be unacceptable for any discretized data interval, we can modify the discretization specifically for that interval by creating a finer grid. This resolved model leads to an improved goodness-of-fit. One may continue in this manner up to a point when the hypothesis is not rejected and it can be concluded that the estimates are adequate.

3.5 An Example of the Computational Process

We have, so far, discussed the probabilistic framework in somewhat abstract terms. This section illustrates those ideas with a concrete example for a specific industry case with real data.

[6] A better approximation can be obtained by some corrections which avoid overestimation of statistical significance for small data.

3.5.1 Solving the nonlinear optimization model

The nonlinear constrained optimization problems, given in (3.23)-(3.33) in §3.4, was coded using the GAMS distribution 22.2 and solved using the CONOPT3 nonlinear programming algorithm, which seems to be well suited for the type of model at hand. The algorithm attempts to find a local optimum. Determining the solution algorithm is of great importance for nonlinear models. CONOPT3 is particularly well suited for models with a large degree of nonlinearity where one experiences the problem of maintaining feasibility during the optimization. It has been designed to be efficient and reliable for large and sparse models where both the number of variables and equations can be large. CONOPT3 will take advantage of the presence of definitional constraints in our nonlinear programs where many equations can be solved one by one, since CONOPT3 has a preprocessing step in which recursive equations and variables are solved and removed from the model. Some specific issues on the solution of the nonlinear models GOF_{k-1} are mentioned in Appendix B.

3.5.2 Numerical results

We use the company data provided in Kayhan et al. (2005) for a fast-moving consumer goods company. The basic setting that they studied is as follows. The operating environment is an integrated manufacturer-buyer system with capacity restrictions and correlated stochastic demands for multiple products. The buyer publishes demand forecasts for a number of future periods on a rolling horizon basis, for which the manufacturer plans its production activities. The forecasting environment is a general one involving both statistical and judgmental forecasting. The production-inventory system is governed by a set of corporate rules, which can be thought as a simple preliminary supply contract. These corporate rules require the forecasts to be updated within the agreed-upon percent revision limits. The company data is a collection of historical forecasts and demand realizations that acquired over a past $T = 125$-period horizon. They are for a particular set of products, which are sold through two main distribution channels with varying production and marketing requirements. They model the time series of forecasts and demands as a multiplicative forecast evolution process through the MMFE technique, intending to come up with a better explanation of the demand structure.

In the discussion that follows, we assume the case where $B = 2$ and $H = 2$. [7] Buyers in this setting can be thought as different products in Kayhan et al. (2005) with different distribution channels but being produced on the same production line. For this particular case, the vector of historical order commitments available at the beginning of period $s = 1, 2, \ldots, T = 125$ is represented by $\mathbf{d_s} = [\,\mathbf{d_s^b},\ \forall\, b \in \{1, 2, \}\,]$ where

$$\mathbf{d_s^b} = [d^b_{s-1,s-1},\, d^b_{s-1,s},\, d^b_{s-1,s+1},\, \mu_{D^b},\, \mu_{D^b},\, \ldots],$$

where $d^b_{s-1,s-1}$ is the realized order from buyer b in period $s-1$, and $d^b_{s-1,s+k}$ is the order commitment for the amount to be ordered in period $s+k-1$, $k = 1, 2$. The corporate rules require buyer b to update her order commitments subject to the following lower and upper percent revision limits [8]

$$(\boldsymbol{\mathcal{W}^b},\, \boldsymbol{\mathcal{A}^b}) \;=\; \{(\omega^b_k,\, \alpha^b_k),\ k = 1, 2\} \;=\; \{(0.15,\, 0.15),\, (0.30,\, 0.30)\}.$$

We first provide the MMFE fitting process for the company data and discuss its implications. The main focus here is in estimating the variance-covariance matrix of the commitment update vectors $\boldsymbol{\mathcal{E}_s}$ from this company data. This enables the manufacturer to distill historical data on orders into useful information on the important characteristics of the information process behind this forecast evolution model, namely the estimated variance-covariance matrix of commitment updates. In the MMFE fitting process we took the following steps: (1) calculate commitment update vectors; (2) eliminate forecasting biases; (3) validate the MMFE assumptions; (4) estimate the variance-covariance matrix of the commitment update vectors. Details of the steps are given briefly as follows:

Step-1: Calculate commitment update vectors

From the historical data on order commitments covering $T = 125$ periods, we calculated the sequence of commitment update vectors $\boldsymbol{\mathcal{E}_s}$ using the multiplicative evolution equation (3.2). The vector of historical commitment updates available at the end of

[7] We assume a two-period commitment horizon, for expository convenience, although the company data is available for a longer horizon length.

[8] The company data actually accommodates the case that the revision limits are violated in the case of unexpected demand conditions. So, for the purposes of our research we first modify the data to conform to censored nature of the distributions.

period $s = 1, 2, \ldots, 125$ is represented by $\boldsymbol{\mathcal{E}_s} = [\,\boldsymbol{\mathcal{E}_s^b},\ \forall\, b \in \{1,\, 2,\}\,]$ where

$$\boldsymbol{\mathcal{E}_s^b} = [\varepsilon_{s,1}^b,\, \varepsilon_{s,2}^b,\, \varepsilon_{s,3}^b,\, 1,\, 1,\, \ldots], \tag{3.34}$$

where $\varepsilon_{s,k}^b$ is the random variable denoting the multiplicative update for the amount to be ordered in period $s+k-1$, $k = 1, 2, 3$, and $\varepsilon_{s,3}^b$ is the update from μ_{D^b} to $d_{s,s+2}^b$. By analyzing these vector values we can obtain a general picture of the evolution characteristics of the underlying forecasting system.

Step-2: Eliminate forecasting biases

Estimation biases inherent in the order commitments, which may be induced by incorrect information or some inherent forecasting behavior, are investigated. If forecasts are unbiased estimates of demand then there is no systematic tendency to either underestimate or overestimate the true value of the demand. We assessed the estimation biases by statistically testing the expected value of prediction error. We observed that 2-step ahead order commitments appear to be generally almost 7% higher than realized orders. This upward bias, hence, was eliminated by adjusting all the historical order commitment values by the scaling value of (100/107) before being used in further calculations. Analysis regarding forecasting bias reflects the predictive performance of the forecasting system considered. Analyzing the systematic errors and making adjustments to obtain unbiased order commitments is critical for the purposes of estimating a variance-covariance matrix of $\boldsymbol{\mathcal{E}_s}$ that is based only on random fluctuations. Subsequently, the censored mean of $\varepsilon_{s,k}^b$ for every b, k, conditioned on the value of $(\boldsymbol{\mathcal{W}^b},\, \boldsymbol{\mathcal{A}^b})$, turns out to be

$$\boldsymbol{\mu_{\mathcal{E}}} \;=\; \begin{bmatrix} \mu_{\varepsilon_k^1},\ \forall k \in \{1,\,2,\,3\} \\ \mu_{\varepsilon_k^2},\ \forall k \in \{1,\,2,\,3\} \end{bmatrix} = \begin{bmatrix} -0.10032 & -0.10218 & -0.11247 \\ -0.15006 & -0.15887 & -0.16580 \end{bmatrix}.$$

Step-3: Validate the MMFE assumptions

The MMFE technique under revision limits produces a model where the commitment update vectors $\boldsymbol{\mathcal{E}_s}$ through time form independent, identically distributed multivariate censored Normal vectors with mean zero. This requires $\{\boldsymbol{\mathcal{E}_s},\ s \geqslant 1\}$ to form a stationary stochastic process.

When data are censored, the standard distributional tests do not apply to testing normality. To check the normality assumption we perform modified Kolmogorov-

Smirnov (K-S) tests with the significance level of 0.05 (Barr and Davidson 1973). The K-S tests indicated that in all but one case the data was most likely consistent with the normality assumption. The degree of violation for that case was not too significant (p-values for that case is 0.048) and hence they were assumed to be normal. The independence assumption in fitting an MMFE is made to specify a model where forecast updates are not predictable using the past data. To test the independence assumption we calculated autocorrelation and partial autocorrelation functions for commitment update variates. The correlation analysis indicated that the independence assumption was not violated.

Step-4: Estimate the variance-covariance matrix of $\boldsymbol{\mathcal{E}_s}$ vectors

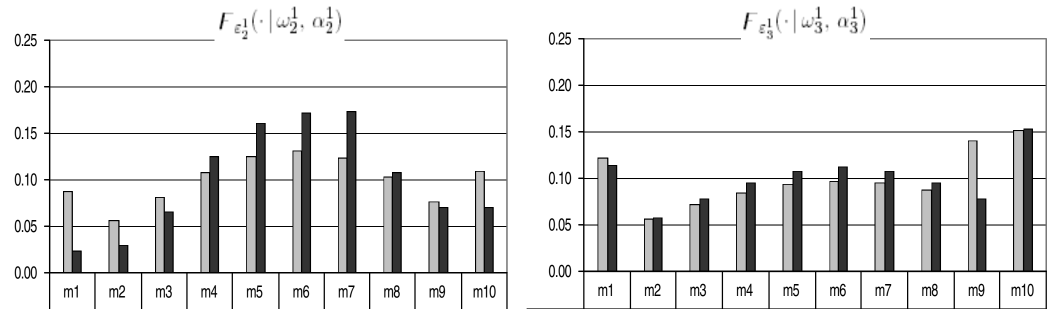
The censored variance-covariance matrix $\Sigma_{\boldsymbol{\mathcal{E}}}$ for the distribution of commitment updates $\boldsymbol{\mathcal{E}_s}$ was estimated from the historical data conditioned on the value of ($\boldsymbol{\mathcal{W}^b}$, $\boldsymbol{\mathcal{A}^b}$) using moment estimators as

$$\Sigma_{\boldsymbol{\mathcal{E}}} \quad = \quad \begin{array}{c} \\ \varepsilon_1^1 \\ \varepsilon_2^1 \\ \varepsilon_3^1 \\ \varepsilon_1^2 \\ \varepsilon_2^2 \\ \varepsilon_3^2 \end{array} \begin{array}{c} \begin{array}{cccccc} \varepsilon_1^1 & \varepsilon_2^1 & \varepsilon_3^1 & \varepsilon_1^2 & \varepsilon_2^2 & \varepsilon_3^2 \end{array} \\ \left[\begin{array}{cccccc} 0.30011 & 0.00103 & -0.00859 & 0.19246 & 0.00012 & 0.00023 \\ & 0.00121 & -0.00009 & 0.00032 & 0.00140 & 0.00019 \\ & & 0.03028 & 0.00009 & 0.00016 & 0.00832 \\ & & & 0.20064 & 0.00254 & 0.00017 \\ & & & & 0.00372 & -0.00189 \\ & & & & & 0.02057 \end{array} \right] \end{array},$$

Dimensionality of the matrix depends on the number of buyers considered, B, and the length of the commitment horizon used, H. Since the number of historical commitment update vectors available is 125 ($= T$), we have a 125×6 data matrix. As a result, we have a $6 \times 6 (= 2 \times 3)$ variance-covariance matrix of the 125×6 data matrix. Note that with more historical data one can estimate the variance-covariance matrix more accurately. When the dimensionality of the variance-covariance matrix is large compared to the sample size of the data, it will result in fewer degrees of freedom. We should note that using the variances for all $k \in \{1, 2, 3\}$ for a particular buyer b the percentages of total order commitment variability that is resolved as the system evolves from one period to the next period are calculated. As can be observed from the variance-covariance matrix, a significant proportion of total order commitment variability is not resolved until the period of realization. In other words, there exists a considerable amount of prediction error in the system. This observation reflects the performance of the forecasting system (or the accuracy of the order commitments provided by the buyers) and indicates that the manufacturer will not be able to effectively respond to demand variability without holding significant amount of stock.

The finite Markovian representation of the cumulative update process $\{(U_k^1, U_k^2),\ 1 \leqslant k \leqslant 2\}$ involves two joint transition probability matrices $\mathbf{JTP_{1,2}}$ and $\mathbf{JTP_{2,3}}$ and hence joint probability mass functions $\mathbf{JP_2}$ and $\mathbf{JP_3}$ are to be estimated. We assumed to discretize the continuous range of every commitment update into $M = 10$ distinct categories. In the resulting optimization procedure, we have two nonlinear programs to be solved successively for the two transitions. Given an initial joint probability mass function $\mathbf{JP_1}$ derived from the company data, we first solved the nonlinear model GOF_1 for the transition probabilities $\mathbf{JTP_{1,2}}$ and the joint probability mass function $\mathbf{JP_2}$. Using these probabilities, we then solved the second nonlinear model GOF_2, which is from time instant $k = 2$ to $k = 3$ for determining the transition probabilities $\mathbf{JTP_{2,3}}$ and the joint probability mass function $\mathbf{JP_3}$.

The marginal probability mass functions $F_{\varepsilon_2^1}(\cdot\,|\,\omega_2^1,\, \alpha_2^1)$ and $F_{\varepsilon_3^1}(\cdot\,|\,\omega_3^1,\, \alpha_3^1)$ that were determined as a result of optimization corresponding to commitment updates ε_2^1 and ε_3^1 from buyer $b = 1$, respectively, are shown in Figure 3.7. These probabilities are labeled by *model* in the figure. Figure also includes the frequencies that were observed in the problem environment (labeled by *obs*), and found by applying the MMFE technique to the company data.



	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10
model	0.08744	0.05554	0.08162	0.10709	0.12551	0.13135	0.12278	0.10247	0.07640	0.10982
obs	0.02268	0.02988	0.06614	0.12559	0.16158	0.17147	0.17407	0.10763	0.07083	0.07013

	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10
model	0.12162	0.05684	0.07135	0.08416	0.09335	0.09734	0.09537	0.08787	0.14100	0.15111
obs	0.11433	0.05782	0.07738	0.09532	0.10802	0.11268	0.10815	0.09554	0.07766	0.15303

Figure 3.7: Goodness-of-fit of $F_{\varepsilon_2^1}(\cdot\,|\,\omega_2^1,\, \alpha_2^1)$ and $F_{\varepsilon_3^1}(\cdot\,|\,\omega_3^1,\, \alpha_3^1)$ for buyer $b = 1$

The absolute deviations in setting the means and covariances are within the error tolerances $(\mathcal{T}_\mu,\, \mathcal{T}_\sigma) = (0.01,\, 0.001)$. We performed goodness-of-fit tests using the G-test to summarize how well our Markov-modulated approach fits the values observed in the environment. We have $M - 1 = 9$ degrees of freedom available. The G-tests produced the test-statistic χ^2_9 values of 5.52491 and 6.05742 for the probability mass functions $F_{\varepsilon_2^1}(\cdot\,|\,\omega_2^1,\, \alpha_2^1)$ and $F_{\varepsilon_3^1}(\cdot\,|\,\omega_3^1,\, \alpha_3^1)$, respectively. The critical values should be drawn from the Chi-square distribution table with nine degrees of freedom. Let $\chi^2_{\alpha,9}$ denote the percentage point or value of the Chi-square random variable with nine degrees of freedom such that the probability that χ^2_9 exceeds this value is α. Since $\chi^2_{0.05,9} = 16.919$ for $\alpha = 0.05$, there is no reason to reject the null hypothesis that there is no significant difference from the observed frequencies. We should note that we may reiterate the optimization procedure by creating a finer grid for intervals labeled $m1$ and $m7$ of $F_{\varepsilon_2^1}(\cdot\,|\,\omega_2^1,\, \alpha_2^1)$ and $m9$ of $F_{\varepsilon_2^1}(\cdot\,|\,\omega_2^1,\, \alpha_2^1)$.

3.6 Summary

In this chapter, we presented the stochastic framework for the decision problems that will be analyzed in what follows. In the first part, we modeled the time series of the buyers' order commitments and realized orders as a multiplicative forecast evolution process through the MMFE technique. The resulting evolution model is a sophisticated one, as we modeled the evolution as a multiplicative process and accommodated the revision limits stipulated in QF contracts. In the second part of the chapter, we introduced a finite Markov chain approximation to the martingale forecast evolution process having some revision limits. The problem of estimating the transition probabilities of the Markov chain was addressed by a general optimization model maximizing the goodness-of-fit to observations. This numerical estimation process imposed some regularity constraints to accommodate the revision limits and the correlations of commitment updates across buyers and through time. At the end, we illustrated these ideas with a concrete example where a computational process was given for a specific industry case with the real data.

Consequently, in the subsequent part of the research, we will use this probabilistic model of the commitment evolution in conjunction with an analytical model of the production and inventory planning under multi-period QF contracts. This integrated use will provide an enhanced variability representation which enables the manufacturer to better capture the underlying forecasting behaviors of the buyers. As regards the finite Markov chain approximation, it will facilitate the probability modeling of the sequential production/inventory decision model under QF contracts. It essentially provides an approach to discretization in the associated stochastic dynamic programming. This will make all the random variables hereinafter discrete.

CHAPTER 4

MULTI-PERIOD STOCHASTIC PRODUCTION/INVENTORY DECISION MODEL

In this chapter, we present the manufacturer's multi-period stochastic production/inventory decision model, where stochastic elements are described by the probabilistic framework discussed in the preceding chapter. In §4.1 we first describe the sequence of events that take place in any period for the execution of the system. We then formulate the manufacturer's problem as a finite-horizon dynamic production/inventory model. The problem is characterized by the presence of random forecast evolution with revision limits, on the demand side, and production capacity restrictions with an option of subcontracting, on the supply side. Order commitments are stated as the explicit component of the state space. In §4.2 we characterize the structure and properties of optimal replenishment policies of the manufacturer. §4.3 discusses a state space compaction.

4.1 The Model

In each time period, the manufacturer is faced with the problems of (*i*) determining whether or not to place a replenishment order and (*ii*) if an order is placed, how much to order in satisfying uncertain demand. The condition of the inventory system that we have already described in §2.2 is to be reviewed in every period. Each time a replenishment decision is made, the manufacturer must plan ahead for L periods since replenishment orders arrive only after a lead-time of L periods. After the replenishment decision has been made, the buyer orders are realized throughout the current period. If the buyer order realizations exceed the on-hand inventory, then unmet order quantity is backlogged with a penalty cost. Otherwise, for each unit of remaining inventory at the end of the current period an inventory holding cost is incurred. The system then progresses to the next time period.

We assume that the manufacturer does not hold back any inventory for future periods (that is, he does not make any inventory rationing). He differentiates between the buyers only in order fulfillment process, allocating his on-hand inventory according to a given precedence relationship among the buyers (that is, total requirement of the highest-priority buyer is simply filled first, then the next highest-priority buyer is satisfied, and so on). The manufacturer tries to choose those actions that will minimize the sum of the costs accumulated as the inventory system progresses.

Assume that the manufacturer executes the agreed-upon supply contracts over a finite time span of length $N + L$, where the number of replenishment decision points is N. The sequence of events that take place at any period $s = 1, 2, \ldots, N$ for the execution of the contracts are then as follows. Figure 4.1 depicts the timing and the decision structure.

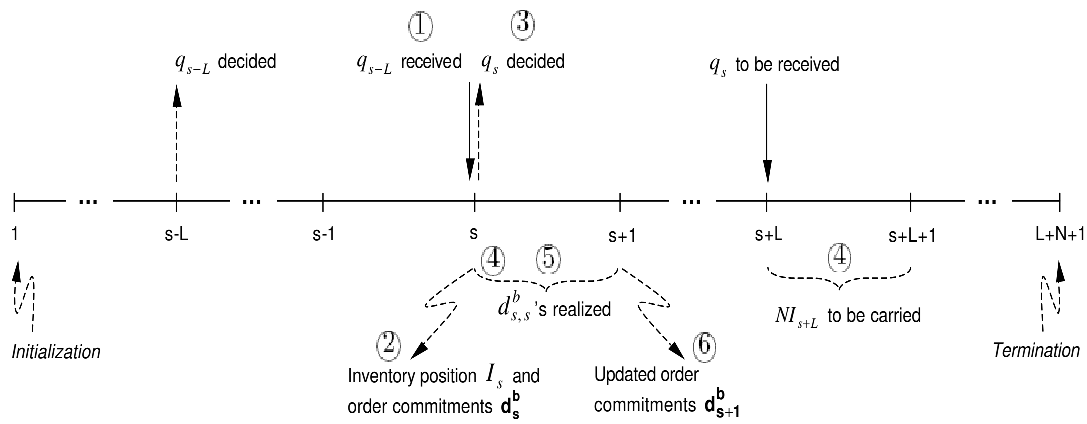


Figure 4.1: The sequence of events

1. At the beginning of period s, the manufacturer receives the replenishment orders from L periods ago. These have been supplied with in-house production and subcontractor [cf. Chapter 2].

2. The manufacturer reviews the state of the system; namely, current inventory position (i.e., all outstanding replenishment orders less all realized buyer orders) and the buyers' order commitments made at the end of period $s-1$. These cover all the periods in the following H-period horizon $t \in [s,\, s+H-1]$. If it is the case that $H < L$, the mean realized order per period is taken as the best available order information for the periods beyond the commitment horizon H.

3. The replenishment decision is made in anticipation of receiving buyer orders during the following L periods. If in-house production capacity is not sufficient to produce all the amount decided (q_s), the manufacturer has an option of subcontracting as mentioned in §2.2.

4. The expected costs associated with period s are incurred. Replenishment cost does not include any fixed cost of ordering but rather variable costs of in-house production and subcontracting. There exist also costs associated with carrying inventory and backorders. We assume, due to the replenishment lead-time of L periods, that these costs depend on the expected net inventory level that will be carried over in period $s+L$. That is to say, the manufacturer incurs an expected holding (backorder) cost charged to period s only for the expected positive (negative) net inventory that will be carried over in period $s+L$.

5. During period s the order realizations $d^b_{s,s}$ are observed from the buyers. These realized orders (plus outstanding backorders if there is any) are served by the manufacturer's on-hand inventory, and any shortages will become backorders.

6. At the end of period s the buyers review and update their available order commitments within the bounds constructed by the agreed-upon quantity flexibility contracts. Then they submit the updated order commitments for all the periods in the following H-period horizon $t \in [s+1,\, s+H]$.

At the end of the planning horizon (i.e., at the end of period $N+L$), if there is positive inventory left at the manufacturer, it will be sold at a unit salvage price being equivalent to unit subcontract procurement cost. Conversely, any inventory shortage at the end of period $N+L$ will be fulfilled by a final subcontract procurement. The use of unit cost of subcontracting in the end-of-horizon transactions is reasonable since after coming to the end of a contract period only the subcontract capacity would be available and the price requested by the manufacturer can reasonably be at most the unit subcontract procurement cost.

Consequently, we have a sequence of L-period rolling horizon problems with one period re-planning frequency over the $(N + L)$-period planning horizon. The number of decision points for the manufacturer thus is N. We shall model this recurring problem to solve for an ordered set of replenishment decisions to be taken in any system state so that the total expected cost of replenishment and inventory will be minimized over the planning horizon. Before giving the description of the model in detail, we provide the following additional notation that is used throughout the exposition:

Decision variables:

q_s : replenishment order placed in period s (in-house production plus subcontract orders, if any) for delivery in period $s + L$.

I_s : inventory position before ordering in period s.

TI_s : inventory position after ordering in period s, $TI_s \geqslant I_s$. Thus, $q_s = TI_s - I_s$.

NI_s : net inventory level (i.e., on-hand inventory or backorders) to be carried over in period s.

Parameters:

K : finite per-period capacity for in-house production.

h : unit cost of carrying inventory per period.

π_b : unit backorder penalty per period for buyer b. We have $\pi_1 > \pi_2$, indicating that $b = 1$ stands for the highest-priority buyer.

c_{pi} : unit cost of in-house production.

c_{ps} : unit cost of subcontracting. We have $c_{ps} \geqslant c_{pi}$.

$\mathbf{1}(A_s)$: indicator function of the event $A_s = \{TI_s - I_s > K\}$, which is

$$\mathbf{1}(A_s) = \begin{cases} 1, & \text{if } TI_s - I_s > K \text{ (i.e., subcontracting is made)} \\ 0, & \text{otherwise.} \end{cases}$$

Additionally, define $Z^b_{[s,s+L)}$ as the random variable denoting total of L order occurrences to be received from buyer b over the time interval $[s,\, s+L)$ such that $Z_{[s,s+L)} = \sum_b Z^b_{[s,s+L)}$, where

$$Z^b_{[s,s+L)} \quad = \quad \sum_{k=1}^{L} D^b_{s+k-1,s+k-1}. \tag{4.1}$$

Its probability distribution function is denoted by $f_{Z^b_L}(\cdot \mid \mathbf{d_s})$, conditioned on the value of order commitment vector $\mathbf{d_s}$ available at the beginning of period s. Similarly, we have $f_{Z_L}(\cdot \mid \mathbf{d_s})$ for $Z_{[s,s+L)}$. In computing the expected costs of carrying inventory and backorders we will refer to these conditional probabilities.

The inventory position at the beginning of period s, I_s, can be written as

$$I_s = I_0 + \sum_{i=1}^{s-1} q_i - \sum_b Z^b_{[1,s)}, \tag{4.2}$$

for $s = 2,\, 3,\, \ldots,\, N+L$ (where $q_i = 0$ for $i > N$ since we have N replenishment decisions). It denotes total on order minus total realized order from the buyers before the manufacturer's ordering in period s, including outstanding backorders if there is any. The replenishment decision made in period s will bring the inventory position to TI_s. The costs associated with period s are of several classes. Let $PC_s(TI_s,\, I_s)$ denote the replenishment cost to be incurred for an amount $q_s = TI_s - I_s$,

$$PC_s(TI_s,\, I_s) = \begin{cases} c_{pi}(TI_s - I_s) + (c_{ps} - c_{pi})(TI_s - I_s - K)\mathbf{1}(A_s), & TI_s > I_s \\ 0, & TI_s \leqslant I_s, \end{cases} \tag{4.3}$$

for $s = 1,\, 2,\, \ldots,\, N$. We assume, due to the L-period replenishment lead-time, that the costs associated with carrying inventory and backorders depend on the expected net inventory level that will be carried over in period $s+L$, NI_{s+L}, which is found to be

$$\begin{aligned} NI_{s+L} \quad &= \quad \sum_{i=1}^{s} q_i - \sum_b Z^b_{[1,s+L)} \\ &= \quad TI_s - Z_{[s,s+L)} \\ &= \quad I_{s+1} - Z_{[s+1,s+L)}, \end{aligned} \tag{4.4}$$

for $s = 1,\, 2,\, \ldots,\, N+1$ (where $q_i = 0$ for $i = N+1$). It denotes total received order minus total satisfied demand before ordering in period $s+L$. Let $L_s(TI_s,\, \mathbf{d_s})$ denote the current L-period costs associated with inventory carrying and backorders, given that the inventory position is set to TI_s after the manufacturer's ordering in period

s. Its formula depends on the number of buyers involved because of the given priority scheme among them. It is readily found, when $B = 2$ (i.e., two contract buyers), to be

$$
\begin{aligned}
L_s(TI_s,\, \mathbf{d_s}) \;\; = & \\
& h\,\mathrm{E}\,[TI_s - z_1 - z_2 \mid z_1 + z_2 < TI_s,\, \mathbf{d_s}] \\
& +\; \pi_1\,\mathrm{E}\,[z_1 - TI_s \mid z_1 > TI_s,\, \mathbf{d_s}] \;+\; \pi_2\,\mathrm{E}\,[z_2 \mid z_1 > TI_s,\, \mathbf{d_s}] \\
& +\; \pi_2\,\mathrm{E}\,[z_1 + z_2 - TI_s \mid z_1 < TI_s,\, \mathbf{d_s}] \qquad (4.5)
\end{aligned}
$$

for $s = 1, 2, \ldots, N$. Due to our finite Markov approximation introduced in §3.2 to §3.4, the continuous cost function $L_s(TI_s,\, \mathbf{d_s})$ in (4.5) is discretized as

$$
\begin{aligned}
L_s(TI_s,\, \mathbf{d_s}) \;\; = & \\
& h \sum_{0}^{TI_s} \sum_{0}^{TI_s - z_1} (TI_s - z_1 - z_2)\, f_{Z_L^2}(z_2 |\, \mathbf{d_s})\, f_{Z_L^1}(z_1 |\, \mathbf{d_s}) \\
& +\; \pi_1 \sum_{TI_s}^{\infty} (z_1 - TI_s)\, f_{Z_L^1}(z_1 |\, \mathbf{d_s}) \;+\; \pi_2 \sum_{TI_s}^{\infty} \sum_{0}^{\infty} z_2\, f_{Z_L^2}(z_2 |\, \mathbf{d_s})\, f_{Z_L^1}(z_1 |\, \mathbf{d_s}) \\
& +\; \pi_2 \sum_{0}^{TI_s} \sum_{TI_s - z_1}^{\infty} (z_2 - (TI_s - z_1))\, f_{Z_L^2}(z_2 |\, \mathbf{d_s})\, f_{Z_L^1}(z_1 |\, \mathbf{d_s}), \qquad (4.6)
\end{aligned}
$$

for $s = 1, 2, \ldots, N$. The discounting factor is conjectured to be unity without loss of generality. $L_s(TI_s,\, \mathbf{d_s})$ comprises of three main components. The first term denotes the expected cost of inventory holding. The second represents the expected backorder penalty costs for the case that the total requirement for the highest-priority buyer is greater than or equal to the manufacturer's availability, which is given by the second and third terms. The complement of the shortage case is addressed in the last term. The lower priority buyer will be served from the balance of TI_s serving the first buyer.

Observe that the manufacturer incurs an expected holding (backorder) cost charged to period s only for the expected positive (negative) net inventory NI_{s+L} that will be carried over in period $s + L$. Thus denote $J_s(TI_s,\, I_s,\, \mathbf{d_s})$ as the current-period cost associated with period s, which is given by

$$
J_s(TI_s,\, I_s,\, \mathbf{d_s}) = PC_s(TI_s,\, I_s) + L_s(TI_s,\, \mathbf{d_s}), \qquad (4.7)
$$

for $s = 1, 2, \ldots, N$. We now state an important preliminary result on the current-period cost function $J_s(TI_s,\, I_s,\, \mathbf{d_s})$.

Proposition 4.1 *Being a newsvendor-type cost function, $J_s(TI_s,\, I_s,\, \mathbf{d_s})$ is convex in TI_s for all values of $(I_s,\, \mathbf{d_s})$.*

Proof: See Appendix A.1.

□

Let $\pi = \{TI_s,\ 1 \leqslant s \leqslant N\} \in \Pi$ denote a policy specifying an ordered set of replenishment decisions, there being one decision for each system state $(I_s,\, \mathbf{d_s})$ at the beginning of period s. Hence Π is the set of all possible policies. Let $G_1(I_1,\, \mathbf{d_1},\, \pi)$ be the total expected cost over a finite time horizon of $N + L$ periods, given that the policy π is being used and the initial system state is observed as $(I_1,\, \mathbf{d_1})$,

$$G_1(I_1,\, \mathbf{d_1},\, \pi) \;=\; E_{\boldsymbol{\mathcal{E}_1},\boldsymbol{\mathcal{E}_2},\ldots,\boldsymbol{\mathcal{E}_N}}\left[\sum_{s=1}^{N} J_s(TI_s,\, I_s,\, \mathbf{d_s}) \;\middle|\; I_1,\, \mathbf{d_1}\right] + J_{N+1}(I_{N+1},\, \mathbf{d_{N+1}}). \quad (4.8)$$

$E_{\boldsymbol{\mathcal{E}_1},\boldsymbol{\mathcal{E}_2},\ldots,\boldsymbol{\mathcal{E}_N}}[\,\cdot \mid I_1,\, \mathbf{d_1}]$ represents the conditional expectation, given the initial inventory position and order commitments. It is taken with respect to commitment update vectors $\boldsymbol{\mathcal{E}_1},\, \boldsymbol{\mathcal{E}_2},\, \ldots,\, \boldsymbol{\mathcal{E}_N}$. This is because the forecast evolution model allows us to express the uncertainty about the future values of inventory position and order commitments in terms of $\boldsymbol{\mathcal{E}_s}$'s, $s \in [1,\, N]$. As a linear end-of-horizon condition, we have $J_{N+1}(\cdot,\, \cdot) \;=\; -\, c_{ps}\, E_{\boldsymbol{\mathcal{E}_{N+1}},\boldsymbol{\mathcal{E}_{N+2}},\ldots,\boldsymbol{\mathcal{E}_{N+L}}}[\, NI_{N+L}\,]$ to be evaluated at period N (i.e., decision epoch N). In other words, at the end of period $N + L$ the manufacturer salvages the left-over inventory (i.e., positive values of NI_{N+L}) for a unit salvage revenue of c_{ps} and satisfies any inventory shortage (i.e., negative values of NI_{N+L}) with a final subcontract procurement which is equal to c_{ps}. As a consequence, the total expected cost $G_1(I_1,\, \mathbf{d_1},\, \pi)$ depends not only upon the initial system state $(I_1,\, \mathbf{d_1})$ and the series of decisions $TI_1,\, TI_2,\, \ldots,\, TI_N$, but also upon the sequence of random commitment update vectors $\boldsymbol{\mathcal{E}_1},\, \boldsymbol{\mathcal{E}_2},\, \ldots,\, \boldsymbol{\mathcal{E}_{N+L}}$.

We are interested in minimizing $G_1(I_1,\, \mathbf{d_1},\, \pi)$ over all $\pi \in \Pi$. We call the policy π^* optimal if

$$G_1(I_1,\, \mathbf{d_1},\, \pi^*) \;=\; \inf_{\pi\in\Pi}\, G_1(I_1,\, \mathbf{d_1},\, \pi). \quad (4.9)$$

Due to finiteness and convergence of the sum in (4.8), the infimum in (4.9) can be replaced by the minimum. Therefore, when the optimal policy π^* is used, the expected value of the total cost is minimized over all potential decision sequences. Let

$V_1(I_1,\, \mathbf{d_1})$ denote the *minimum* expected total cost from the beginning of period 1 to the end of period $N+L$, given that the system state is observed as $(I_1, \mathbf{d_1})$ at the beginning of period 1 and that an optimal decision is made in each period 1, 2, ..., N. Consequently, we can derive the optimality equation as

$$V_s(I_s,\, \mathbf{d_s}) \;=\; \min_{TI_s \geqslant I_s} \left\{ J_s(TI_s,\, I_s,\, \mathbf{d_s}) + E_{\boldsymbol{\mathcal{E}_s}}[V_{s+1}(TI_s - \sum_b D^b_{s,s},\, \mathbf{D_{s+1}})] \right\}, \tag{4.10}$$

which follows since all the terms inside the expectation in (4.8) are nonnegative, integrable functions, and the vector sequence $\{\boldsymbol{\mathcal{E}_s},\ 1 \leqslant s \leqslant N+L\}$ are independent and identically distributed due to the assumptions of the forecast evolution model, as mentioned in §3.1. Using the optimality equation we may recursively solve for $V_s(I_s,\, \mathbf{d_s})$. The main theoretical result is that the policy determined by the (optimality) equation (4.10) will be optimal.

4.2 Properties of the Optimal Policy

In this section, we obtain structural results about the finite horizon model introduced in §4.1. The characterization of the optimal policy for the problem under study is complicated as the manufacturer can place two types of orders in each period, in-house production and subcontracting, depending on its capacity level and the available order commitments. A replenishment source is characterized by its unit cost of replenishment the manufacturer incurs and the level of capacity it has. As we have already mentioned in §2.2, the manufacturer has a restriction on the maximum amount that can be ordered from the in-house production, but he has a more costly subcontracting option with infinite supply. Orders are assumed to arrive only after a lead-time of L periods, independent of the replenishment source.

Let $G_s(TI_s,\, I_s,\, \mathbf{d_s})$ denote the *(suboptimal)* expected total cost from the beginning of period s to the end of period $N+L$, given that the system state is observed as $(I_s,\, \mathbf{d_s})$ at the beginning of period s,

$$G_s(TI_s,\, I_s,\, \mathbf{d_s}) \;=\; J_s(TI_s,\, I_s,\, \mathbf{d_s}) + E_{\boldsymbol{\mathcal{E}_s}}[\, V_{s+1}(TI_s - \sum_b D^b_{s,s},\, \mathbf{D_{s+1}})\,]. \tag{4.11}$$

It can be readily derived from Proposition 4.1 that the function $G_s(TI_s,\, I_s,\, \mathbf{d_s})$ is convex in TI_s for given values of I_s and $\mathbf{d_s}$. This will be proved later in this section.

Since $G_s(\cdot)$ in (4.11) is convex a minimizer exists [1]. Define TI_s^* to be the optimal value of TI_s at which the minimal value is attained by $G_s(TI_s,\, I_s,\, \mathbf{d_s})$; that is,

$$G_s(TI_s^*,\, I_s,\, \mathbf{d_s}) \;=\; \min_{TI_s} G_s(TI_s,\, I_s,\, \mathbf{d_s}).$$

In view of the finite in-house capacity level K, we may define the following two cases on the value of $q_s = TI_s - I_s$, replenishment order quantity in period s.

Case1: $q_s \leqslant K$.

It accounts for the case where the inventory position before ordering, I_s, is relatively higher, and hence only the in-house production source is engaged to satisfy the expected buyer orders. Let $G_s^{inh}(TI_s,\, I_s,\, \mathbf{d_s})$ be the expected total cost from period s through $N+L$, given that the system is in state $(I_s,\, \mathbf{d_s})$ and only the in-house capacity is used at the beginning of period s.

$$\begin{aligned} G_s^{inh}(TI_s,\, I_s,\, \mathbf{d_s}) \;&= \\ &c_{pi}(TI_s - I_s) + L_s(TI_s,\, \mathbf{d_s}) + E_{\boldsymbol{\mathcal{E}_s}}[\, V_{s+1}(TI_s - \sum_b D_{s,s}^b,\, \mathbf{D_{s+1}})\,], \end{aligned} \qquad (4.12)$$

which is convex in TI_s for given values of I_s (when $I_s \leqslant TI_s \leqslant I_s + K$) and $\mathbf{d_s}$. Define TI_s^{inh} to be the minimizer of $G_s^{inh}(TI_s,\, I_s,\, \mathbf{d_s})$,

$$G_s^{inh}(TI_s^{inh},\, I_s,\, \mathbf{d_s}) \;=\; \min_{TI_s} G_s^{inh}(TI_s,\, I_s,\, \mathbf{d_s}).$$

If it is the case that $q_s = 0$ or $q_s = K$, then $\partial G_s^{inh}(TI_s,\, I_s,\, \mathbf{d_s})/\partial TI_s$ vanishes at $TI_s = I_s$ and $TI_s = I_s + K$, respectively, by the optimality of TI_s^{inh}.

Case2: $q_s > K$.

It is the case where I_s is not sufficient to satisfy the expected buyer orders internally, and hence both in-house production source and subcontract capacity are engaged. Let $G_s^{sub}(TI_s,\, I_s,\, \mathbf{d_s})$ be the expected total cost from period s through $N+L$, given that the system is in state $(I_s,\, \mathbf{d_s})$ and both in-house and subcontract capacity are engaged at the beginning of period s.

$$G_s^{sub}(TI_s,\, I_s,\, \mathbf{d_s}) \;=\; G_s^{inh}(TI_s,\, I_s,\, \mathbf{d_s}) + (c_{ps} - c_{pi})(TI_s - I_s - K), \qquad (4.13)$$

[1] The minimizer might not be unique as the cost function $L_s(\cdot)$ in (4.5) is discretized by our finite Markov approximation introduced in §3.2 to §3.4.

which is also convex in TI_s for given values of I_s and $\mathbf{d_s}$. Define TI_s^{sub} to be the minimizer of the cost function $G_s^{sub}(TI_s,\, I_s,\, \mathbf{d_s})$,

$$G_s^{sub}(TI_s^{sub},\, I_s,\, \mathbf{d_s}) \;=\; \min_{TI_s}\, G_s^{sub}(TI_s,\, I_s,\, \mathbf{d_s}).$$

Note that $TI_s^{sub} \geqslant I_s + K$.

Therefore we have two critical levels, TI_s^{inh} and TI_s^{sub}, associated with two different replenishment sources, which is due to the separability of the cost function $G_s(TI_s,\, I_s,\, \mathbf{d_s})$ in (4.11) as shown in (4.12) and (4.13). The following result states the relationship between these critical levels.

Proposition 4.2 *We have* $TI_s^{sub} \leqslant TI_s^{inh}$.

Proof: We have $G_s^{sub}(TI_s,\, I_s,\, \mathbf{d_s}) \geqslant G_s^{inh}(TI_s,\, I_s,\, \mathbf{d_s})$ and $G_s^{sub} - G_s^{inh}$ is increasing in TI_s since $(c_{ps} - c_{pi})$ and $(TI_s - I_s - K)$ on the right-hand side of (4.13) are nonnegative. Hence the convexity of them implies that their minimizers have $TI_s^{sub} \leqslant TI_s^{inh}$. □

Karlin (1958) demonstrates that for multi-period problem with backorders and no fixed production cost, optimal policy is of order-up-to type under a strictly convex cost function. Thus, in view of these results, the following theorem identifies the structure of the optimal policy that specifies the manufacturer's replenishment decisions.

Theorem 4.1 *By using Proposition 4.1 and 4.2, for any decision period* $s = 1,\, 2,\, \ldots,\, N$ *we have*

(i) $G_s(TI_s,\, I_s,\, \mathbf{d_s})$ *is convex in* TI_s*,*

(ii) $V_s(I_s,\, \mathbf{d_s})$ *is convex in* I_s*,*

(iii) We have a staircase optimal policy due to the presence of different replenishment sources with a finite in-house capacity level K*. It is of state-dependent order-up-to type, given a state vector* $(I_s,\, \mathbf{d_s})$ *at the beginning of period* $s = 1, 2, \ldots, N$*. Then the optimal order-up-to level,* $TI_s^*(I_s,\, \mathbf{d_s})$*, is*

$$TI_s^*(I_s,\, \mathbf{d_s}) = \begin{cases} TI_s^{sub}(\mathbf{d_s}) & I_s \leqslant TI_s^{sub}(\mathbf{d_s}) - K \\ I_s + K & TI_s^{sub}(\mathbf{d_s}) - K \leqslant I_s \leqslant TI_s^{inh}(\mathbf{d_s}) - K \\ TI_s^{inh}(\mathbf{d_s}) & TI_s^{inh}(\mathbf{d_s}) - K \leqslant I_s \leqslant TI_s^{inh}(\mathbf{d_s}) \\ I_s & TI_s^{inh}(\mathbf{d_s}) \leqslant I_s, \end{cases} \tag{4.14}$$

where $TI_s^{inh}(\mathbf{d_s})$ *and* $TI_s^{sub}(\mathbf{d_s})$ *depend only on the commitment state* $\mathbf{d_s}$*, independent of the inventory state* I_s*, whereas* $TI_s^*(I_s,\, \mathbf{d_s})$ *is a function of* $(I_s,\, \mathbf{d_s})$*.*

Proof: See Appendix A.2.

□

Figure 4.2 helps in understanding the staircase structure of the optimal policy.

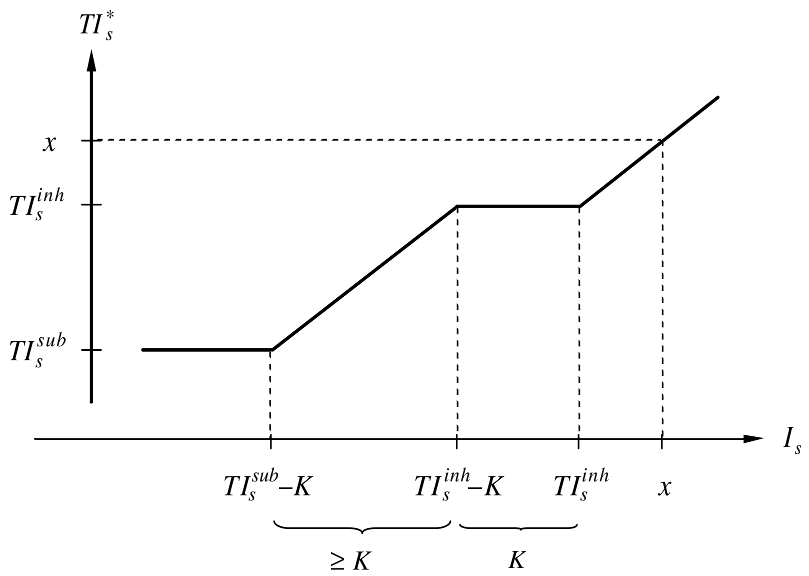


Figure 4.2: Staircase structure of the optimal policy with two replenishment sources

Note that we will write TI_s^{sub}, TI_s^{inh} and TI_s instead of $TI_s^{sub}(\mathbf{d_s})$, $TI_s^{inh}(\mathbf{d_s})$ and $TI_s(I_s,\, \mathbf{d_s})$, respectively, throughout the study, unless stated otherwise. The optimal replenishment policy is described in terms of two critical order-up-to levels,

TI_s^{inh} and TI_s^{sub}, with $TI_s^{sub} \leqslant TI_s^{inh}$, corresponding to two replenishment sources. Bradley (2004), Feng et al. (2006), and Tan and Alp (2008) have shown the optimality of such a policy type in different problem settings for a number of discrete and continuous-time inventory models with two replenishment sources. They also base on the separability of the recursive equations for two sources.

The meaning of (4.14) should be emphasized. The optimal replenishment policy essentially classifies the state space (which is $(BH+1)$-dimensional space) in any period s into several regions in a two-dimensional space, where each region specifies its own optimal order-up-to level. Given any period $s \in [1, N]$, the two dimensions are namely, I_s and $\sum_b \sum_{k=1}^{H} d_{s-1,s+k-1}^b$. The number of allowable regions in this two-dimensional space depends on the in-house capacity level K relative to the total of order commitments available at that period. If it is the case that $KH < \sum_b \sum_{k=1}^{H} d_{s-1,s+k-1}^b$ (i.e., the in-house capacity falls short), then we have only two allowable regions in a two-dimensional space, otherwise four regions are involved. Figure 4.3 depicts possible cases on the value of I_s, for a given value of $\mathbf{d_s}$, together with the corresponding optimal ordering decisions.

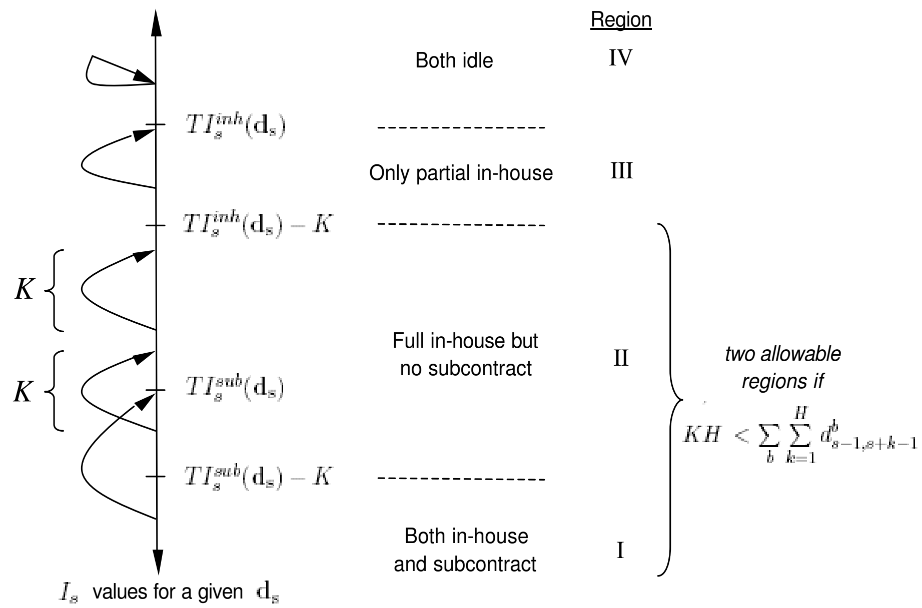


Figure 4.3: Cases on the value of I_s and the corresponding optimal decisions

In Figure 4.3, the region I; i.e., $(-\infty,\, TI_s^{sub}(\mathbf{d_s}) - K]$, corresponds to cases where both the in-house and subcontract production sources are engaged. In the region II, the full in-house capacity and no subcontract capacity are used for replenishment. In the region III, on the other hand, the in-house production source is partially engaged. Finally, when $I_s \in [TI_s^{inh}(\mathbf{d_s}),\, \infty)$; i.e., in the region IV, neither production sources are called for. An important point that can also be observed in Figure 4.3 is that we are imposing the condition that $TI_s^{inh}(\mathbf{d_s}) - TI_s^{sub}(\mathbf{d_s}) \geqslant K$.

For illustrative purposes, we consider a restricted problem size. Figure 4.4 gives the optimal order-up-to levels of period 1, TI_1^*, for an allowable range of values of state vector $(I_1,\, \mathbf{d_1})$. The staircase structure of the optimal policy given in Theorem 4.1 is apparent in this figure where y-axis is for different values of the sum $\sum_b \sum_{k=1}^{H} d_{0,k}^b$.

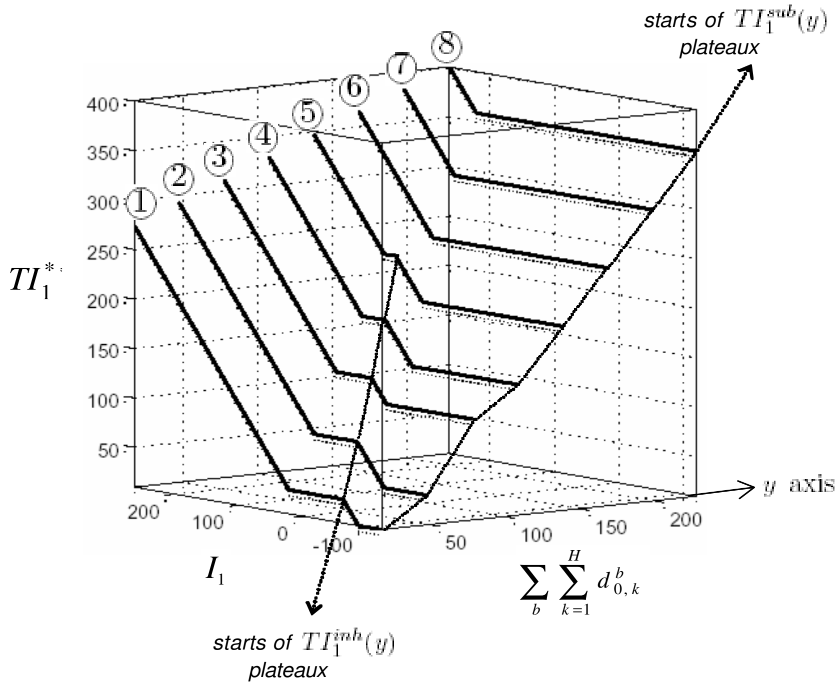


Figure 4.4: Optimal order-up-to levels TI_1^* of period 1 for various system state $(I_1,\, \mathbf{d_1})$

The first five plots [① – ⑤] correspond to the case when $KH \geqslant \sum_b \sum_{k=1}^{H} d_{0,k}^b$, and the last three [⑥ – ⑧] when otherwise. In the first group, the optimal order-up-to level is defined by two critical levels; i.e., $TI_1^{sub}(y)$ and $TI_1^{inh}(y)$ with $TI_1^{sub}(y) \leqslant TI_1^{inh}(y)$, which depend only on K and $\mathbf{d_1}$ (or equivalently, values on axis y). These two groups of plots are observed since the in-house capacity level K relative to the total order commitments determines the number of allowable regions in the two-dimensional (I_1 and $\sum_b \sum_{k=1}^{H} d_{0,k}^b$) domain in Figure 4.4.

We shall now provide some theoretical properties about the behavior of the optimal order-up-to levels. The following theorem is given to conclude a certain monotonicity result about the optimal policy. More specifically, for a fixed value of order commitment vector, $\mathbf{d}$, it relates the optimal order-up-to level in period s, $TI_s^*(x, \mathbf{d})$ to the one in period $s+1$, $TI_{s+1}^*(x, \mathbf{d})$.

Theorem 4.2 *For a given value of order commitment vector,* $\mathbf{d}$*, we have the following monotonicity statements for* $s = 1, 2, \ldots, N$*, where* x *and* y *denote inventory position before ordering and after ordering, respectively.*

(i) $\partial V_s(x, \mathbf{d})/\partial x \geqslant \partial V_{s+1}(x, \mathbf{d})/\partial x \ \ \forall x$

(ii) $\partial G_s(y, x, \mathbf{d})/\partial y \geqslant \partial G_{s+1}(y, x, \mathbf{d})/\partial y \ \forall y$

(iii) $TI_s^*(x, \mathbf{d}) \leqslant TI_{s+1}^*(x, \mathbf{d})$

Proof: See Appendix A.3.

□

Statements (i) and (ii) of Theorem 4.2 say that the effect of the replenishment decision y on the cost in the future gets suppressed as time to termination gets close. Hence, statement (iii) states that the optimal order-up-to level of period s is less than that of period $s+1$, assuming the same value of order commitment vector in period s is repeated in period $s+1$; i.e., $\mathbf{d_s} \equiv \mathbf{d_{s+1}}$.

The following theorem states another form of monotonicity about the optimal order-up-to levels. To be specific, it relates the optimal order-up-to levels in period s to different values of order commitment vectors available in that period.

Theorem 4.3 *Suppose two distinct values* $\bar{\mathbf{d}}$ *and* $\mathbf{d}$ *that the order commitment vector at the beginning of period* s *can take, with* $\sum_b \sum_{k=1}^{H} \bar{d}_{s-1,s+k-1}^b \leqslant \sum_b \sum_{k=1}^{H} d_{s-1,s+k-1}^b$*. Thus we have the following regularity statements for* $s = 1, 2, \ldots, N$*.*

(i) $\partial V_s(x, \bar{\mathbf{d}})/\partial x \geqslant \partial V_s(x, \mathbf{d})/\partial x \quad \forall x$

(ii) $\partial G_s(y, x, \bar{\mathbf{d}})/\partial y \geqslant \partial G_s(y, x, \mathbf{d})/\partial y \quad \forall\, y$

(iii) $TI_s^*(x,\, \bar{\mathbf{d}}) \leqslant TI_s^*(x,\, \mathbf{d})$

Proof: See Appendix A.4.

□

Statements (i) and (ii) of Theorem 4.3 say that the replenishment decision y in period s affects the costs less for a larger total of order commitments available in that period. Statement (iii) says that the optimal order-up-to level increases as the total of order commitments increases as expected.

These results will help to make the computation associated with the dynamic programming recursions less demanding in the sense that the corresponding state space can be searched more efficiently. We will elaborate on the use of these properties in Section 4.3.

4.3 State Space Compaction

As our solution procedure will be enumerative, any bounds on the optimal order-up-to levels might make the computation associated with the recurrence relations less demanding.

In deriving bounds on the optimal-order-up-to levels, it appears to be more effective to develop bounds on the first-order condition function $\partial\, G_s(y,\, x,\, \mathbf{d}) \,/\, \partial\, y$. Then the solutions to these bounding functions provide the bounds for the optimal order-up-to levels. The following corollary to Theorems 4.2 and 4.3 may be used to show that a relaxed upper bound exists on the optimal order-up-to levels.

Corollary 4.1 (to Theorems 4.2 and 4.3) *For a given order commitment vector,* $\mathbf{d}$*, the optimal order-up-to level* $TI_N^*(x,\, \mathbf{d})$ *of the last decision period is an upper bound for the optimal order-up-to levels* $TI_s^*(x,\, \bar{\mathbf{d}})$ *of all other periods* $s < N$ *for those values of* $\bar{\mathbf{d}}$ *such that* $\sum_b \sum_{k=1}^{H} \bar{d}^b_{s-1,s+k-1} \leqslant \sum_b \sum_{k=1}^{H} d^b_{N-1,N+k-1}$.

Corollary 4.1 follows since we have $\partial G_s(y, x, \bar{\mathbf{d}})/\partial y \geqslant \partial G_s(y, x, \mathbf{d})/\partial y \geqslant \partial G_N(y, x, \mathbf{d})/\partial y$ due to Theorems 4.2 and 4.3. This result allows us to set a recursive upper bound in backward progress. To derive the upper bound, we need to study the final-period problem covering the L periods extending out from period N. Consider the first-order condition function of period N, $\partial G_N(y,\, x,\, \mathbf{d})/\partial y$,

$$\begin{aligned}
&\partial\, G_N(y,\, x,\, \mathbf{d}) \,/\, \partial\, y \\
&\quad = \partial J_N(y,\, x,\, \mathbf{d})/\partial\, y \,+\, \partial\, E_{\boldsymbol{\mathcal{E}_N}}[\, V_{N+1}(y - \sum_b D^b_{N,N},\, \mathbf{D_{N+1}})\,] \,/\, \partial\, y \\
&\quad = \partial J_N(y,\, x,\, \mathbf{d})/\partial\, y \,+\, \partial\, E_{\boldsymbol{\mathcal{E}_N}}[\, -c_{ps}(y - \sum_b D^b_{N,N} - Z_{[N+1,N+L-1)})\,] \,/\, \partial\, y \\
&\quad = \partial J_N(y,\, x,\, \mathbf{d})/\partial\, y \,-\, c_{ps} \\
&\quad = c_{pi} \,+\, (c_{ps} - c_{pi})\mathbf{1}(A_N) \,+\, h F_{Z_L}(y\,|\,\mathbf{d}) \,-\, \pi_1 \Big[\, 1 - F_{Z^1_L}(y\,|\,\mathbf{d})\,\Big] \\
&\qquad\qquad -\, \pi_2 \Big[\, F_{Z^1_L}(y\,|\,\mathbf{d}) \,-\, F_{Z_L}(y\,|\,\mathbf{d})\,\Big] \,-\, c_{ps}.
\end{aligned} \tag{4.15}$$

Let TI^*_N be the solution of the first-order condition $\partial G_N(y,\, x,\, \mathbf{d})/\partial y = 0$. We may rearrange $\partial\, G_N(y,\, x,\, \mathbf{d})/\partial\, y$ in (4.15) to yield $\partial\, G_N(TI^\star_N,\, x,\, \mathbf{d})/\partial\, TI^\star_N = 0$ as

$$\begin{aligned}
&c_{pi} \,+\, (c_{ps} - c_{pi})\mathbf{1}(A_N) \,+\, h F_{Z_L}(TI^*_N\,|\,\mathbf{d}) \;\; = \\
&\qquad \pi_1 \Big[1 - F_{Z^1_L}(TI^*_N\,|\,\mathbf{d})\Big] \,+\, \pi_2 \Big[F_{Z^1_L}(TI^*_N\,|\,\mathbf{d}) - F_{Z_L}(TI^*_N\,|\,\mathbf{d})\Big] \,+\, c_{ps}.
\end{aligned} \tag{4.16}$$

Upper Bounding from the Optimal Terminal Decision

The derivation of the upper bound from the expression (4.16) is not straightforward as two distinct cumulative probability functions, $F_{Z^1_L}(\,\cdot\,|\,\mathbf{d})$ and $F_{Z_L}(\,\cdot\,|\,\mathbf{d})$, are involved associated with two contract buyers [2]. Since we know that $F_{Z^1_L}(y\,|\,\mathbf{d}) \geqslant F_{Z_L}(y\,|\,\mathbf{d})$ for any $y \geqslant 0$, we can obtain a relaxed upper bound

$$TI^*_N \;\;=\;\; F^{-1}_{Z_L}\left(\frac{\pi_1 + c_{ps} - c_{pi} - (c_{ps} - c_{pi})\mathbf{1}(A_N)}{h + \pi_1}\,|\,\mathbf{d}\right) \tag{4.17}$$

by simply assuming that $F_{Z^1_L}(y\,|\,\mathbf{d}) = F_{Z_L}(y\,|\,\mathbf{d})$. This envelope, in addition to the condition that $\sum_b \sum_{k=1}^{H} \bar{d}^b_{s-1,s+k-1} \leqslant \sum_b \sum_{k=1}^{H} d^b_{N-1,N+k-1}$, leads to an absolute upper bound for $TI^*_s(x,\, \bar{\mathbf{d}})$ of all $s < N$. Note that this upper bound can be calculated for any $\bar{\mathbf{d}}$ value as long as it satisfies the stochastic magnitude relation, hence it may be employed as an experimental upper bound on the allowable range of the order-up-to levels in any period.

[2] Note that one can readily generalize the first-order condition for more than two buyers.

We now interpret the expression (4.16). It allows one to see clearly the trade-off between the marginal cost in the current period and the marginal cost for the future periods. Target inventory decision is made by the manufacturer such that the marginal revenue of increasing the target position by one more unit is equal to the marginal cost. The left-hand side of expression (4.16) denotes the expected marginal cost of increasing the target position by one more unit, including unit replenishment cost and unit inventory holding cost incurred in the case that total requirement for all buyers is less than or equal to the manufacturer's availability. The right-hand side of expression (4.16), on the other hand, denotes the expected marginal revenue. It includes the unit gains that would be obtained in backorder penalty in case a backordering situation arose (reflected in the first two terms). The last term in the right-hand side is for the end-of-horizon condition. It accounts for the expected salvage revenue when one more unit of target position results in left-over inventory or the expected cost of outstanding backorders at the end of the planning horizon (i.e., at the end of period $N+L$).

Tightening the Upper Bounds by Inventory Position Regions

In addition to the bound described by (4.17), we might derive tighter upper bounds by developing tighter lower bounds (than $\partial G_N(y, x, \mathbf{d})/\partial y$) on $\partial\, G_s(y,\, I_s,\, \bar{\mathbf{d}}_{\mathbf{s}})/\partial\, y$ due to a tighter evaluation of the optimal cost function in $s+1$. To derive a tighter lower bound on $\partial\, G_s$, we study the constituents of the first order condition function

$$\partial\, G_s(y,\, x,\, \mathbf{d})/\partial\, y \;\;=\;\; \partial\, J_s(y,\, x,\, \mathbf{d})/\partial\, y \,+\, E_{\boldsymbol{\mathcal{E}}_s}\left[\partial\, V_{s+1}(y - \sum_b D^b_{s,s},\, \mathbf{d})/\partial\, y\right].$$

Now it is possible to show that $\partial\, E_{\boldsymbol{\mathcal{E}}_s}[V_{s+1}(.,.)]/\partial\, y \;=\; E_{\boldsymbol{\mathcal{E}}_s}[\partial\, V_{s+1}(.,.)/\partial\, y]$. Let us consider the decomposition of $E_{\boldsymbol{\mathcal{E}}_s}[\partial\, V_{s+1}(.,.)/\partial\, y]$. In view of Theorem 4.1, the $\partial\, V_{s+1}(y - \sum_b D^b_{s,s},\, \mathbf{d})/\partial\, y$ term can take four different forms corresponding to regions on the value of $y - \sum_b D^b_{s,s}$ ($= I_{s+1}$ values), with the associated probability of being in that particular region. Let $x = y - \sum_b D^b_{s,s}$. Hence, since $V_{s+1}(x, \mathbf{d})$ is determined by minimizing $G_{s+1}(TI_{s+1},\, x,\, \mathbf{d})$ over $x \leqslant TI_{s+1}$, we may write

$$V_{s+1}(x,\, \mathbf{d}) \;\;=\;\; \begin{cases} G_{s+1}(TI^{sub}_{s+1},\, x,\, \mathbf{d}) & x \leqslant TI^{sub}_{s+1} - K \\ G_{s+1}(x+K,\, x,\, \mathbf{d}) & TI^{sub}_{s+1} - K \leqslant\; x \leqslant TI^{inh}_{s+1} - K \\ G_{s+1}(TI^{inh}_{s+1},\, x,\, \mathbf{d}) & TI^{inh}_{s+1} - K \leqslant\; x \leqslant TI^{inh}_{s+1} \\ G_{s+1}(x,\, x,\, \mathbf{d}) & TI^{inh}_{s+1} \leqslant\; x, \end{cases}$$

where

$$G_{s+1}(\,\cdot\,,\,x,\,\mathbf{d}) \;\;=\;\; J_{s+1}(\,\cdot\,,\,x,\,\mathbf{d}) \,+\, E_{\boldsymbol{\mathcal{E}}_{\mathbf{s+1}}}[\,V_{s+2}(\,\cdot\,-\sum_b D^b_{s+1,s+1},\,\mathbf{D_{s+2}})\,].$$

Taking the partial derivative with respect to x,

$$\begin{aligned}
&\partial\, V_{s+1}(x,\,\mathbf{d})/\partial\, y \\
&= \begin{cases} -c_{ps} & x \leqslant TI^{sub}_{s+1} - K \\ \partial\, G_{s+1}(x+K,\,x,\,\mathbf{d})/\partial\, y & TI^{sub}_{s+1} - K \leqslant x \leqslant TI^{inh}_{s+1} - K \\ -c_{pi} & TI^{inh}_{s+1} - K \leqslant x \leqslant TI^{inh}_{s+1} \\ \partial\, G_{s+1}(x,\,x,\,\mathbf{d})/\partial\, y & TI^{inh}_{s+1} \leqslant x. \end{cases}
\end{aligned} \tag{4.18}$$

We now show that the functions $\partial\, G_{s+1}(x{+}K,\,x,\,\mathbf{d})/\partial\, y$ and $\partial\, G_{s+1}(x,\,x,\,\mathbf{d})/\partial\, y$ are higher than $\partial G_N(y,x,\mathbf{d})/\partial y$ (i.e., tighter lower bounds than ∂G_N). Observe that $\partial\, V_{s+2}(\,\cdot\,-\sum_b D^b_{s+1,s+1},\,\mathbf{D_{s+2}})/\partial\, y$ vanishes for the first and the third regions. They account for the cases where the critical order-up-to levels TI^{sub}_{s+1} and TI^{inh}_{s+1}, respectively, are attainable in period $s+1$ (i.e., the inventory position before ordering, x, is no greater than TI^{sub}_{s+1} for the first region and than TI^{inh}_{s+1} for the second). This means that the cost of periods $s+2$ and beyond is insensitive to the decision made in period s (i.e., y) for those order realizations $\sum_b D^b_{s,s}$ that make $x = y - \sum_b D^b_{s,s}$ not larger than the respective critical order-up-to level, namely TI^{sub} or TI^{inh}.

In the second region, on the other hand, the TI^{inh}_{s+1} level in period $s+1$ is not attainable since the use of the full in-house capacity sets inventory position to the point $x+K$. Hence the decision made in period s; i.e., y, does affect the cost of period $s+1$ and $\partial\, G_{s+1}(x+K,\,x,\,\mathbf{d})/\partial\, y$ denotes the corresponding effect on the cost in the future periods. We have similar arguments for the fourth region.

Let $\mathcal{I}^i_{s+1}$ for $i = 1,\,2,\,3,\,4$ be the event that inventory position before ordering in period $s+1$, x, falls in the ith region in (4.18) after the inventory position has been set to y in period s. We number the regions in an increasing order of the break points, where the first region, $(-\infty,\,TI^{sub}_{s+1} - K]$, is given the number $i = 1$. By inserting (4.18) into $E_{\boldsymbol{\mathcal{E}}_{\mathbf{s}}}[\partial\, V_{s+1}(y - \sum_b D^b_{s,s},\,\mathbf{d})/\partial\, y]$,

$$
\begin{aligned}
E_{\boldsymbol{\mathcal{E}_s}}[\partial V_{s+1}(x,\,\mathbf{d})/\partial y] \;=\;& E_{\boldsymbol{\mathcal{E}_s}}[\;-\mathbf{1}(\mathcal{I}^1_{s+1})\,c_{ps} \\
&+\mathbf{1}(\mathcal{I}^2_{s+1})\,\partial G_{s+1}(x+K,\,x,\,\mathbf{d})/\partial y \\
&-\mathbf{1}(\mathcal{I}^3_{s+1})\,c_{pi} \\
&+\mathbf{1}(\mathcal{I}^4_{s+1})\,\partial G_{s+1}(x,\,x,\,\mathbf{d})/\partial y\;\;], \qquad (4.19)
\end{aligned}
$$

where $\mathbf{1}(.)$ is the indicator function. The structure of (4.19) does not allow any closed-form solutions to the first-order condition $G_s(y,\,x,\,\mathbf{d_s}) = 0$.

The occurrence probability of the event that inventory position before ordering in period $s+1$, $x = y - \sum_b D^b_{s,s}$, falls in the ith region in (4.18), $i = 1, 2, 3, 4$, after the inventory position has been set to y in period s is given by

$$
\begin{aligned}
Pr\{\mathbf{1}(\mathcal{I}^1_{s+1})\} &= Pr\{\sum_b D^b_{s,s} \geqslant y - TI^{sub}_{s+1} + K\} \\
&= Pr\{Z_{[s,s+1)} \geqslant y - TI^{sub}_{s+1} + K\} \\
&= \int_{y-TI^{sub}_{s+1}+K}^{\infty} f_{Z_1}(\cdot\,|\,\mathbf{d_s}) \\
Pr\{\mathbf{1}(\mathcal{I}^2_{s+1})\} &= \int_{y+K-TI^{inh}_{s+1}}^{y+K-TI^{sub}_{s+1}} f_{Z_1}(\cdot\,|\,\mathbf{d_s}) \\
Pr\{\mathbf{1}(\mathcal{I}^3_{s+1})\} &= \int_{y-TI^{inh}_{s+1}}^{y+K-TI^{inh}_{s+1}} f_{Z_1}(\cdot\,|\,\mathbf{d_s}) \\
Pr\{\mathbf{1}(\mathcal{I}^4_{s+1})\} &= \int_{0}^{y-TI^{inh}_{s+1}} f_{Z_1}(\cdot\,|\,\mathbf{d_s}). \qquad (4.20)
\end{aligned}
$$

Likewise, the joint occurrence probability of the events $\mathcal{I}^4_{s+1}$ and $\mathcal{I}^4_{s+2}$, for example, can be given by

$$
\begin{aligned}
Pr\{\mathbf{1}(\mathcal{I}^4_{s+1}\mathcal{I}^4_{s+2})\} &= Pr\{\mathbf{1}(\mathcal{I}^4_{s+2})\;\mid\;\mathbf{1}(\mathcal{I}^4_{s+1})\} \\
&= Pr\{Z_{[s,s+2)} \leqslant TI^{inh}_{s+2} - y \;\mid\; Z_{[s,s+1)} \leqslant TI^{inh}_{s+1} - y\} \\
&= \int_0^{TI^{inh}_{s+1}-y}\int_0^{TI^{inh}_{s+2}-y-z_1} f_{\mathbf{D}}(\cdot,\cdot\,|\,\mathbf{d_s}) \;/\; \int_0^{TI^{inh}_{s+1}-y} f_{Z_1}(\cdot\,|\,\mathbf{d_s}). \qquad (4.21)
\end{aligned}
$$

The conditional probabilities for other joint event occurrences can be written similarly.

Joint Occurrence of Events for Inventory Position

In a similar manner, we may further decompose the $\partial G_{s+1}(x+K,\,x,\,\mathbf{d})/\partial y$ and

$\partial G_{s+1}(x, x, \mathbf{d})/\partial y$ terms. For each of these expressions, we again have four possible ranges on the value of inventory position before ordering in period $s+2$ (i.e., I_{s+2}). This implies that total enumeration of all possible cases increases with combinations of ranges from $(s+1)$ and $(s+2)$ for periods $s+2$ and beyond. Nevertheless, applying Theorem 4.2 to these terms, we can eliminate some cases from further consideration.

To illustrate, suppose $\mathbf{1}(\mathcal{I}^2_{s+1}) = 1$; that is, the inventory position before ordering in period $s+1$ ($x = y - \sum_b D^b_{s,s}$) falls in the second region given the decision y in period s. Then the conditional probability that the inventory position before ordering in period $s+2$ falls in the first region, given that $\mathbf{1}(\mathcal{I}^2_{s+1}) = 1$, is equal to zero. This follows from the relative values of K, $\sum_b D^b_{s+1,s+1}$ and $TI^{sub}_{s+2} - TI^{sub}_{s+1}$ since the value of the difference $TI^{sub}_{s+2} - TI^{sub}_{s+1}$ depends on the value of $\sum_b D^b_{s+1,s+1}$. We can apply similar argument to the fourth region. Thus we have

$$\begin{aligned}
Pr\{x + K - \textstyle\sum_b D^b_{s+1,s+1} \in (-\infty, TI^{sub}_{s+2} - K] \mid \mathbf{1}(\mathcal{I}^2_{s+1}) = 1\} &= 0 \\
Pr\{x - \textstyle\sum_b D^b_{s+1,s+1} \in (-\infty, TI^{sub}_{s+2} - K] \mid \mathbf{1}(\mathcal{I}^4_{s+1}) = 1\} &= 0 \\
Pr\{x - \textstyle\sum_b D^b_{s+1,s+1} \in [TI^{sub}_{s+2} - K, TI^{inh}_{s+2} - K] \mid \mathbf{1}(\mathcal{I}^4_{s+1}) = 1\} &= 0 \\
Pr\{x - \textstyle\sum_b D^b_{s+1,s+1} \in [TI^{inh}_{s+2} - K, TI^{inh}_{s+2}] \mid \mathbf{1}(\mathcal{I}^4_{s+1}) = 1\} &= 0,
\end{aligned}$$

restricting the potential state variables in stage $s+1$. We have non-zero conditional probabilities for all the other four cases. Consequently, using these results, we may state (4.19) as

$$\begin{aligned}
\mathrm{E}_{\boldsymbol{\mathcal{E}_s}}[\,\partial V_{s+1}(y - Z_{[s,s+1)}, \mathbf{D_{s+1}})/\partial y\,] &= \\
&- c_{ps}\, Pr\{\mathbf{1}(\mathcal{I}^1_{s+1})\} \;-\; c_{pi}\, Pr\{\mathbf{1}(\mathcal{I}^3_{s+1})\} \;-\; c_{pi} \sum_{j=2}^{N-s} Pr\{\mathbf{1}(\mathcal{I}^2_{s+2} \ldots \mathcal{I}^2_{s+j})\} \\
&+ \sum_{j=1}^{N-s} Pr\{\mathbf{1}(\mathcal{I}^2_{s+1} \ldots \mathcal{I}^2_{s+j})\}\, \mathrm{E}\Big[\partial J_{s+j}(y + jK - Z_{[s,s+j)}, I_{s+j}, \mathbf{D_{s+j}})/\partial y\Big] \\
&+ \sum_{j=1}^{N-s} \sum_{k=j+1}^{N-s} Pr\{\mathbf{1}(\mathcal{I}^2_{s+1} \ldots \mathcal{I}^2_{s+k}\mathcal{I}^4_{s+k+1} \ldots \mathcal{I}^4_{s+j})\} \\
&\qquad\qquad \mathrm{E}\Big[-c_{ps} + \partial J_{s+j}(y + kK - Z_{[s,s+j)}, I_{s+j}, \mathbf{D_{s+j}})/\partial y\Big] \\
&+ \sum_{j=1}^{N-s} Pr\{\mathbf{1}(\mathcal{I}^4_{s+1} \ldots \mathcal{I}^4_{s+j})\}\, \mathrm{E}\Big[[-c_{ps} + \partial J_{s+j}(y - Z_{[s,s+j)}, I_{s+j}, \mathbf{D_{s+j}})/\partial y\,]\Big],
\end{aligned} \tag{4.22}$$

where $\partial J_{s+j}(\,\cdot\,,\, I_{s+j},\, \mathbf{D_{s+j}})/\partial y$ is given by

$$\begin{aligned}\partial J_{s+j}(\,\cdot\,,\, I_{s+j},\, \mathbf{D_{s+j}})/\partial y \;\; &= \\ & c_{pi} + (c_{ps} - c_{pi})\mathbf{1}(A_{s+j}) \;+\; h\, F_{Z_L}(\,\cdot \mid \mathbf{D_{s+j}}) \\ & - \pi_1 \left[1 - F_{Z_L^1}(\,\cdot \mid \mathbf{D_{s+j}})\right] \;-\; \pi_2 \left[F_{Z_L^1}(\,\cdot \mid \mathbf{D_{s+j}}) - F_{Z_L}(\,\cdot \mid \mathbf{D_{s+j}})\right], \quad (4.23)\end{aligned}$$

and $Pr\{\mathbf{1}(\mathcal{I}_a\mathcal{I}_b\ldots\mathcal{I}_r)\}$ [cf. Eqs. (4.20) and (4.21)] states the joint occurrence probability of the events a to r. The probabilities for joint occurrences of the events in (4.22) can readily be evaluated using the finite Markov chain structure introduced in §3.2 to §3.4. Thus, it is trivial to conclude that the expression (4.22) results in a higher bounding function than $\partial G_N(y, x, \mathbf{d})/\partial y$ can [cf. Eq. (4.15)]. The solution to (4.22) then gives a tighter upper bound (than the relaxed upper bound in (4.17)) for the optimal order-up-to level of period s.

CHAPTER 5

AN EFFICIENT APPLICATION OF STOCHASTIC DYNAMIC PROGRAMMING

In this chapter, we suggest an efficient application of stochastic dynamic programming to the multi-period stochastic decision model under study. In §5.1, we discuss the stochastic state transitions and the associated dimensionality problem. In §5.2, we present how we will quantify the probabilities involved in the stochastic dynamic recursions. In §5.3, we introduce a reasonably accurate way of reducing state dimensionality, aiming to circumvent the associated computational and storage requirements encountered in solving the dynamic program. Finally, an example computational process is given in §5.4. We also discuss how to validate our way of reducing state dimensionality.

5.1 Stochastic State Transitions

The multi-period stochastic production/inventory problem under study is a Markov decision process where the replenishment decision made in any period depends on the current state of the system, and transitions to next-period state are independent of all previously-visited states. The problem has the properties needed to be structured as an equivalent stochastic dynamic programming problem whose optimal solution provides an optimal replenishment policy for the original problem.

The focal point that one associates with a dynamic programming problem is its state transition function. We shall now elaborate on state transitions that occur in our production/inventory system. We have a finite problem horizon involving N decision periods where the manufacturer observes the state of the system at the beginning of period $s = 1, 2, \ldots, N$ as $(I_s, \mathbf{d_s})$ where

- I_s is one-dimensional inventory state representing inventory position before the manufacturer's replenishment decision in period s. It is one-dimensional since the manufacturer does not keep track of separate inventory pools for buyers.

- $\mathbf{d_s} = \left[\mathbf{d_s^b},\, \forall\, b \in \{1, 2, \ldots, B\}\right]$ is BH-dimensional commitment state (where B is the number of buyers involved and H is the commitment horizon. $H \leqslant L$ where L is the manufacturer's replenishment lead-time). The vector $\mathbf{d_s^b} = \left[d^b_{s-1,s}, \ldots, d^b_{s-1,s+H-1}, \mu_{D^b}, \mu_{D^b}, \ldots\right]$ denotes random order commitments from buyer b available to the manufacturer at the beginning of period s [1]. Note that $\mathbf{d_s}$ denotes observed values (realizations) of $\mathbf{D_s^b}$.

We would like to select an ordered set of replenishment decisions which will minimize the total cost incurred for each possible initial state $(I_1, \mathbf{d_1})$. Such a minimum cost thus can be computed, in principle, using the stochastic dynamic programming, which has proven to be of great utility in the solution of inventory problems. Consequently, we have the following recursive functional equations, for $s = N, N-1, \ldots, 2, 1$,

$$\begin{aligned} V_s(I_s, \mathbf{d_s}) &= \min_{TI_s \geqslant I_s} \left\{ J_s(TI_s, I_s, \mathbf{d_s}) + E_{\boldsymbol{\mathcal{E}_s}}[V_{s+1}(TI_s - \sum_b D^b_{s,s}, \mathbf{D_{s+1}})] \right\}, \\ V_{N+1}(I_{N+1}, \mathbf{d_{N+1}}) &= -\,c_{ps} E_{\boldsymbol{\mathcal{E}_{N+1}}, \boldsymbol{\mathcal{E}_{N+2}}, \ldots, \boldsymbol{\mathcal{E}_{N+L}}}\left[NI_{N+L} \right], \end{aligned} \quad (5.1)$$

where $J_s(TI_s, I_s, \mathbf{d_s})$, which is given by (4.7), is the current-period cost associated with state $(I_s, \mathbf{d_s})$ when replenishment decision TI_s is selected. $V_s(I_s, \mathbf{d_s})$ is the minimum expected cost-to-go from the beginning of period s to the end of period $N + L$, assuming that the system is in state $(I_s, \mathbf{d_s})$ in period s and an optimal replenishment decision is made in every period $s, s + 1, \ldots, N$.

The dynamic programming recursions in (5.1) consist of stochastic state transitions in the sense that the current state vector $(I_s, \mathbf{d_s})$ and a replenishment decision to result in TI_s in period s are not adequate to determine the next-period state with certainty. Instead, we have a collection of possible next-period states as a random variable dependent on the choice of TI_s, and state transitions from one period to the next is described by a probability distribution. Thus we have the random next-period

[1] Note that for periods beyond the commitment horizon $H \leqslant L$ order commitment is assumed to be the mean realized order, which are all identical and given by μ_{D^b}.

state denoted by $(TI_s - \sum_b D^b_{s,s},\ \mathbf{D_{s+1}})$, where $TI_s - \sum_b D^b_{s,s}$ is the random next-period inventory state and $\mathbf{D_{s+1}}$ is random commitment state vector that will be observed at the beginning of the next period.

The forecast evolution model that we have adopted enables us to express the source of randomness here in terms of the random commitment update vector, $\boldsymbol{\mathcal{E}_s} = [\,\boldsymbol{\mathcal{E}^b_s},\ \forall\, b \in \{1,\, 2,\, \ldots,\, B\}\,]$ where

$$\boldsymbol{\mathcal{E}^b_s} = \left[\varepsilon^b_{s,1},\, \varepsilon^b_{s,2},\, \ldots,\, \varepsilon^b_{s,H+1},\, 0,\, 0,\, \ldots\right],$$

where $\varepsilon^b_{s,k}$ is the random variable corresponding to the k-period ahead commitment update (on the logarithmic scale) from buyer b for the amount to be ordered in period $s + k$, $k = 1,\, 2,\, \ldots,\, H + 1$. Thus, the expectation $E_{\boldsymbol{\mathcal{E}_s}}[\,\cdot\,]$ in (5.1) is with respect to all possible realizations of $\boldsymbol{\mathcal{E}_s}$ and is conditioned on the value of $\mathbf{d_s}$.

An illustrative example shown in Figure 5.1 helps to clarify the nature of the associated computational process. Suppose we will follow value iteration using backward induction algorithm as a solution technique for the dynamic programming recursions in (5.1). Consider an $(N - s)$-period problem extending from the beginning of period s. Suppose we would like to compute $V_s(I_s,\, \mathbf{d_s})$, the minimum expected cost-to-go from period s for a specific choice $(I_s,\, \mathbf{d_s})$ of the current state. Suppose at the beginning of period s a certain replenishment decision aiming at a specific TI_s is made. We first evaluate the current costs associated with period s for this choice of TI_s; that is, $J_s(TI_s,\, I_s,\, \mathbf{d_s})$, considering all possible realizations $(z^1_L,\, \ldots,\, z^B_L)$ of the random variables $(Z^1_{[s,s+L)},\, \ldots,\, Z^B_{[s,s+L)})$ (sums of L realized orders) [①]. Each of the resulting costs $J_s(TI_s,\, I_s,\, \mathbf{d_s} \,|\, z^1_L,\, \ldots,\, z^B_L)$ is weighted by the probability $Pr\{z^1_L,\, \ldots,\, z^B_L\}$ of obtaining that particular realization of $(Z^1_{[s,s+L)},\, \ldots,\, Z^B_{[s,s+L)})$. We then make a certain transition from the state $(I_s,\, \mathbf{d_s})$ to a next-period state $(TI_s - \sum_b d^b_{s,s},\, \mathbf{d_{s+1}})$ with probabilities $Pr\{\boldsymbol{\mathcal{E}^1_s},\, \ldots,\, \boldsymbol{\mathcal{E}^B_s}\}$ [②]. We draw the minimum expected cost-to-go $V_{s+1}(TI_s - \sum_b d^b_{s,s},\, \mathbf{d_{s+1}} \,|\, \boldsymbol{\mathcal{E}^1_s},\, \ldots,\, \boldsymbol{\mathcal{E}^B_s})$ associated with this particular next-period state from the list of $V_{s+1}(I_{s+1},\, \mathbf{D_{s+1}})$ values, which have been already evaluated by the previous backward iteration for all possible realizations of the random next-period state $(I_{s+1},\, \mathbf{D_{s+1}})$ [③]. We can similarly obtain all the remaining minimum expected cost-to-go values associated with the other transitions to the next-period state. As a consequence, we continue repeating this entire calculation for every allowable choice of replenishment decision TI_s, and we can then choose the minimum among them, which becomes $V_s(I_s,\, \mathbf{d_s})$ [④].

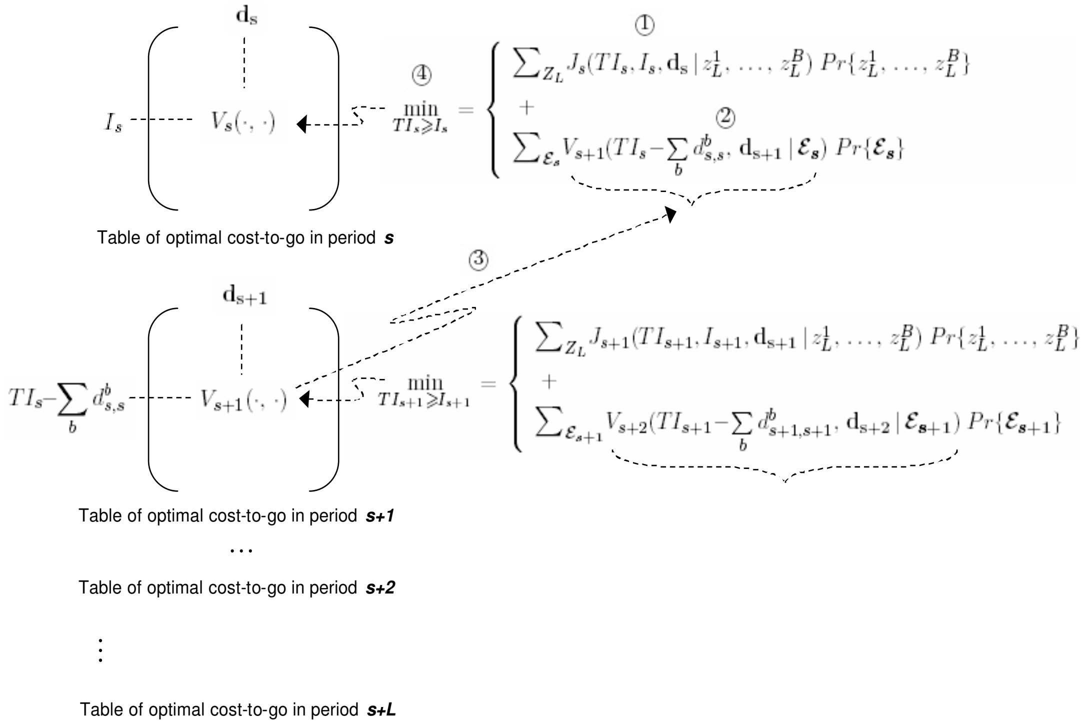


Figure 5.1: Computing $V_s(I_s,\, \mathbf{d_s})$

A pseudo-code for this application of the dynamic programming is given in Figure 5.2. We number the lines therein in an increasing order, where the lines at the same level are assigned the same number. This pseudo-code begins with $(B+3)$ nested loops; (i) one for decision periods s in the planning horizon [*line 1*], (ii) one for the set of possible inventory states I_s [*line 2*], (iii) B for the sets of possible commitment state vectors $\mathbf{d_s}$ across buyers each being H-dimensional [*lines 3 to 5*], and (iv) the last one for the set of allowable replenishment decisions TI_s [*line 6*]. Next, we have two successive blocks of loops each having B additional nested loops. The first block [*the first loop at line 7*] is for evaluating the expected costs of carrying inventory and backorders where the expectations are taken with respect to all possible realizations of $Z_{[s,s+L)}$, conditioned on the value of $\mathbf{d_s}$. The second is for moving to the random next-period state and taking the future cost out of a list of previously evaluated minimum costs [*the second loop at line 7*]. The lowest level to which the nested loops extend is $2B + 3$ where each commitment state vector has H elements, resulting in high dimensionality in the state space.

---

1. **For** period s backwards from $s = N$ to $s = 1$
2. **For** state I_s values
3. **For** state $\mathbf{d_s^1} = [d^1_{s-1,s}, d^1_{s-1,s+1}, \ldots, d^1_{s-1,s+L-1}]$ values from buyer 1
4. $\ldots$
5. **For** state $\mathbf{d_s^B} = [d^B_{s-1,s}, d^B_{s-1,s+1}, \ldots, d^B_{s-1,s+L-1}]$ values from buyer B
6. **For** decision TI_s

7. Compute $PC_s(TI_s,\, I_s)$
7. **For** realizations of $Z^1_{[s,s+L)}$ from buyer 1
8. $\ldots$
9. **For** realizations of $Z^B_{[s,s+L)}$ from buyer B
10. Find probability $Pr\{z^1_L,\, \ldots,\, z^B_L\}$ for these realizations
10. Compute $J_s(TI_s,\, I_s,\, \mathbf{d_s} \mid z^1_L,\, \ldots,\, z^B_L)$
9. **End for**
8. $\ldots$
7. **End for**
7. Compute $J_s(TI_s,\, I_s,\, \mathbf{d_s})$ over all realizations

7. **For** state $\mathbf{D^1_{s+1}} = [D^1_{s,s}, D^1_{s,s+1}, \ldots, D^1_{s,s+L}]$ values from buyer 1
8. $\ldots$
9. **For** state $\mathbf{D^B_{s+1}} = [D^B_{s,s}, D^B_{s,s+1}, \ldots, D^B_{s,s+L}]$ values from buyer B
10. Compute $TI_s - \sum_b D^b_{s,s}$, next-period inventory state
10. Find joint probability $Pr\{\boldsymbol{\mathcal{E}_s^1},\, \ldots,\, \boldsymbol{\mathcal{E}_s^B}\}$ for next-period states
10. Get $V_{s+1}(TI_s - \sum_b D^b_{s,s},\, \mathbf{D_{s+1}} \mid \boldsymbol{\mathcal{E}_s^1},\, \ldots,\, \boldsymbol{\mathcal{E}_s^B})$ from a list of values
9. **End for**
8. $\ldots$
7. **End for**
7. Compute expected cost-to-go $E_{\boldsymbol{\mathcal{E}_s}}[\,V_{s+1}(TI_s - \sum_b D^b_{s,s},\, \mathbf{D_{s+1}})\,]$
7. Compute total cost $G_s(TI_s,\, I_s,\, \mathbf{d_s}) = J_s(\,\cdot,\, \cdot,\, \cdot) + E_{\boldsymbol{\mathcal{E}_s}}[\,V_{s+1}(\,\cdot,\, \cdot)\,]$ for TI_s

6. **End for**
6. Find the minimum cost, $V_s(I_s,\, \mathbf{d_s})$ over all TI_s
5. **End for**
4. $\ldots$
3. **End for**
2. **End for**
1. **End for**

Figure 5.2: A pseudo-code of dynamic recursions (5.1)

5.2 Probabilities in the Stochastic Dynamic Programming

The dynamic programming recursive relations in (5.1) involve two types of probability distributions, one is for the expected L-period costs of carrying inventory and backorders and the other is for stochastic state transitions. This section discusses how we determine these probabilities.

In the multi-period stochastic decision model under study, the current order commitment is an explicit component of the state space that evolves from one period to the next. The random component of the decision making environment should be described by this evolution. Thus, the probabilities governing the stochastic state transitions for the system under the forecast evolution model are quantified by the distribution of commitment updates. This corresponds to the multivariate distribution $f_{\boldsymbol{\mathcal{E}}_s}(\cdot\,|\,\boldsymbol{\mathcal{W}},\,\boldsymbol{\mathcal{A}})$ of the commitment update vector $\boldsymbol{\mathcal{E}}_s = \left[\boldsymbol{\mathcal{E}}_s^b,\ \forall\, b \in \{1,\, 2,\, \ldots,\, B\}\right]$ (on the logarithmic scale), comprising $B(H+1)$ interval censored random variables, $\varepsilon_{s,k}^b$'s, as mentioned before in §3.1. As true mathematical form for this multivariate distribution is complicated, we have suggested a finite Markov chain approximation in §3.2 to §3.4.

The other probability distribution involved in (5.1) should quantify the uncertainty associated with the sum of L realized orders in the future. This necessitates the distribution $f_{Z_L^b}(\cdot\,|\,\mathbf{d_s})$ of $Z_{[s,s+L)}^b = \sum_{k=1}^{L} D_{s+k-1,s+k-1}^b$, conditioned on the value of $\mathbf{d_s}$. Evaluating the distribution $f_{Z_L^b}(\cdot\,|\,\mathbf{d_s})$ accurately is difficult because of the censored nature of the underlying random variables and the correlations among $D_{s+k-1,s+k-1}^b$'s.

The following proposition is fundamental in overcoming this difficulty. It essentially structures $f_{Z_L^b}(\cdot\,|\,\mathbf{d_s})$ into an (approximately) equivalent probability distribution function, which is readily amenable to being estimated by means of the Markov chain approximation introduced in §3.2 to §3.4. Note that the proposition assumes, for the sake of clarity, that $L = H$ without loss of generality. Otherwise, we would need to decompose $Z_{[s,s+L)}^b$ into two components, one for the interval $[s, s+H)$ and the other for $[s+H, s+L)$.

Proposition 5.1 *For a given value of the order commitment vector* $\mathbf{d_s}$ *at the beginning of period* s*, the probability distribution* $f_{Z_L^b}(\cdot\,|\,\mathbf{d_s})$ *can be reduced to the following alternative form.*

$$
\begin{aligned}
f_{Z_L^b}(y \,|\, \mathbf{d_s}) &= Pr\{\ Z^b_{[s,s+L)} \ = \ y \,|\, \mathbf{d_s}\ \} \\
&\simeq Pr\{ \sum_{k=1}^{L} \lambda_k^b \varepsilon_{s,k}^b \ = \ \ln \frac{y}{\sum\limits_{k=1}^{L} \beta_k^b} + \sum_{k=1}^{L} \theta_k^b \ \ln \frac{\beta_k^b}{d^b_{s-1,s+k-1}} \ \} \\
&= f_{\vartheta_L^b}(\ \frac{\ln y}{\sum\limits_{k=1}^{L} \beta_k^b} + \sum_{k=1}^{L} \theta_k^b \ \ln \frac{\beta_k^b}{d^b_{s-1,s+k-1}} \ ),
\end{aligned} \tag{5.2}
$$

where

(*i*) β_k^b is called *the expected order after k updates from buyer b* (i.e., single period order k periods into the commitment horizon),

$$
\beta_k^b = \mu_{D^b} \ e^{(\mu_{\varepsilon_1^b} + \mu_{\varepsilon_2^b} + \cdots + \mu_{\varepsilon_k^b})}. \tag{5.3}
$$

β_k^b has the same units as $Z^b_{[s,s+L)}$. It does not depend on s since μ_{D^b} is assumed to be stationary.

(*ii*) θ_k^b is called *the proportional expected order after k updates.* It is the proportion of β_k^b in the expected aggregate L-period buyer order,

$$
\theta_k^b = \beta_k^b / \sum_{m=1}^{L} \beta_m^b. \tag{5.4}
$$

θ_k^b is dimensionless.

(*iii*) λ_k^b is called *the weight factor of k-step commitment update* $\varepsilon_{s,k}^b$,

$$
\lambda_k^b = \sum_{m=k}^{L} \theta_m^b. \tag{5.5}
$$

$\lambda_k^b \in (0,1]$ is dimensionless. $\lambda_1^b = 1$ and it decreases as k goes from 1 to L. Weight factors increase towards the period of order realization (i.e., as the period moves to L-periods into the commitment horizon). Note that the sequence of λ_L^b, λ_{L-1}^b, $\ldots$, λ_1^b is independent of s and $\{\boldsymbol{\mathcal{E}_s},\ s \geqslant 1\}$. Hence each weight is a constant.

(*iv*) ϑ_L^b is called *the weighted aggregate L-period commitment update.* It denotes a weighted sum of commitment updates (on the logarithmic scale) to be received from buyer b over the time interval $[s,\ s+L)$,

$$
\vartheta_L^b = \sum_{k=1}^{L} \lambda_k^b \varepsilon_{s,k}^b. \tag{5.6}
$$

ϑ^b_L has the units of ln $Z^b_{[s,s+L)}$ (similar to $\varepsilon^b_{s,k}$). Note that ϑ^b_L does not depend on s since the vector sequence $\{\boldsymbol{\mathcal{E}_s},\ s \geqslant 1\}$ forms a stationary stochastic process due to Assumption 3.3 of the forecast evolution structure.

Proof: See Appendix A.5.

□

Proposition 5.1 transforms the sum of L realized orders over the time interval $[s,\ s+L)$ onto a weighted sum of logarithmic commitment updates. If it is the case when $L = 1$, the first two lines in (5.2) represent the probabilities of two equivalent events. Logarithmic transformations therein are used to replace difficult operations on the products of censored random variables by easier operations on the weighted sum of their logarithms. This allows us to operate with a simpler form of $f_{Z^b_L}(\cdot\,|\,\mathbf{d_s})$. The only approximation in (5.2) applies to the logarithm of a sum of random variables employing the linear Taylor series expansion.

We are now interested in the chance of realizing particular values for ϑ^b_L in order to approximate the discrete probabilities for specific $Z^b_{[s,s+L)}$ occurrences. The form of ϑ^b_L permits quantifying its probability function $f_{\vartheta^b_L}(\cdot\,)$ by means of the finite Markov chain approximation introduced in §3.2 to §3.4. This will be a straightforward extension generalizing the empirical approach to the weighted sums of commitment updates, since each weight λ^b_k is a given constant.

The Markovian representation of §3.3 translates into the multivariate process $\{(\vartheta^1_k,\ \ldots,\ \vartheta^B_k),\ 1 \leqslant k \leqslant L\}$. Hence, allowable range of commitment updates by the flexibility limits, their expected values, and covariances all need to be converted to incorporate this weighting structure. At the end, we will obtain the joint probability mass function of $(\vartheta^1_L, \ldots, \vartheta^B_L)$ and hence the marginal distributions, $f_{\vartheta^b_L}(\cdot\,)$'s in Proposition 5.1.

For the important special case where $L = H = 1$ (this is a simple model where each buyer submits one order commitment in every period under the one-period replenishment lead-time), we have $Z^b_{[s,s+1)} = D^b_{s,s}$ and hence $\vartheta^b_1 = \varepsilon^b_{s,1}$. The expression (5.2) then turns out to be

$$
\begin{aligned}
f_{Z_1^b}(y \,|\, \mathbf{d_s}) &= Pr\{\, Z^b_{[s,s+1)} \;=\; y \,|\, \mathbf{d_s} \,\} \\
&= Pr\{\, \lambda_1^b \varepsilon_1^b \;=\; \ln \frac{y}{\beta_1^b} + \theta_1^b \,\ln\, \frac{\beta_1^b}{d^b_{s-1,s}} \,\} \\
&= Pr\{\, \varepsilon_1^b \;=\; \ln y - \ln d^b_{s-1,s} \,\} \qquad \text{since } \theta_1^b = \lambda_1^b = 1 \\
&= f_{\varepsilon_1^b}(\, \ln \frac{y}{d^b_{s-1,s}} \,|\, \boldsymbol{\mathcal{W}},\, \boldsymbol{\mathcal{A}}). \qquad (5.7)
\end{aligned}
$$

It is important to mention that $\{Z^b_{[s,s+1)} = y \,|\, \mathbf{d_s}\}$ and $\{\varepsilon_1^b = \ln y - \ln d^b_{s-1,s}\}$ in (5.7) are equivalent events, not equivalent quantities. The transformation is exact in this case, and the probabilities $f_{Z_1^b}(y \,|\, \mathbf{d_s})$ are readily available as the probabilities $f_{\varepsilon_1^b}(\cdot \,|\, \boldsymbol{\mathcal{W}},\, \boldsymbol{\mathcal{A}})$ of one-period ahead commitment updates ε_1^b. Note that the precision of the approximation in (5.2) decreases as L increases (or as H increases for $H < L$ with a fixed L).

5.3 An Approach for Reducing the State Dimensionality

A direct application of the dynamic programming to problems for which stochastic state variable is multidimensional is quite challenging due to the resulting dimensionality of the state space. The dynamic programming recursive relations in (5.1) require $BH + 1$ state variables at each period, one for inventory state and BH for the order commitments received from B buyers for the following H periods. This implies that an application of the dynamic programming will bring about impractical computational and storage requirements as B and/or H increase. We have approximated the continuous space by a finite number of possible discrete states that the stochastic process may realize, as mentioned before in §3.3. The number of disjoint sub-intervals in discretization was arbitrarily specified as M. The discretization will cause yet another level of combinatorial explosion. Figure 5.3 illustrates a typical enumeration of possible discrete values that a certain commitment state (out of BH commitment states) can take over the H-period horizon. Note that M^H distinct discrete values are possible even for a single commitment state.

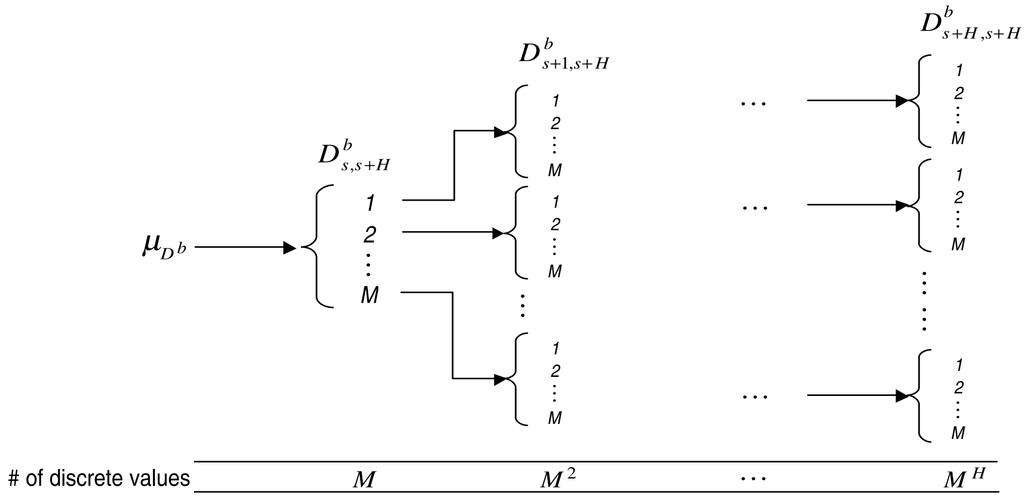


Figure 5.3: Enumeration of possible discrete values for a single commitment state

This curse of dimensionality limits the applicability of the dynamic programming methodology. In this section we suggest a method to circumvent the time and memory requirements by reducing the high dimensionality in the state variables, particularly in BH-dimensional commitment states. Reducing the number of states corresponds to decreasing the number of transitions of order commitments. The main advantage of the reduced problem will be that it substantially reduces the number of states necessary to find an optimal solution and, thus, makes it possible to solve much larger problems.

Define $\mathfrak{D}_{s,H}$ to be a statistic computed from a possible realization $\mathbf{d_s}$ of the order commitment vector to be available in period s,

$$\mathfrak{D}_{s,H} \;=\; \sum_b \sum_{k=1}^{H} \theta^b_k \ln d^b_{s-1,s+k-1}, \tag{5.8}$$

where $\theta^b_k = \beta^b_k / \sum_{m=1}^{L} \beta^b_m$ and $\beta^b_k = \mu_{D^b}\; e^{(\mu_{\varepsilon^b_1}+\mu_{\varepsilon^b_2}+\cdots+\mu_{\varepsilon^b_k})}$, as given by (5.4) and (5.3), respectively. This statistic is an *indicator of weighted effects* by possible realizations of order commitments. In the calculation of $\mathfrak{D}_{s,H}$, the formula consists of summing the logarithms of all the order commitments, each being weighted by a known constant θ^b_k. When the mean realized order is stationary over time we have $\theta^b_k = 1/L$, otherwise the weights θ^b_k imply relatively larger weight for higher expected updates. $\mathfrak{D}_{s,H}$ relates to the structure in Proposition 5.1. Both are in logarithm of order units, both are θ

weighted, both are accumulation over the commitment horizon.

Consider all possible realizations (discrete values) of a random commitment vector $\mathbf{D_s}$ in period s. We first compute $\mathfrak{D}_{s,H}$ statistic for each of these discrete vector values. Those $\mathfrak{D}_{s,H}$ values that are sufficiently close to each other (at the analyst's discretion) are unified as one value. The corresponding commitment realizations $\mathbf{d_s}$ can then be partitioned into clusters. We refer to such clusters as *commitment clusters.* This allows us to state the following fundamental theorem, suggesting our way of reducing high dimensionality in state variables. The theorem says that for a given value of inventory state I_s at the beginning of period s, all possible commitment state $\mathbf{d_s}$ realizations partitioned into the same commitment cluster will result in identical replenishment decision.

Theorem 5.1 *Suppose a certain system state* $(I_s,\, \mathbf{d_s})$ *is observed at the beginning of period* s. *The following properties hold.*

(*i*) *All possible commitment state realizations* $\mathbf{d_s}$ *that can be partitioned into the same commitment cluster, by the statistic* $\mathfrak{D}_{s,H}$, *yield the same optimal replenishment decision.*

(*ii*) *For the purposes of determining optimal replenishment policy of each commitment cluster, the minimum expected cost-to-go from period* s, $V_s(I_s,\, \mathbf{d_s})$ *can be replaced by the modified cost-to-go* $\widehat{V}_s(\widehat{I}_s,\, \mathfrak{D}_{s,H})$, *where*

$$\widehat{I}_s \triangleq I_s - \sum_b \left[(1 - \sum_{k=1}^{L} \theta_k^b \ln \frac{\beta_k^b}{d^b_{s-1,s+k-1}}) \sum_{k=1}^{L} \beta_k^b \right], \tag{5.9}$$

$$\mathfrak{D}_{s,H} = \sum_b \sum_{k=1}^{H} \theta_k^b \ln d^b_{s-1,s+k-1}, \tag{5.10}$$

where β_k^b *and* θ_k^b *are given by (5.3) and (5.4), respectively.* $(\widehat{I}_s,\, \mathfrak{D}_{s,H})$ *is called the modified system state.*

Proof: See Appendix A.6.

□

Theorem 5.1 reveals that, upon observing the system state $(I_s,\, \mathbf{d_s})$ at the beginning of period s, we can eliminate the remaining commitment states in the same commitment cluster with $\mathbf{d_s}$ from further consideration. In turn, this enables us to consider the problem in terms of the modified cost-to-go function $\widehat{V}_s(\cdot)$, which is a function of the *modified inventory position* $\widehat{I}_s$ and the $\mathfrak{D}_{s,H}$ statistic. $\widehat{I}_s$ has the same units as I_s, and $\mathfrak{D}_{s,H}$ is in logarithm of order units. Thus the state space dimensionality is reduced in the sense that we have a calculation in only two-dimensional state variable $(\widehat{I}_s,\, \mathfrak{D}_{s,H})$ with Theorem 5.1, instead of $(BH+1)$-dimensional state variable $(I_s,\, \mathbf{d_s})$.

The modified inventory position $\widehat{I}_s$ in (5.9) decreases the observed inventory state I_s by an amount equal to the expected aggregate L-period buyer order, $\sum\limits_b \sum\limits_{k=1}^{L} \beta_k^b$. As a simple numerical example, consider a 1-buyer case where $H = L = 2$. Suppose we have $\mu_{D^b} = 30$, $\mu_{\varepsilon_m^b} = 0$. This yields $\theta_k^b = 0.5$, $\beta_k^b = 30$, and the expected aggregate L-period buyer order of 60. Suppose the system is observed as $I_s = 80$ and $\mathbf{d_s} = [30,\, 30]$ at the beginning of period s. This results in the modified inventory position $\widehat{I}_s$ of 20. The term in the parentheses in (5.9), on the other hand, is a correction factor, which further reduces $\widehat{I}_s$ if order commitments are greater than μ_{D^b}. To see this, assume two distinct alternative $\underline{\mathbf{d}}_\mathbf{s} = [28,\, 30]$ and $\bar{\mathbf{d}}_\mathbf{s} = [32,\, 30]$ for the order commitment vectors at the beginning of period s. Continuing with previous example, the modified inventory position $\widehat{I}_s$ is further reduced to 18.1 when $\bar{\mathbf{d}}_\mathbf{s} = [32,\, 30]$, and it becomes 22.1 when $\underline{\mathbf{d}}_\mathbf{s} = [28,\, 30]$.

Consequently, the use of Theorem 5.1 in solving the dynamic programming recursions of (5.1) enables us to obtain considerable computational savings. It operates by identifying commitment clusters of the state space that leads to identical replenishment decisions. We treat those clusters by a representative state. Consequently, the policy evaluation in the studied problem is approximate on three counts: (i) the discretization of the continuous space, (ii) the approximation of the censored joint probabilities using a finite Markov chain, (iii) the Taylor's approximation made in evaluating the logarithm of a sum of random variables in the use of commitment clusters. An important question that naturally arise is whether these reduction techniques result in a sufficiently accurate representation of the actual problem being studied. This will be mentioned later in §5.4.

The following corollary to Theorem 5.1 states how stochastic state transitions under the state-space reduction take place from modified state $(\widehat{I}_s,\, \mathfrak{D}_{s,H})$ in period s to the random next-period state in the reduced space, $(\widehat{I}_{s+1},\, \mathfrak{D}_{s+1,H})$.

Corollary 5.1 (to Theorem 5.1) *Suppose a certain replenishment decision* $\widehat{TI}_s$ *is made when the system is in modified state* $(\widehat{I}_s,\, \mathfrak{D}_{s,H})$ *at the beginning of period* s*. Then the random next-period state in the reduced space,* $(\widehat{I}_{s+1},\, \mathfrak{D}_{s+1,H})$ *can be specified as*

$$\begin{aligned} \widehat{I}_{s+1} &= \widehat{TI}_s - \sum_b \Big(\sum_{k=1}^{H+1} \beta_k^b \varepsilon_{s,k}^b - \beta_1^b - \beta_{H+1}^b \ln \frac{\beta_{H+1}^b}{\mu_{D^b}} \Big), && (5.11) \\ \mathfrak{D}_{s+1,H} &= \mathfrak{D}_{s,H} + \sum_b \Big(\sum_{k=2}^{H+1} \beta_k^b \varepsilon_{s,k}^b + \beta_{H+1}^b \ln \mu_{D^b} - \beta_1^b \ln d_{s-1,s}^b \Big) \Big/ \sum_{k=2}^{H+1} \beta_k^b , && (5.12) \end{aligned}$$

where $\varepsilon_{s,k}^b$ *is the random variable denoting the* k*-period ahead commitment update made in period* s*.*

Proof: See Appendix A.7.

□

The logic in Corollary 5.1 can be seen by noting the term in the parentheses on the right-hand side of (5.11). It incorporates the renewed random commitment state vector $\mathbf{D_{s+1}}$ to be observed at period $s+1$ (through the random variables $\varepsilon_{s,2}^b$, $\varepsilon_{s,3}^b$, …, $\varepsilon_{s,H+1}^b$) and the random one-period commitment update to be made in obtaining the realized order in period s, $D_{s,s}$ (through the random variable $\varepsilon_{s,1}^b$). Similarly, the right-hand side of (5.12) contains ($\varepsilon_{s,2}^b$, $\varepsilon_{s,3}^b$, …, $\varepsilon_{s,H+1}^b$), random commitment updates made in period s. Hence (5.11) and (5.12) are for the transitions of the modified states. As a consequence, the dynamic programming recurrence relation given in (5.1) under the state-space reduction becomes

$$\begin{aligned} \widehat{V}_s(\widehat{I}_s,\, \mathfrak{D}_{s,H}) &= \min_{\widehat{TI}_s \geqslant \widehat{I}_s} \Big\{ \widehat{J}_s(\widehat{TI}_s,\, \widehat{I}_s,\, \mathfrak{D}_{s,H}) + E_{\boldsymbol{\mathcal{E}_s}}[\, \widehat{V}_{s+1}(\widehat{I}_{s+1},\, \mathfrak{D}_{s+1,H}) \,] \Big\} \\ \widehat{V}_{N+1}(\widehat{I}_{N+1},\, \mathfrak{D}_{N+1,H}) &= -\, c_{ps} E_{\boldsymbol{\mathcal{E}_{N+1}}, \boldsymbol{\mathcal{E}_{N+2}}, \ldots, \boldsymbol{\mathcal{E}_{N+L}}}[\, \widehat{NI}_{N+L} \,], && (5.13) \end{aligned}$$

for $s = N,\, N-1,\, \ldots,\, 2,\, 1$. $\widehat{J}_s(\widehat{TI}_s,\, \mathfrak{D}_{s,H})$ is the current-period costs associated with the modified state $(\widehat{I}_s,\, \mathfrak{D}_{s,H})$ when action $\widehat{TI}_s$ is selected, and $\widehat{V}_{s+1}(\widehat{I}_{s+1},\, \mathfrak{D}_{s+1,H})$ is trivially the optimal cost-to-go for later periods.

A pseudo-code involved with evaluating these dynamic recursions is given in Figure 5.4. It begins with 4 nested loops; (*i*) one for the decision periods in the planning horizon [*line 1*], (*ii*) one for the set of possible modified inventory states $\widehat{I}_s$ [*line 2*], (*iii*) one for the set of possible modified commitment states $\mathfrak{D}_{s,H}$ [*line 3*], and (*iv*) one for the set of allowable replenishment decisions [*line 4*]. Subsequently, we have two successive blocks of loops each having B additional nested loops. The first block [*the first loop at line 5*] is for evaluating expected costs of carrying inventory and backorders where the expectations are taken with respect to all possible realizations of $\boldsymbol{\mathcal{E}_s}$. The second is for moving to the random next-period modified state and getting the future cost which involves only two state variables [*the second loop at line 5*].

---

1. **For** period s backwards from $s = N$ to $s = 1$
2. **For** modified state $\widehat{I}_s$ values
3. **For** modified state $\mathfrak{D}_{s,H}$ values
4. **For** decision $\widehat{TI}_s$

5. Compute $PC_s(\widehat{TI}_s,\, \widehat{I}_s)$
5. **For** realizations of ϑ_L^1 for buyer 1
6. ...
7. **For** realizations of ϑ_L^B for buyer B
8. Find probability $Pr\{v_L^1,\, \ldots,\, v_L^B\}$ for these realizations
8. Compute $\widehat{J}_s(\widehat{TI}_s,\, \widehat{I}_s,\, \mathfrak{D}_{s,H} \mid v_L^1,\, \ldots,\, v_L^B)$
7. **End for**
6. ...
5. **End for**
5. Compute $\widehat{J}_s(\widehat{TI}_s,\, \widehat{I}_s,\, \mathfrak{D}_{s,H})$ over all realizations

5. **For** realizations of $(\varepsilon_{s,1}^1,\, \ldots,\, \varepsilon_{s,H+1}^1)$ for buyer 1
6. ...
7. **For** realizations of $(\varepsilon_{s,1}^B,\, \ldots,\, \varepsilon_{s,H+1}^B)$ for buyer B
8. Compute $\widehat{I}_{s+1}$ and $\mathfrak{D}_{s+1,H}$, next-period modified states
8. Find joint probability $Pr\{\boldsymbol{\mathcal{E}_s^1},\, \ldots,\, \boldsymbol{\mathcal{E}_s^B}\}$ for next-period modified states
8. Get $\widehat{V}_{s+1}(\widehat{I}_{s+1},\, \mathfrak{D}_{s+1,H} \mid \boldsymbol{\mathcal{E}_s^1},\, \ldots,\, \boldsymbol{\mathcal{E}_s^B})$ from a list of values
7. **End for**
6. ...
5. **End for**
5. Compute expected cost-to-go $E_{\boldsymbol{\mathcal{E}_s}}[\,\widehat{V}_{s+1}(\widehat{I}_{s+1},\, \mathfrak{D}_{s+1,H})\,]$
5. Compute total cost $\widehat{G}_s(\widehat{TI}_s,\, \widehat{I}_s,\, \mathfrak{D}_{s,H}) = \widehat{J}_s(\cdot,\, \cdot,\, \cdot) + E_{\boldsymbol{\mathcal{E}_s}}[\,\widehat{V}_{s+1}(\cdot,\, \cdot)\,]$ for $\widehat{TI}_s$

4. **End for**
4. Find the minimum cost, $\widehat{V}_s(\widehat{I}_s,\, \mathfrak{D}_{s,H})$ over all $\widehat{TI}_s$
3. **End for**
2. **End for**
1. **End for**

Figure 5.4: A pseudo-code of dynamic recursions in (5.13)

One can see the benefit from using the algorithm in Figure 5.4 if the amount of calculation required is compared with that of Figure 5.2. By using the algorithm in Figure 5.4, we have a calculation in only two state variables with $B + 4$ nested loops at the lowest level, instead of $BH + 1$ state variables with $2B + 3$ nested loops.

5.4 An Example of the Computational Process

In this section, we discuss loss of accuracy resulted with the state-space reduction together with the associated gain in the computational requirements.

5.4.1 Solving the stochastic dynamic recursions

In order to solve the stochastic dynamic programming recursions, given by (5.1) and (5.13), we coded them in Matrix Laboratory (MATLAB) Release 14 with Service Pack 1. MATLAB is a high-performance and a high-level language whose basic data element is an array that does not require dimensioning. This allowed us to solve our dynamic programming relations, which requires matrix and vector formulations to store optimal costs and order-up-to-levels evaluated by previous backward iterations for all possible realizations of random system states. Furthermore, MATLAB incorporates two main functionalities allowing us to solve our dynamic problems conveniently. One is its family of add-on application-specific solutions called toolboxes, which extends the MATLAB environment to solve particular classes of problems. The other is its mathematical function library, which is a comprehensive collection of computational algorithms ranging from elementary functions to more sophisticated functions.

5.4.2 Validating the state-space reduction

The state-space reduction technique that we have proposed transforms the dynamic recursions into a simpler one which involves only two state variables. On the other hand, there arises the question of whether the reduction technique results in an accurate representation of the actual problem being studied.

In this section we now describe how to determine whether the reduction technique is valid. The goal of validation is two-fold. (1) First, we try to show that the optimal replenishment policies under the reduced state space agree with the ones that are obtained by using the original state space. (2) The second is to demonstrate whether the state-space reduction is effective in improving the computational and storage requirements. Thus, the claim is that the reduced model formulation, if it can be solved efficiently, does what the original formulation asks us to do. A powerful verification of this claim is to implement both formulations on a set of problem instances. Hence, the approach taken to validate the state-space reduction is an empirical one where we solve various problem instances using the stochastic dynamic programming for both model formulations.

We generate seven problem instances as seen in Table 5.1. They are differentiated by the number of buyers involved (B), the length of the manufacturer's replenishment lead-time assumed (L), the number of the manufacturer's decision periods (N), and the number of discrete values assumed for a continuous variable (M). Note that larger values of these parameters would amplify computational requirements. This constitutes the reason for studying these restricted problem sizes. The instance labelled $INS1$ is generated to represent those cases where multiple buyers exist, and $INS2$ to observe the effect of the replenishment lead-time. The problem instances $INS3$ to $INS7$, on the other hand, are taken into consideration to observe the effects of expanding grid of discrete points (i.e., increasing the value of M).

Table 5.1: Problem instances used in validating state-space reduction

Parameter	$INS1$	$INS2$	$INS3$ to $INS7$
B	2	1	1
$L = H$	2	3	3
N	4	6	6
M	5	5	from 6 to 10 by 1 unit steps

All other parameters and settings (including the capacity level, and costs) are kept the same across all problem instances. Suppose that buyers have the same set of contract flexibility limits in updating their order commitments. They are given by $\{\, \alpha_k^b = \omega_k^b,\ k = 1, 2, 3 \,\} = \{\, 0.4, 0.6, 0.8 \,\}$. The mean realized orders and the mean logarithmic commitment update are taken as $[\mu_{D^1}, \mu_{D^2}] = [30, 20]$ and $\mu_{\varepsilon_k^b} = 0$, for $b = 1, 2$ and $k = 1, 2, 3$, respectively. The covariance matrix of $\boldsymbol{\mathcal{E}_s}$ are assumed to be correspond to a certain correlation structure. More specifically, it is assumed that (i) $(\varepsilon_k^b, \varepsilon_l^b)$'s for the same buyer b but different lags $k \neq l$ have positive correlation of 0.5, (ii) $(\varepsilon_k^b, \varepsilon_k^r)$'s for different buyers $(b \neq r)$ but identical lag (k for both) are negatively correlated with correlation coefficient of -0.3, and (iii) $(\varepsilon_k^b, \varepsilon_l^r)$'s for different buyers $(b \neq r)$ and lags $(k \neq l)$ have zero correlation.

For assessing the effects of state-space reduction, we calculate (i) the maximum percentage deviation in order-up-to levels TI_s (maximum taken over N periods), (ii) the percentage deviation in minimum expected cost-to-go from period 1 (i.e., deviation in $V_1(\cdot, \cdot)$ values) and (iii) the percentage decrease in run time for the reduced model formulation as compared to the original formulation. Table 5.2 gives these results. For problem instance $INS2$ we have a higher percentage deviation since the precision of the approximation in (5.2) reduces for larger L.

Table 5.2: Performance measures

Performance Measure	$INS1$	$INS2$
Maximum percentage deviation in TI_s	1.70%	2.30%
Percentage deviation in V_1	1.09%	1.63%
Percentage decrease in run time	87.0%	88.15%

Figure 5.5, on the other hand, shows the effects of expanding grid of discrete points. Even small increases in M result in an increased computational burden for the original formulation (hence % decrease in run time gets better after state-space reduction), but the gain in the maximum percentage deviation in order-up-to levels does not look much.

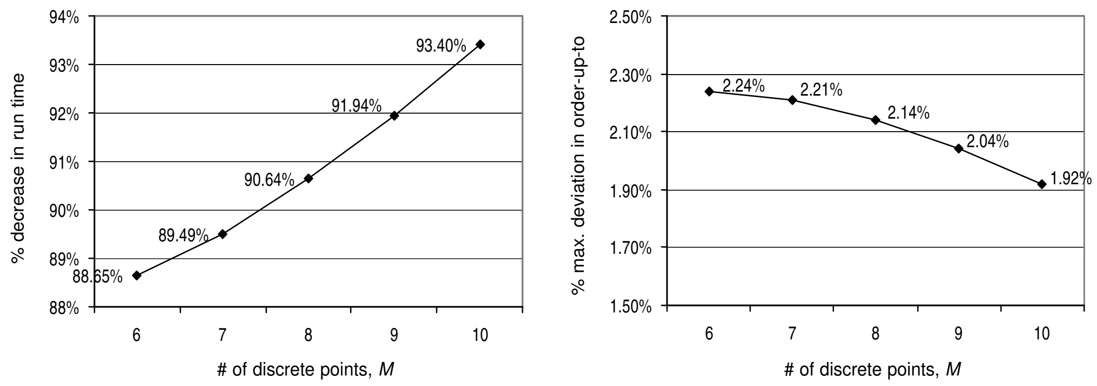


Figure 5.5: Effects of increasing M (for problem instances $INS3$ to $INS7$)

As a consequence, the empirical validation demonstrate the agreement of the replenishment decisions, and the effectiveness of stochastic dynamic recursions under the state-space reduction.

CHAPTER 6

DESCRIPTION OF THE COMPUTATIONAL STUDY

This chapter is devoted to a detailed description of the computational approach used for the manufacturer's multi-period stochastic production/inventory decision problem. We shall perform optimization experiments, each minimizing the expected total cost of the manufacturer, directed toward exploring the relationship between the experimental factors of interest and the measures of performance. §6.1 provides the central research questions of this dissertation that will be addressed through the experimentation. In §6.2, we define factors and identify their levels so that various system scenarios can be experimented. §6.3 describes the base case experiments that are needed as a reference to compare different conditions in the production/inventory system. We discuss, in §6.4, the performance measures that are used to evaluate the performance of different system scenarios. Finally, we consider a comparable alternative to the model under the forecast evolution in §6.5, where we build upon the results from Graves (1999).

6.1 Objectives of the Computational Study

In this research, we investigate a multi-period stochastic production/inventory problem under quantity flexibility (QF) supply contracts. The problem involves a capacitated manufacturer, with an option of subcontracting, who satisfies periodically updated order commitments from multiple contract buyers having stochastic demands. This problem environment is a complicated one that has multiple factors affecting it. In this environment, we would like to make an integrated analysis of the QF contract parameter setting and the production/inventory planning. It is of great value to understand the underlying interplay of the decisions. We thus attempt to investigate the effects, on the decisions and costs of the manufacturer, of possible action options.

In this research we do not attempt to make a specification of the buyers' objective function and constraints, as mentioned in §2.2. The manufacturer makes optimal decisions based on the minimization of his expected total cost subject to the QF contracts and capacity restrictions. This is accomplished using an efficient enumerative application based upon an analytical optimization approach rather than an experimental method such as simulation techniques. The associated cost functions (which are known to be continuous and display differentiability and convexity properties) are evaluated by using discrete values and their approximate probabilities suggested by our finite Markov chain approximation introduced in §3.2 to §3.4.

The computational study thus is intended to delineate ways that this model might be elaborated on to capture various important features of the production and inventory system. The following are the specific issues to be investigated through the computational study.

1. How the appropriate scheme of QF contracts to be offered to the buyers can be determined? What are the effects, on the buyers, of the manufacturer's decisions about the QF contracts?

2. How can the manufacturer's in-house capacity investment decisions benefit from the presence of early order commitments and subcontracting option? What are their implications on the manufacturer's ability to meet the demand?

3. How much can the manufacturer benefit from practicing early order commitments under QF contracts? How much revision flexibility can he tolerate with regard to changes in the capacity level and cost structure?

4. How much benefit can the manufacturer derive from integrating his production/inventory planning with the evolutions of order commitments through the MMFE structure? How does this respond to changes in the variability of demand and the correlation structure inherent in the system?

To address these issues, in the following section we first discuss the action options to be experimented with.

6.2 Experimental Factors and Factor Levels

In this section, we define the experimental factors to be used in our computational study. We classify the experimental factors of interest as either environmental settings or controllable action options. We shall conduct experiments on a variety of system scenarios by varying combinations of levels of these factors. There are a large number of parameters to test. However we have limited our research questions. Hence, there will be two sets of parameters regarding the environments tested. One set of parameters are set throughout the experimentation, we call it environmental settings.

6.2.1 Environmental settings

Environmental settings are of two classes. The first includes those experimental factors which may not be possible for the manufacturer to change. Correlations of commitment updates (CO), coefficient of variation for demand (CV), and the mean demand from the buyers (μ_{D^b}) constitute the uncontrollable environmental settings. The second class of environmental settings may include those not being immediately relevant to the research questions under study although they are controllable. The number of contract buyers (B), the number of decision points for the manufacturer (N), and the manufacturer's replenishment lead-time (L) constitute these. For the complete list of the environmental settings, their levels and level values refer to Table 6.1.

Table 6.1: Environmental settings

Experimental Factor	Label	Levels
Number of buyers	B	2
Number of decision points	N	72
Replenishment lead-time	L	12
Correlations of commitment updates	CO	0, $\pm$ 0.8, $\pm$ 0.5, $\pm$ 0.3
Coefficient of variation for demand	CV	0.20, 0.50
Expected value of demands from buyers	$[\mu_{D^1}, \mu_{D^2}]$	[30, 20]

As mentioned in §4.1, we have restricted the problem environment to only two contract buyers to maintain analytical tractability ($B = 2$). When more contract buyers are committing orders earlier, it allows the manufacturer to pool the demands together to smooth production, depending on the underlying correlation structure across the buyers. A two-buyer case will suffice to understand the consequences of this well-known pooling effect, and including more buyers in the analysis would not provide extra information.

We study a periodic-review inventory system with one-period re-planning frequency over a finite horizon. The number of decision points with replenishment decisions is taken to be $N = 72$ in our experimental runs. In determining the N value, we heed the computational efficiency and still allow modeling of a realistic environment. Although the appropriate choice of N would appear to be extremely model-dependent, we think that given the replenishment lead-time, L, our choice provides a time span long enough to reliably capture all the relevant facts.

The replenishment orders placed by the manufacturer require a certain amount of time until they arrive. The replenishment lead-time is assumed to be $L = 12$ periods. The manufacturer makes inventory replenishment decisions in every period for servicing stochastic orders of the buyers over L periods into the future.

The correlation structure inherent in the system may influence the effectiveness of the actions to be taken. We assume that there exist correlations among commitment updates for a number of future periods in the same period. This allows us to model the correlation among the commitments as well as the current realized order through time and across buyers. The correlations of commitment updates occurring in the same period of time are of two classes: (i) Correlations within commitment updates for a particular buyer $b \in \{1, 2\}$ are all assumed to be positive. The further apart the commitment updates are, the smaller the correlations between them. They are assumed to range from 0.8 to 0.3. (ii) Correlations across buyers for the commitment updates made for the same future period are all taken to be -0.8. We also assume that k-period ahead commitment updates from a buyer are uncorrelated with l-period ahead updates from the other buyer for $k \neq l \in \{1, 2, \ldots, L\}$. It is important to keep in mind that these correlation values are for the *uncensored (intended) commitment updates*. The revision limits are incorporated and the associated censored distributional parameters are obtained by (3.16) and (3.17). We do not consider other correlation structures although this will constitute an interesting issue to study.

We vary the coefficient of variation for demand from low ($CV = 0.20$) to high ($CV = 0.50$) for each buyer, although the CV value may depend on the particulars of the situation. The expected value of demand is assumed to be constant over the planning horizon and taken to be $\mu_{D^1} = 30$ and $\mu_{D^2} = 20$ for the first and second buyers, respectively.

Furthermore, as the results for a terminating experimentation depend on the state of the system at the beginning of the experimentation horizon, initial conditions are of vital importance. We initiate an initial inventory position of $I_0 = L \sum_b \mu_{D^b}$ units. We assume that initially the manufacturer observes a complete set of order commitments $\mathbf{d_1^b} = \left[d_{0,1}^b,\, d_{0,2}^b,\, \ldots,\, d_{0,H}^b \right]$, $b = 1, 2$ initialized to be equivalent to μ_{D^b}'s such that

$$\mathbf{d_1^b} = [\, \mu_{D^b},\, \mu_{D^b},\, \ldots,\, \mu_{D^b} \,] .$$

Thus the first replenishment decision being made at the beginning of period 1 relies on the commitment vector $\mathbf{d_1^b}$. The corresponding replenishment order q_1 will arrive at the beginning of period $L+1$, and will constitute the first arrival. Throughout period 1, the realized orders $d_{1,1}^b$ ($b = 1, 2$) are observed. The manufacturer starts satisfying realized orders consuming his on-hand inventory from period 1 onwards. Period 2 is the first time that updated commitments $d_{1,k+1}^b$ are available to the manufacturer.

The key element in the dynamics of computational experiments is the way in which the order commitments evolve randomly from one period to the next. Given the forecast evolution model, a theoretical distribution could not be found. We use an empirical distribution described in §3.3 and §3.4 to approximate the probability distribution function by a probability function for the chosen points. The number of discrete points used in approximating any relevant distribution is taken to be $M = 5$.

6.2.2 Controllable action options

Controllable action options are those experimental factors the manufacturer does have control over, and their levels are effective for the research questions under study. We identify five main action options for the system under study. They are the number of periods (H) of early order commitments, the flexibility limit per period (FL) in order commitment updates, the manufacturer's in-house capacity level (K), the cost differential between in-house production and subcontracting ($\Delta_c = c_{ps} - c_{pi}$), and the backorder-to-holding cost ratio ($\Delta_\pi = \pi_b/h$). We shall investigate how these factors react with each other and have an effect on the system performance. For the complete list of the controllable action options refer to Table 6.2.

Table 6.2: Controllable factors

Experimental Factor	Label
Commitment horizon	H
Flexibility limit per period	FL
In-house capacity level	K
Cost differential in production	Δ_c
Backorder-to-holding cost ratio	Δ_π

Commitment Horizon (H)

We define the *commitment horizon* (H) to be the number of periods for which the contract buyers submit order commitments in each replenishment decision point. The buyers are committed to purchase from the manufacturer by submitting flexible order quantities as signals before the actual order realizations. The commitment horizon is one of the focal points of the QF contracts, affecting the execution of the production/inventory system and its performance. Order commitment for a future period first becomes available H periods in advance and is immediately exploited in determining the manufacturer's replenishment decision as a signal for the upcoming order sizes during the replenishment lead-time.

The commitment horizon does not need to coincide with the L-period replenishment lead-time. If it is the case that $H < L$, the manufacturer takes the mean realized order to be the best available order information for the periods beyond the commitment horizon. Practicing early order commitments are expected to allow the manufacturer to obtain greater accuracy at an earlier point in time due to the variance reduction effect of committing. On the contrary, from the buyers' perspective, committing an order early on is accompanied by a risk of either underestimating or overestimating the true value of the future orders. The risk increases in the demand uncertainty and decreases in the quality of forecasting machinery used. Therefore, it is valuable to quantify these implications so as to determine whether the manufacturer can derive sufficient benefit from exercising early order commitments. Furthermore, how early should the order commitments be made and which other factors may have significant impacts on such a decision are questions of interest in the computational study.

We vary the commitment horizon from $H = 1$ (minimal demand signal) to $H = L = 12$ (maximal demand signal) in the computational investigation. In the minimal commitment environment, the manufacturer receives order commitments only for the next period, and no other signal for the lead-time demand. Under maximal commitment, the manufacturer would have full signal for the lead-time demand, and completely act on order commitments from the buyers. The factor levels and the values for these levels can be seen in Table 6.3.

Table 6.3: Commitment horizon levels and level values

Experimental Factor	Label	Levels	Level Values
Commitment Horizon	H	minimal commitment	1
		k-period partial commitment	$k \in \{2, 3, \ldots, 11\}$
		maximal commitment	$L = 12$

Flexibility Limit per Period (FL)

We define the *flexibility limit per period* (FL) to be the percentage revision limit on how much room buyers will have in updating their order commitments from period to period. It implies that the buyers are capable of adapting order commitments up or down on the basis of changing circumstances and their own forecasts in the future. The QF contracts attach a degree of commitment to the orders by stipulating flexibility constraints on the buyer's ability to revise them over time. The main role of flexibility in the QF contracts is to balance risks that could be incurred as a result of their prediction errors, by the manufacturer's costs in attaining a wider range of responses.

Not all supply chain partners benefit equally in terms of this flexibility. Given the dynamic and uncertain nature of the operating environment, buyers would like to have a greater flexibility, allowing them to satisfy stochastic market demand at a lower cost. The greater the flexibility, the greater the likelihood buyers can effectively adjust to changes in market conditions.

The manufacturer, on the other hand, demands a lower flexibility level in attaining smoother production schedules at high capacity utilization levels. We examine the implications of order commitments (and hence realized orders) available to the manufacturer from a certain range defined by contract flexibility limits.

We assume that lower and upper flexibility limits are equal and they do not vary over time (i.e., $\omega_k^b = \alpha_k^b = FL$ for all $k = 1, 2, \ldots, H$). Hence, a buyer can not revise k-step order commitment (cumulatively) upward by a fraction of more than $(1 + FL)^k$ or downward by more than $(1 - FL)^k$; that is, the uncertainty about order commitments grows with time periods. The factor levels and the values for these levels are given in Table 6.4.

Table 6.4: Flexibility limit levels and level values

Experimental Factor	Label	Levels	Level Values
Flexibility limit per period	FL	no flexibility	0.01
		tight	0.05
		loose	0.10
		no limit	∞

Capacity Level (K)

We define the *capacity level* (K) to be the maximum amount that can be produced internally by the manufacturer in any period for realized delivery after L periods. The capacity level is a relatively long-term decision, which greatly impacts the manufacturer's ability to match supply with demand. Because of the guaranteed supply up to a certain percentage update associated with the QF contracts, the manufacturer is assumed to have an option of subcontracting as a supplementary capacity. We shall make an integrated analysis of the capacity level in the presence of the QF contracts and the subcontracting option. It is to the manufacturer's benefit to understand the association between his additional capacity costs, contract parameters, and subcontracting.

The terms of QF contracts and the option of subcontracting change the essence of K in terms of manufacturer's benefit. Through managing supply contracts with quantity flexibility, the manufacturer bears some portion of the buyers' risks caused by uncertainty she faces in market demand. The buyer has a certain degree of downward flexibility for updating order commitments while the manufacturer guarantees to deliver up to a certain percentage above it. This complicates the capacity investment decisions. The subcontracting option, on the other hand, serves as a means of hedging due to the buyers' flexibilities (it is naturally possible that subcontracting occurs even if the buyers have no revision flexibility). Hence, all these need to be taken into account at the time of the capacity investment decision.

In the computational investigation, different capacity levels are represented by varying the level of *capacity slack*. The capacity slack identifies the amount of excess capacity over the expected total of orders to be received from all the buyers per lead-time. A certain level of excess capacity is essential for flexibility if fast reaction to change is an important operational requirement. Suppose an amount of excess capacity is measured by a multiple Δ_K of the standard deviation of buyers' orders. Values of Δ_K often depend on how the manufacturer cares about stockout risk. Thus the in-house capacity level K is denoted by

$$K = \mu_D + \Delta_K \sigma_D, \tag{6.1}$$

where μ_D is the expected total of orders to be received from all the buyers in any period, and σ_D is the associated standard deviation. It is important to mention that σ_D in (6.1) is calculated by means of our finite Markov chain approximation introduced in §3.2 to §3.4 (rather than being an experimental or theoretical value), which takes the non-stationary nature of the order commitments into account. We will consider a variety of environments, each corresponding to a level of Δ_K. The capacity slack is varied from a Δ_K value of 0 (no excess capacity) to that of 2.5 (loose capacity) in the computational study. We assume that the in-house capacity is not allowed to be adjusted from period to period during the experimentation horizon N. The factor levels and the values for these levels are summarized in Table 6.5.

Table 6.5: Capacity slack levels and level values

Experimental Factor	Label	Levels	Level Values
Capacity slack	Δ_K	no capacity slack	0
		very tight	0.5
		tight	1
		moderate	1.5
		loose	2.5

Cost Structure **(Δ_c, Δ_π)**

We define the *cost structure* of the manufacturer to be the relative values of his costs. It can be identified by means of two classes of cost drivers; structural and executional. Structural cost drivers of the manufacturer reflect the technology available and the technical complexity of the production system, and both are assumed to be reflected in the relative values of unit costs of in-house production (c_{pi}) and subcontracting (c_{ps}). Executional cost drivers, on the other hand, are concerned with the manufacturer's attitude to stockout risk, which affects his preference for the levels of backordering cost (π_b) and inventory carrying cost (h). The cost structure identifies the perceived relationship between the manufacturer's structural costs and the level of supply and customer service he can provide.

Two issues arise in determining the relative values of the manufacturer's costs. The first is related to the costs of in-house production and subcontracting. How the manufacturer structures his business operations has a direct impact on his ability to compete and deliver service to the buyers. The magnitude of c_{pi} indicates the extent of whether internal business processes are efficiently designed and executed. On the other hand, the guaranteed supply up to a certain percentage under the QF contracts may encourage the manufacturer to supplement his limited production capacity in contingencies. As a form of such supplementary capacity, the manufacturer is assumed to subcontract part of his due shipments to other firms for capacity reasons. That is, the manufacturer is capable of producing the product, but he does not have all the production capacity required to produce all the amount ordered. Subcontracting enables the manufacturer to take advantage of smoothing his releases when facing non-stationary stochastic orders. As a consequence, the magnitude of cost differential between the manufacturer and subcontractor, $\Delta_c = c_{ps} - c_{pi}$, plays the key role in the manufacturer's choice.

The second issue is an executional one which arises in the costs of carrying inventory and backorders. The manufacturer essentially chooses his service level by selecting the relative values of π_b and h. If the manufacturer desires a higher service level to buyers, hence larger inventories, he can simply select a value of backorder-to-holding cost ratio that might be higher than what it is used to be. This preference is reflected through the backorder-to-holding cost ratio $\Delta_\pi = \pi_b/h$.

We shall investigate the attitudes of the manufacturer in various system scenarios represented by different (Δ_c, Δ_π) combinations. We interpret the unit cost of carrying inventory as a base for determining the values of the other costs. The factor levels and the values for these levels are summarized in Table 6.6. The values of Δ_c and Δ_π are relative to $c_{ps} = 2.5$ and $h = 1$, respectively.

Table 6.6: Δ_c and Δ_π levels and level values

Experimental Factor	Label	Levels	Values
Cost differential in production	Δ_c	very low	0.1
		low	0.5
		medium	1
		high	1.5
Backorder-to-holding cost ratio	Δ_π	low	1
		high	5

6.3 Base Case Experiments

The computational investigation that is carried out in this research is fairly complicated, involving a stochastic environment with many experimental factors and possibly a large number of levels for some factors. Thus, some base case experiments are needed as a reference to compare typical operating conditions in the production/inventory system. We believe that a parametric experimentation with respect to some predefined base cases is the most appropriate computational approach. The later results are derived from comparing those of the alternative scenarios to the system performance associated with the base case. We have three classes of base case experiments.

The first class is intended to evaluate various scenarios for commitment horizon. It corresponds to a minimal-early-commitment environment where $H = 1$ and $FL = \infty$. We call this base case the *minimal-commitment base case.* In this base case, the buyers provide only one-period ahead order commitments. They commit themselves to the mean order sizes and update freely. There is only one commitment update possible at the point of order realization and no restriction is imposed on these updates (imposing no restriction nullifies all the commitment values other than μ_{D^b}). So, any value is possible like $\mu_{D^b} = 10$ and $d^b_{1,1} = 1000$ with 990 as the update. This base case offers the buyers utmost opportunity to modify their initial commitments since they are not required to be accurate. This means that there is virtually no commitment. Thus, at the time of replenishment decisions, the manufacturer has the minimal early signal for the order realizations over the following replenishment lead-time.

We establish the second class of base case experiments to investigate various scenarios for flexibility arrangements. It corresponds to a situation where $FL = 0.01$ for a given level $H = k$ of commitment horizon, for $k = 1, 2, \ldots, L$. We call this base case the *minimal-flexibility base case.* In this base case, the buyers do provide order commitments for all the periods in the following k-period commitment horizon. Initial order commitments are to be updated k times until the period of order realization, but with a very restrictive flexibility. This base case offers the buyers much less opportunity to modify their initial commitments. The maximum upward update would amount to less than 13% of the initial commitment (where $13\% = (1 + 0.01)^{H+1}/100$ for the largest horizon of $H = 12$).

The third class of base case experiments serves as a reference to evaluate various system scenarios differing in the capacity level K. We call this the *capacity base case.* This base case assumes that the available in-house capacity is constant and equal to the mean order per lead-time (i.e., the case when there is no excess capacity, $\Delta_K = 0$).

All the other environmental settings, controllable action options and initial system conditions are kept the same across the base case experiments. For these three base case experiments, the complete list of the values for the factor levels are summarized in Table 6.7.

Table 6.7: Base case experiments and factor levels

Experimental Factor	Label	Minimal Commitment Base Case	Minimal Flexibility Base Case	Capacity Base Case
Correlations of updates	CO	0.3, - 0.8	0.3, - 0.8	0.3, - 0.8
Coefficient of variation	CV	0.2	0.2	0.2
Number of buyers	B	2	2	2
Number of decision points	N	72	72	72
Replenishment lead-time	L	12	12	12
Commitment horizon	H	1	$k \in \{1, 2, \ldots, L\}$	12
Flexibility limit per period	FL	∞	0.01	0.10
Capacity slack	Δ_K	1	1	0
Cost differential in production	Δ_c	1.5	1.5	1.5
Backorder-to-holding cost ratio	Δ_π	5	5	5

6.4 Performance Measures

In this section, measures of system performance that will be used to evaluate different system scenarios are discussed. We now describe seven performance measures that are most relevant in our case in Table 6.8.

Table 6.8: Performance measures

Performance Measures	Label
Cost improvements	CI_H^+, CI_H^-, CI_K
Mean order-up-to deviation	TI_{dev}
Order-up-to instability	TI_{ins}
Capacity utilization	CU
Fill-rate	φ

The measures of system performance are computed as an average performance observed over the N periods on a given experimental run. Since the initial conditions for a terminating experimentation generally affect the desired measures of performance, we exclude the first L immaterial periods to avoid the effects of initial inventory chosen in all these measures, and use those periods from $L + 1$ to $N + L$. Figure 6.1 helps explain this setting. Our modeling assumption is that initially the system has a positive inventory level due to an initial order of $q_{1-L} = \sum_b \mu_{D^b}$ to arrive at the beginning of period 1. The first replenishment decision relying on buyers' commitments $\mathbf{d_1}$ is made at the beginning of period 1 to arrive at the beginning of period $L+1$. And, the manufacturer starts satisfying buyers' realized orders consuming I_1.

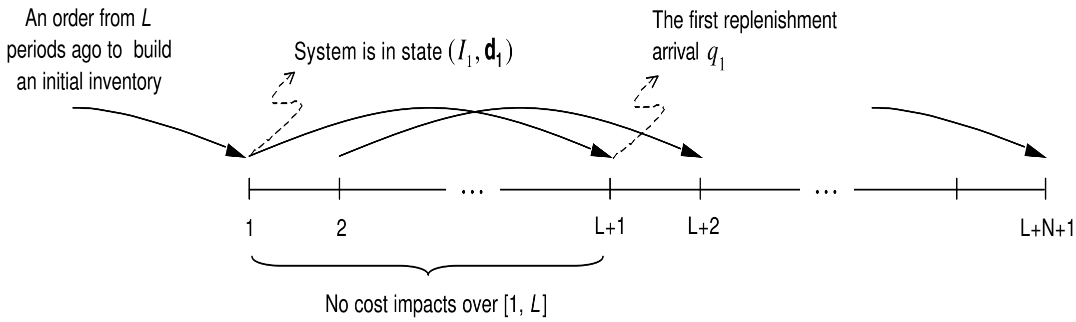


Figure 6.1: Illustrating the first L immaterial periods

Cost Improvements (CI_H^+, CI_H^-, CI_K)

We define the *cost improvement* (CI) to be the manufacturer's percentage cost change over a certain base case of the considered scenario. The manufacturer's cost is assumed to be the minimum expected cost-to-go from the beginning of the planning horizon, assuming that the system is in state (I_1, $\mathbf{d_1}$) in period $s = 1$ and an optimal replenishment decision is made in every period thereon $s = 1, 2, \ldots, N$ [cf. Fig. 6.1]. The formula for CI is

$$CI \quad = \quad 100\, \frac{V_{L+1}(base\ case) - V_{L+1}(scenario)}{V_{L+1}(base\ case)}, \tag{6.2}$$

where $V_{L+1}(base\ case)$ is the minimum expected cost-to-go from the beginning of period $L+1$ to the end of period $N+L$ corresponding to a base case of interest. There are three types of cost measures.

The first is calculated against the minimal-commitment base case (i.e., when $H = 1$ and $FL = \infty$), and is referred to as *percentage cost saving* (CI_H^+). The minimal-commitment base case corresponds to a cost upper bound for other alternative (H, FL) combinations with $H > 1$ and $FL < \infty$. CI_H^+ thus represents the relative percentage cost advantage that the manufacturer could attain if commitment horizon H is extended and flexibility limit per period FL is restricted. A cost advantage arises since practicing early order commitments allows the manufacturer to obtain greater accuracy at an earlier point in time as compared to the minimal-commitment base case.

The second type of cost measure, which is referred to as *percentage cost increase* (CI_H^-), is calculated against the minimal-flexibility base case (i.e., when $FL = 0.01$ for a given level $H = k$ of commitment horizon, $k = 1, 2, \ldots, L$). The minimal-flexibility base case corresponds to a cost lower bound for other alternative (H, FL) combinations with $H = k$ and $FL > 0.01$. CI_H^- represents the relative percentage cost disadvantage due to using the optimal replenishment policy which is obtained when the buyers have larger revision flexibility. A cost disadvantage arises since this is likely to cause inefficiency and extra costs at the manufacturer because of responding to a wider range of orders, as compared to the minimal-flexibility case.

The third type of cost measure, which is denoted by CI_K, is calculated against the capacity base case (i.e., when $\Delta_K = 0$). The capacity base case corresponds to a cost upper bound for other alternative Δ_K combinations with $\Delta_K > 0$. It represents the relative percentage cost advantage that could be attained if the manufacturer has a certain amount of excess capacity over the mean order per lead-time. This provides a sort of capacity flexibility with the manufacturer, in turn he could better match supply with demand, as compared to the capacity base case.

Mean Order-up-to Deviation (TI_{dev})

We define the *mean order-up-to deviation* (TI_{dev}) to be the average ratio of optimal order-up-to positions to the mean order per lead-time. It is calculated by

$$TI_{dev} = \frac{1}{N-L} \sum_{s=L+1}^{N} \frac{TI_s^*}{L \sum_b \mu_{D^b}}, \tag{6.3}$$

where s is time index over the number of decision points N, TI_s^* is the optimal order-up-to position of period s, and $L \sum_b \mu_{D^b}$ is the mean order per lead-time.

Optimal order-up-to position over the mean order per lead-time measures how planned target inventory positions match the mean realized order. The more the manufacturer cares about backorders (by higher backorder penalty), the larger the TI_{dev}.

Order-up-to Instability (TI_{ins})

We define the *order-up-to instability* (TI_{ins}) to be the average absolute deviation between optimal order-up-to positions, in fraction of the mean order per lead-time, of consecutive decision periods on a given experimental run of the system. The formula for computing the order-up-to instability is

$$TI_{ins} \;=\; \frac{1}{N-L} \sum_{s=L+1}^{N-1} \frac{|\, TI_s^* - TI_{s+1}^* \,|}{L \sum_b \mu_{D^b}}, \tag{6.4}$$

where TI_s^* is the optimal order-up-to position of period s. We first calculate the absolute deviation in optimal order-up-to positions of consecutive decision periods. These deviations are divided by the respective mean order from all buyers per lead-time, and averaged over all s except the first L periods.

The stability of order-up-to positions is of interest because the more variable (less stable) order-up-to levels are expected to require expend more managerial and physical production resources (since production decisions are highly dependent on order-up-to positions). This would imply more fluctuating production quantities from one period to the next, which in turn results in higher penalties associated with period-to-period production variation and more stringent (hence costly) administrative action/control. A more stable policy, on the other hand, justifies a base-stock type standard policy (as an approximation).

Capacity Utilization (CU)

We define the *capacity utilization* (CU) to be the extent to which the manufacturer actually uses his in-house production capacity on a given experimental run of the system. It is measured as an average percentage rate

$$CU \;=\; \frac{1}{N-L} \sum_{s=L+1}^{N} \frac{\min\{TI_s^* - I_s,\, K\}}{K}, \tag{6.5}$$

where K is the maximum amount that can be produced internally, and $TI_s^* - I_s$ is the replenishment amount in period s, q_s [cf. Section 4.1].

The capacity utilization is of interest because the effective level of customer service should be interpreted together with the level of capacity utilization. The buyers would like to have greater flexibility, allowing them to satisfy the uncertain market demand at a lower cost. Whereas, the manufacturer demands a lower flexibility with the aim of attaining smoother production schedules at high capacity utilization. Low capacity utilization leads to additional manufacturing expenses created by excess capacity. High capacity utilization, on the other hand, gives rise to higher risk of lost customers or penalties associated with unsatisfied buyers' orders. As a middle way between these two extremes, medium capacity utilization (i.e., a certain amount of capacity slack) potentially offers more flexibility without much costly administrative control.

Fill Rate (φ)

We define the *fill rate*, as type-2 service level, to be the expected proportion of total realized order over all the buyers that is satisfied immediately from the manufacturer's finished-goods inventory. It is measured as an average proportion observed over a given experimental run of the system

$$\varphi \quad = \quad \frac{1}{N-L} \sum_{s=L+1}^{N} \left(1 - \frac{\mathrm{E}[TI_s^* - \sum\limits_b \sum\limits_{t=s}^{s+L-1} D_{t,t}^b]^-}{\sum\limits_b \mu_{D^b}} \right). \tag{6.6}$$

The expression $\mathrm{E}[TI_s^* - \sum\limits_b \sum\limits_{t=s}^{s+L-1} D_{t,t}^b]^-$ reduces to $\mathrm{E}[NI_{s+L}]^-$ due to Eq. (4.4). It gives the expected number of backorders (i.e., current inventory position not capable to meet) over all buyers that occur at the end of period $s + L - 1$. Thus, the ratio $\mathrm{E}[TI_s^* - \sum\limits_b \sum\limits_{t=s}^{s+L-1} D_{t,t}^b]^- / \sum\limits_b \mu_{D^b}$ gives the fraction of total orders that are expected to stock out each period.

φ in (6.6) is not the traditional fill-rate service level for the buyer since the allowable range of order realizations are restricted due to the presence of contract flexibility limits. As we have already mentioned in §2.2, the manufacturer offers contracts that entice buyers to commit their orders in advance while limiting their ability to revise order commitments over time. Each contract buyer first determines her *intended order commitments* in each period as intended future self plans. The revision limits requested by the manufacturer are then applied to these self plans. The intended order commitments may hit one of the limits. In such a case, the limit value is submitted to the manufacturer. This brings about a loss to the contract buyer on those intended orders over the upper limit.

There is a clear relationship between the contract buyer's loss and the extent of restrictions on her commitment updates. Any additional restriction could be expected to have significant effects in terms of further buyer losses. Suppose buyer b can not revise a k-step order commitment $d^b_{s,s+k}$ upward by a fraction of more than α^b_k for $k = 1, 2, \ldots, H$ [1]. Thus, buyer b is required to restrict her order realizations in period $s + H$ to be within the range

$$d^b_{s,s+H} \leqslant D^b_{s+H,s+H} \leqslant \prod_{k=1}^{H} (1 + \alpha^b_k)\, d^b_{s,s+H}.$$

This results in the following range for cumulative commitment updates from that buyer over the H-period commitment horizon

$$1 \leqslant R^b_{s,H} R^b_{s+1,H-1} \ldots R^b_{s+H-1,1} \leqslant \prod_{k=1}^{H} (1 + \alpha^b_k),$$

where $R^b_{s+H-k,k}$ is the random variable denoting the multiplicative update made to k-step order commitment from buyer b in period $s + H - k$. Thus, the loss associated with this restriction is a random variable and depends on the commitment horizon, the flexibility limits, and the distribution of commitment updates.

There are many ways in which the restrictions on order commitments could be stipulated in the QF contracts. So, we may need some criterion for choosing among various flexibility alternatives. This can be accomplished by introducing the concept of an expected loss. Let $\daleth^b(\boldsymbol{\mathcal{A}^b})$ denote the *expected loss due to limited flexibility* buyer b may suffer when she is offered the set $\boldsymbol{\mathcal{A}^b} = \{\, \alpha^b_k, \quad k = 1, 2, \ldots, H \,\}$ of flexibility limits.

Let $\tilde{R}^b_{s,k}$ denote the k-step intended commitment update from buyer b in period s. Note that it corresponds to the uncensored counterpart of $R^b_{s,k}$. Thus, we have the vector $\tilde{\mathbf{R}}^{\mathbf{b}}_{\mathbf{s}} = [\tilde{R}^b_{s,1}, \tilde{R}^b_{s,2}, \ldots, \tilde{R}^b_{s,H}]$. Since the commitment updates $\ln \tilde{R}^b_{s,k}$ are normally distributed, the expected loss may be written as

$$\daleth^b(\boldsymbol{\mathcal{A}^b}) \;=\; 100 \, \frac{1}{\mathrm{E}[\prod_{k=1}^{H} \tilde{R}^b_{s,k}]} \sum_{\prod_{k=1}^{H} (1+\alpha^b_k)}^{\infty} \left[\prod_{k=1}^{H} \tilde{R}^b_{s,k} - \prod_{k=1}^{H} (1 + \alpha^b_k) \right] f_{\tilde{\mathbf{R}}^{\mathbf{b}}}(\cdot\,), \tag{6.7}$$

[1] Note that downward flexibility is assumed to be not restricted (i.e., $\omega^b_k = 1$) since losses are meaningful for the right-tail of the distribution. However, it would also be an interesting issue to calculate overstocking cost of the manufacturer by quantifying the left-tail of the distribution separately.

where $\mathrm{E}[\prod_{k=1}^{H} \tilde{R}^b_{s,k}]$ is the expected value of the intended cumulative update, and $f_{\tilde{\mathbf{R}}^\mathbf{b}}(\cdot)$ is the multivariate lognormal distribution function of the intended commitment updates from buyer b. As the expected loss function (6.7) does not depend on (targeted or realized) inventory random variables, the level of expected loss is not affected by the manufacturer's inventory policy. $\daleth^b$ is a nonnegative and non-increasing function of flexibility limits, $\prod_{k=1}^{H}(1+\alpha^b_k)$.

Clearly, a buyer considering such a contractual agreement needs to evaluate whether she had a relative advantage in reducing the expected losses. Obviously, a preferable contract, on the buyer's side, would be one that minimizes the expected loss. Hence, the buyers would like to have flexibility as large α^b_k's early on (lower k's) as possible.

The central problem in estimating $\daleth^b$ is to find the discrete values and probabilities of commitment updates that fall outside the allowable range; that is, the tail values for the multivariate distribution function $f_{\tilde{\mathbf{R}}^\mathbf{b}}(\cdot)$. We address this problem by means of our finite Markovian structure, introduced in §3.3 and §3.4. We obtain the discrete values and probabilities beyond the allowable range by applying the finite Markovian structure while the revision limits are taken to be very loose (i.e., $\alpha^b_k = \infty$ virtually). As a consequence, the expected loss is an approximate value, and it is the general order of magnitude the manufacturer and his contract buyers should be concerned with, not by its precise value.

6.5 A Comparative Alternative to Inventory Model under Forecast Evolution

In Chapter 4, we have suggested an enhanced production/inventory planning that knows of the operating rules in the order commitments submitted by contract buyers, and how they eventually become the order realizations [cf. Eq. (4.14)]. The optimal replenishment policy is a state-dependent order-up-to type, in which order commitments are explicit component of the state space. It was characterized by formulating a complicated dynamic program, and referred to as the *inventory model under the forecast evolution* (MUFE).

The claim is that the manufacturer benefits significantly from using the MUFE, and the value of resultant optimality will be considerable. We believe that this is due to a better handling of the order uncertainty in deciding on target inventory. That is attained by taking the variability of commitment updates into account.

As a way of substantiating this claim, this section presents a comparative alternative to MUFE. We build upon the results from Graves (1999) in developing the alternative inventory control model. Graves (1999) introduces an adaptive base-stock policy where the demand process is described by an autoregressive integrated moving average (ARIMA) process of order (0, 1, 1). The policy is adaptive since he adjusts the base stock as the order forecast changes. The forecast function is a first-order exponential-weighted moving average (EWMA), which provides the minimum-mean-squared-error forecast for this demand process. Thus, in developing an alternative to MUFE, (*i*) we replace the martingale forecast evolution model with an ARIMA(0, 1, 1) process in the manufacturer's forecasting engine, (*ii*) we develop a novel way of estimating unknown parameters of the related ARIMA(0, 1, 1) process, and (*iii*) we use the adaptive base-stock policy instead of our staircase state-dependent policy for replenishment ordering decisions. We call this alternative the *inventory model under the ARIMA process* (MUMA).

The practical significance of MUMA is that it serves as a comparable alternative to our proposal. This is due to the following reasons: (*i*) The MUMA assumes a similar operating environment with MUFE. Graves (1999) studies a periodic-review inventory system, where there exists a fixed and known replenishment lead-time. The demand process is non-stationary, and any order not satisfied by inventory is backordered. There is no fixed replenishment ordering cost for the manufacturer, and the inventory carrying and backorder penalty costs are linear. (*ii*) There is an analytical relation between stationary time series models and forecast evolution models. More specifically, Güllü (1993) shows that every forecast evolution model corresponds to a series of particular moving average (MA) models and vice versa. He suggests methods of obtaining one representation given the other one. Since the buyers' orders are modeled as an ARIMA(0, 1, 1) in MUMA, we can identify a forecast evolution model that corresponds to the MA representation of orders [2]. Güllü (1993) also states that any suitable stationary time series model (e.g., a class of ARMA(p, q) processes) can be

[2] This is due to that an ARIMA model of order (0, 1, 1) is one in which the observations follow a stationary MA(1) process after they have been differenced.

approximated by a finite order moving average model, which in turn corresponds to a forecast evolution model. Thus, ARIMA models are (approximately) special cases of the forecast evolution models. This analytical correspondence will allow us to estimate the unknown parameters of the ARIMA(0, 1, 1) process on the basis of the covariances between commitment updates of the forecast evolution in §6.5.1.

We should also mention that Graves (1999) differs in several aspects, but these can be suitably modified for a more comparable treatment. (*i*) The QF contracts are not explicitly considered in Graves (1999), and hence no restriction is imposed on the order quantity submitted by the buyer. However, we will take the forecast error for the ARIMA process as bounded and solve the inventory model as such. (*ii*) Graves (1999) adopts an adaptive base-stock policy where the base stock is adjusted as the order forecast changes (first-order EWMA forecast being identical over the lead time). We have, on the other hand, an order-up-to type policy where each period a sufficient order is placed to restore the inventory position to a target level. The target levels may vary by period depending on extended advance order commitments. Hence the difference between the adaptive base-stock policy of Graves (1999) and the policy used in this research is the way in which order quantities are determined. (*iii*) Graves (1999) assumes no capacity restrictions. However, for comparison purposes, MUFE can assume a suitably low value for the cost differential between the in-house production and subcontracting, Δ_c. This will induce an unlimited capacity. (*iv*) Graves (1999) assumes a single-item inventory system where the manufacturer satisfies the demand from a single buyer. So, the corresponding ARIMA process is univariate. However, we readily extend the results from Graves (1999) to the multivariate ARIMA processes.

In characterizing the demand process for this alternative inventory control model, we will not follow the usual procedures in the iterative Box-Jenkins modeling strategy (Box et al. 1994). This is recommended for constructing an ARIMA model from a given time series. Rather, the model structure is set forth in advance (eliminating the model identification stage) and the ARIMA(p, d, q) model is identified as ARIMA(0, 1, 1). Appendix C describes the related bivariate ARIMA(0, 1, 1) process and its forecast model in more detail. §6.5.1 is devoted to our way of estimating unknown parameters of this process. §6.5.2 gives the general ordering of realizing the comparison. Finally, in §6.5.3 we discuss the policy under the ARIMA model.

6.5.1 Estimating the parameters of the related ARIMA(0, 1, 1) process

This section introduces how we develop estimators for the unknown moving average parameters $\mathbf{\Theta}$ and the unknown disturbance covariances $\Sigma_{\boldsymbol{\eta}}$ in the bivariate ARIMA(0, 1, 1) process [cf. Eq. (C.1)]. In the literature various techniques are available for dealing with this estimation problem (Box et al. 1994). These range from simple descriptive techniques to complex inferential models combining regression analysis. Our way of estimating the unknown parameters, however, does not attempt to fit a regression model maximizing the likelihood of the fitted values given a set of observed time series data. Rather, it is an optimization technique to suggest those estimators that provide an adequate description of the correlation structure in the system. The motivation behind this approach is to produce comparable evolution results given by the time series pattern studied, maintaining the stationary mean of the series.

Let $\Sigma_{\mathbf{D}}(k)$ denote the covariance matrix of lag k, $k \geqslant 0$ for the bivariate ARIMA(0, 1, 1) process [cf. Eq. (C.1)],

$$\Sigma_{\mathbf{D}}(k) \quad = \quad \begin{bmatrix} \mathrm{Cov}(D^1_{s+k,s+k},\, D^1_{s,s}) & \mathrm{Cov}(D^1_{s+k,s+k},\, D^2_{s,s}) \\ \mathrm{Cov}(D^2_{s+k,s+k},\, D^1_{s,s}) & \mathrm{Cov}(D^2_{s+k,s+k},\, D^2_{s,s}) \end{bmatrix}. \tag{6.8}$$

The diagonal elements of $\Sigma_{\mathbf{D}}(k)$ are called the *kth-order auto-covariances* between the order realizations submitted by the same buyer. The off-diagonal elements are the *kth-order cross-covariances* across different buyers. We impose a restriction that the kth-order cross-covariances for $k \geqslant 1$ are zero (i.e., cross-covariance of only lag 0 is assumed to be nonzero), which is due our major process assumption [cf. Section 3.1.1].

We now derive closed-form expressions that relate the covariance matrix $\Sigma_{\mathbf{D}}(k)$ to the unknown parameters (i.e., $\mathbf{\Theta}$ and $\Sigma_{\boldsymbol{\eta}}$) in the bivariate ARIMA(0, 1, 1) process [cf. Eq. (C.1)]. These expressions are called the *theoretical covariance functions.* The theoretical auto-covariance function of lag k corresponding to buyer $b \in \{1, 2\}$ is derived as

$$
\begin{aligned}
\mathrm{Cov}^{\boldsymbol{\Theta}}_{bb}(k) &= \mathrm{Cov}(D^b_{s+k,s+k},\, D^b_{s,s}) \qquad (6.9)\\
&= \begin{cases} (1+(1-\theta_{bb})^2)\,\sigma^2_{\eta^b} \;+\; \theta^2_{br}\,\sigma^2_{\eta^r} \;-\; 2\,(1-\theta_{bb})\,\theta_{br}\,\sigma_{\eta^b,\eta^r} & \text{if } k=0\\ -\,(1-\theta_{bb})\,\sigma^2_{\eta^b} \;+\; \theta_{br}\,\sigma_{\eta^b,\eta^r} & \text{if } k=1\\ 0 & \text{if } k \geqslant 2 \end{cases}
\end{aligned}
$$

where the expression for $k = 0$ corresponds to the variance of $D^b_{s,s}$. Similarly, the theoretical cross-covariance function of lag k for buyers $b \neq r \in \{1, 2\}$ can be derived as

$$
\begin{aligned}
\mathrm{Cov}^{\boldsymbol{\Theta}}_{br}(k) &= \mathrm{Cov}(D^b_{s+k,s+k},\, D^r_{s,s}) \qquad (6.10)\\
&= \begin{cases} -\,(1-\theta_{bb})\,\theta_{rb}\,\sigma^2_{\eta^b} \;-\; (1-\theta_{rr})\,\theta_{br}\,\sigma^2_{\eta^r} & \\ \qquad +\,[1+\theta_{br}\,\theta_{rb}+(1-\theta_{bb})\,(1-\theta_{rr})]\,\sigma_{\eta^b,\eta^r} & \text{if } k=0\\ -\,(1-\theta_{bb})\,\sigma_{\eta^b,\eta^r} \;+\; \theta_{br}\,\sigma^2_{\eta^r} & \text{if } k=1\\ 0 & \text{if } k \geqslant 2 \end{cases}
\end{aligned}
$$

Note that the theoretical covariance functions display a distinctive feature of vanishing at lags $k \geqslant 2$. In general, a moving average process with an order of q has non-zero covariances only up to lag q and zero covariances for lags $k > q$ (Box et al. 1994).

As we have already mentioned in §6.2.1, the correlation structure inherent in the system is assumed to be characterized by the correlations of *intended commitment updates*, $\tilde{\boldsymbol{\mathcal{E}}}_{\mathbf{s}}$ (that is, uncensored commitment updates). Accordingly, we have a $B(H+1)\times B(H+1)$ covariance matrix $\Sigma_{\tilde{\boldsymbol{\mathcal{E}}}}$ of the intended commitment updates for the largest commitment horizon of $H = 12$. The covariance matrix $\Sigma_{\mathbf{D}}(k)$ for the bivariate ARIMA(0, 1, 1) process can then be expressed in terms of the uncensored covariance matrix $\Sigma_{\tilde{\boldsymbol{\mathcal{E}}}}$. This follows from the forecast evolution equation [cf. Eq. (3.19)] and Assumption 3.2 of the forecast evolution structure (i.e., independence assumption for the commitment updates at different points in time). We call these expressions the *target covariance functions*.

It is important to note that the uncensored covariance matrix $\Sigma_{\tilde{\boldsymbol{\mathcal{E}}}}$ of intended commitment updates is for a multivariate *normal distribution*, whereas the vector $\mathbf{D_{s,s}}$ of order realizations [cf. Eq. (C.1)] is *lognormally distributed*. Thus, we need the following definition in deriving the target covariance functions.

Definition 6.1 : *Suppose* $\ln \mathbf{X} = \left[\begin{smallmatrix}\ln X_i\\ \ln X_j\end{smallmatrix}\right]$ *is a random vector with bivariate Normal* *density such that* $\mathrm{E}[\mathbf{X}] = \left[\begin{smallmatrix}\mu_i\\ \mu_j\end{smallmatrix}\right]$ *and* $\mathrm{Var}(X_k) = \sigma_k^2$ *for* $k = i$ *and* j*. Then, we have the following relationship between the covariance terms of the Normal vector* $\ln \mathbf{X}$ *and the respective Lognormal vector* $\mathbf{X}$ *(see Law and Kelton 2000)*

$$\mathrm{Cov}(X_i,\, X_j) \quad = \quad e^{\mu_i + \mu_j + (\sigma_i^2 + \sigma_j^2)/2}\ (e^{\overbrace{\mathrm{Cov}(\ln X_i,\, \ln X_j)}^{\zeta}} - 1). \tag{6.11}$$

As we have already mentioned in §3.1, note that $\sigma^2_{\tilde{\varepsilon}^b_k}$ is the variance of the kth component of $\tilde{\boldsymbol{\mathcal{E}}}_{\boldsymbol{s}}$ corresponding to bth buyer. $\sigma_{\tilde{\varepsilon}^b_k,\, \tilde{\varepsilon}^r_l}$ is the covariance between the kth component of $\tilde{\boldsymbol{\mathcal{E}}}_{\boldsymbol{s}}$ corresponding to bth buyer and the lth component of $\tilde{\boldsymbol{\mathcal{E}}}_{\boldsymbol{s}}$ corresponding to rth buyer. Thus, the target auto-covariance function of lag k for buyer $b \in \{1,\, 2\}$ can be derived as

$$\begin{aligned}
&\mathrm{Cov}(\ln D^b_{s+k,s+k},\, \ln D^b_{s,s} \mid \Sigma_{\tilde{\boldsymbol{\mathcal{E}}}}) \ = \\
&= \ \mathrm{Cov}(\ln \mu_{D^b} + \sum_{m=1}^{H+1} \tilde{\varepsilon}^b_{s+k-m+1,m},\, \ln \mu_{D^b} + \sum_{n=1}^{H+1} \tilde{\varepsilon}^b_{s-n+1,n}) \\
&= \ \begin{cases} \sum_{m=1}^{H+1} \sigma^2_{\tilde{\varepsilon}^b_m} & \text{if } k = 0, \\ \sum_{m=1}^{H+1-k} \sigma_{\tilde{\varepsilon}^b_m,\, \tilde{\varepsilon}^b_{m+k}} & \text{if } 1 \leqslant k \leqslant H, \\ 0 & \text{if } k \geqslant H+1, \end{cases} \\
&\triangleq \ \zeta \text{ in Eq. (6.11)} \ \xrightarrow{\text{by Def. 6.1}} \ \mathrm{Cov}^{\tilde{\boldsymbol{\mathcal{E}}}}_{bb}(k).
\end{aligned} \tag{6.12}$$

Similarly, the target cross-covariance function of lag k for buyers $b \neq r \in \{1,\, 2\}$ can be derived as

$$\begin{aligned}
&\mathrm{Cov}(\ln D^b_{s+k,s+k},\, \ln D^r_{s,s} \mid \Sigma_{\tilde{\boldsymbol{\mathcal{E}}}}) \ = \\
&= \ \mathrm{Cov}(\ln \mu_{D^b} + \sum_{m=1}^{H+1} \tilde{\varepsilon}^b_{s+k-m+1,m},\, \ln \mu_{D^r} + \sum_{n=1}^{H+1} \tilde{\varepsilon}^r_{s-n+1,n}) \\
&= \ \begin{cases} \sum_{m=1}^{H+1} \sigma_{\tilde{\varepsilon}^b_m,\, \tilde{\varepsilon}^r_m} & \text{for } b \neq r \text{ and } k = 0, \\ 0 & \text{for } b \neq r \text{ and } k \geqslant 1, \end{cases} \\
&\triangleq \ \zeta \text{ in Eq. (6.11)} \ \xrightarrow{\text{by Def. 6.1}} \ \mathrm{Cov}^{\tilde{\boldsymbol{\mathcal{E}}}}_{br}(k).
\end{aligned} \tag{6.13}$$

Consequently, in estimating the unknown parameters $\boldsymbol{\Theta}$ and $\Sigma_{\boldsymbol{\eta}}$ of the bivariate ARIMA process, we are looking for the identity

$$\begin{array}{ccc}
\mathrm{Cov}^{\tilde{\boldsymbol{\mathcal{E}}}}_{br}(k) & = & \mathrm{Cov}^{\boldsymbol{\Theta}}_{br}(k) \\
\nearrow & & \nwarrow \\
\text{from (6.12), (6.13) given } \Sigma_{\tilde{\boldsymbol{\mathcal{E}}}} & & \text{from (6.10), (6.11) for } \boldsymbol{\Theta},\, \Sigma_{\boldsymbol{\eta}}
\end{array}$$

where $b, r \in \{1, 2\}$. Therefore, to create an equivalence between the ARIMA(0, 1, 1) process and our forecast evolution model, we are interested in minimizing the deviations of $\text{Cov}_{br}^{\boldsymbol{\Theta}}(k)$ from $\text{Cov}_{br}^{\tilde{\boldsymbol{\mathcal{E}}}}(k)$ for lags $k = 0, 1$ and buyers $b, r \in \{1, 2\}$. This will be addressed by minimizing the sum of squared deviations from the target auto- and cross-covariances under certain regularity constraints. The optimization model, referred to as $\text{PAR}_{\boldsymbol{\Theta},\boldsymbol{\eta}}$, can then be constructed as follows:

Optimization model $\text{PAR}_{\boldsymbol{\Theta},\boldsymbol{\eta}}$***:***

The input parameters of the model are

- $\text{Cov}_{br}^{\tilde{\boldsymbol{\mathcal{E}}}}(k)$, the *target* auto- and cross-covariance functions of lags $k = 0$ and 1 for the buyers $b, r \in \{1, 2\}$.

The decision variables of the model are

- $\boldsymbol{\Theta}$, the 2×2 matrix of moving average parameters,
- $\Sigma_{\boldsymbol{\eta}}$, the 2×2 covariance matrix of normally-distributed random disturbance vector $\boldsymbol{\eta_s}$.

The resultant variables of the model are

- $\text{Cov}_{br}^{\boldsymbol{\Theta}}(k)$, the *theoretical* auto- and cross-covariance functions of lags $k = 0$ and 1 for the buyers $b, r \in \{1, 2\}$,
- $\Delta_{br}(k)$, the deviation of $\text{Cov}_{br}^{\boldsymbol{\Theta}}(k)$ from $\text{Cov}_{br}^{\tilde{\boldsymbol{\mathcal{E}}}}(k)$, of lag $k \in \{0, 1\}$ and buyers $b, r \in \{1, 2\}$.

The model $\text{PAR}_{\boldsymbol{\Theta},\boldsymbol{\eta}}$ is defined as follows:

$$\text{Minimize} \quad \sum_{k=0}^{1} \left[\sum_{b=1}^{2} \Delta_{bb}^2(k) + \sum_{b \neq r=1}^{2} \Delta_{br}^2(k) \right] \tag{6.14}$$

subject to

$$\Delta_{bb}(k) = \mathrm{Cov}_{bb}^{\boldsymbol{\Theta}}(k) - \mathrm{Cov}_{bb}^{\tilde{\boldsymbol{\mathcal{E}}}}(k) \qquad \forall\, b \in \{1,\, 2\},\ k \in \{0,\, 1\} \tag{6.15}$$

$$\Delta_{br}(k) = \mathrm{Cov}_{br}^{\boldsymbol{\Theta}}(k) - \mathrm{Cov}_{br}^{\tilde{\boldsymbol{\mathcal{E}}}}(k) \qquad \forall\, b \neq r \in \{1,\, 2\},\ k \in \{0,\, 1\} \tag{6.16}$$

$$0 < \theta_{bb} \leqslant 1 \qquad \forall\, b \in \{1,\, 2\} \tag{6.17}$$

$$-1 \leqslant \sigma_{\eta^b,\eta^r}/(\sigma_{\eta^b}\, \sigma_{\eta^r}) \leqslant 1 \qquad \forall\, b \neq r \in \{1,\, 2\} \tag{6.18}$$

$$\sigma_{\eta^b} \geqslant 0 \qquad \forall\, b \in \{1,\, 2\} \tag{6.19}$$

The objective function minimizes the sum of squared deviations between the estimated and target auto- and cross-covariances. The estimated ARIMA(0, 1, 1) model is more parsimonious than the forecast evolution model. Its adequacy is subject to theoretical limitations of ARIMA models [3]. The estimated ARIMA model will match the target covariance functions, but does not necessarily result in forecasts satisfying a martingale. It is also important to keep in mind that the estimated covariances $\Sigma_{\boldsymbol{\eta}}$ are *uncensored*, as revision limits are not accommodated in the estimation process (due to the use of the uncensored covariance matrix $\Sigma_{\tilde{\boldsymbol{\mathcal{E}}}}$ of intended commitment updates).

Definitional constraints (6.15) compute the deviation of the estimated auto-covariances from the target auto-covariances corresponding to buyer $b \in \{1, 2\}$ at lag $k \in \{0, 1\}$. Constraints (6.16), on the other hand, compute the deviation of the estimated cross-covariances from the target cross-covariances corresponding to buyers $b \neq r \in \{1, 2\}$ at lag $k \in \{0, 1\}$. Constraints (6.17) ensure that the bivariate ARIMA(0, 1, 1) process is a non-stationary process. Constraint (6.18) ensures that the covariance matrix $\Sigma_{\boldsymbol{\eta}}$ of disturbances $\boldsymbol{\eta_s}$ forms a positive semi-definite matrix. Constraints (6.19) are nonnegativity constraints corresponding to the variances of $\boldsymbol{\eta_s}$.

6.5.2 Realization steps of the comparison

In this section, we give the general ordering of realizing the comparison between the inventory models MUFE and MUMA.

[3] Specifically, the ARIMA(0, 1, 1) induces at most one-period lag correlations, and ignores higher order correlations inherent in the environment.

Step-1: The related bivariate ARIMA(0, 1, 1) process and its forecast model

The first-order exponential-weighted moving average (EWMA) forecasting scheme given in Appendix C corresponds to an environment in which the buyers provide only one-period ahead order commitments (i.e., $H = 1$ in our notational convention), which are their realized order in the previous period. There is only one update possible at the point of order realization. The basis of the update is the EWMA forecast, $\mathbf{F_s}$ (with an additive update $\boldsymbol{\eta_s}$ on $\mathbf{F_s}$ such that $\mathbf{D_{s,s}} = \mathbf{F_s} + \boldsymbol{\eta_s}$).

Step-2: Estimating the unknown parameters

For estimating the unknown parameters of the bivariate ARIMA(0, 1, 1) process, we do not use a least-squares estimation algorithm with an empirical data series. Hence, the differences between the fitted values and the observed time series values are not of primary interest. Instead, the parameter estimation is accomplished by an optimization technique minimizing the deviations from the correlation structure inherent in our underlying assumed update system. We are doing so since reflecting the variation structure and maintaining the mean will suffice to generate the candidate policy. We do not attempt to initiate the buyer orders, but the response to them.

Step-3: Incorporating the contract flexibility limits

The estimated covariances $\Sigma_{\boldsymbol{\eta}}$ in (7.1) are *uncensored*, as revision limits are not accommodated in the estimation process. This is due to the use of the uncensored covariance matrix $\Sigma_{\tilde{\boldsymbol{\varepsilon}}}$ of intended commitment updates. However, the MUMA corresponds to an operating environment where the buyers are committing orders one-period ahead (i.e., $H = 1$) and the corresponding forecast errors $\boldsymbol{\eta_s}$ are restricted by some flexibility limits. Thus, the estimated covariances $\Sigma_{\boldsymbol{\eta}}$ need to be modified for incorporating those revision limits. This will result in censored distributional parameters using expressions (3.16) and (3.17).

Our way of incorporating the restricted revision flexibility in MUMA is as follows. Suppose MUFE with the commitment horizon of H and the flexibility limit FL per period. We should ask the question of how to specify MUMA corresponding to this MUFE. We assume that the forecast errors $\boldsymbol{\eta_s}$ ($= \mathbf{D_{s,s}} - \mathbf{F_s}$; i.e., error from immediate EWMA forecast) for the bivariate ARIMA process of MUMA occur within the certain range [4],

$$\ln(1 - FL)^H \leqslant \eta_s^b \leqslant \ln(1 + FL)^H, \tag{6.20}$$

for $b = 1,\, 2$ and $s = 2,\, 3,\, \ldots,\, N + L$. The accumulation of FL over the commitment horizon H in MUFE corresponds to the revision flexibility allowed under MUMA (in forecast errors η_s^b). The larger the H of MUFE, the wider the range of η_s^b in MUMA. In this way we assure the comparable quality of the two alternative inventory models. The revision limits in (6.20) are accommodated by modifying the disturbance covariances $\Sigma_{\boldsymbol{\eta}}$ estimated as a result of optimizing $\mathrm{PAR}_{\boldsymbol{\Theta},\boldsymbol{\eta}}$, [cf. Eqs. (3.16) and (3.17)]. The probability distributions of random disturbances η_s^b thus become censored.

Step-4: Quantifying the replenishment ordering policy

We use the adaptive base-stock policy of Graves (1999) (assuming no capacity restrictions) in generating replenishment orders under MUMA. §6.5.3 discuss this in more detail. Instead, we would use our staircase state-dependent replenishment policy, and hence generalize the results for capacitated environments. But, the corresponding alternative would compare only the manufacturer's way of generating order forecasts (comparing MMFE versus ARIMA(0, 1, 1)), and ignore benchmarking our proposal for replenishment ordering decisions.

Finally, as for the dynamics of computing the costs, we use our finite Markov chain approximation introduced in §3.2 to §3.4. This facilitates probability modeling of censored disturbances η_s^b, and the evaluating the associated expectations.

6.5.3 The replenishment policy under the related ARIMA(0, 1, 1) process

Graves (1999) characterizes the replenishment policy of the manufacturer as an adaptive base-stock type when the first-order EWMA forecasting scheme [cf. Eq. (C.2)] is applied to the bivariate ARIMA(0, 1, 1) process [cf. Eq. (C.1)]. Accordingly, the replenishment order placed in period s for delivery in period $s + L$ is determined by

$$\begin{aligned} q_s^{ma} &= \sum_b D_{s-1,s-1}^b + \sum_b L\,(F_s^b - F_{s-1}^b) \\ &= \sum_b D_{s-1,s-1}^b + L\,(\theta_{11}\,\eta_{s-1}^1 + \theta_{12}\,\eta_{s-1}^2), \end{aligned} \tag{6.21}$$

[4] Logarithms are used since the forecast errors $\boldsymbol{\eta_s}$ under MUMA are additive.

for $s = 1, 2, \ldots, N$. This is called the *policy under the ARIMA process* (PUMA). It is a myopic policy as q_s^{ma} minimizes the expected current-period cost for the L-period lead-time into the future. It is an adaptive policy since it compensates for anticipated changes in the inventory position due to the changes in the forecasts. That is; it adapts $D_{s-1,s-1}^b$, $b = 1, 2$ up or down by the most recent shift up or down. The state of the inventory system under PUMA at the beginning of period s is $(I_s, \mathbf{F_s})$.

We assume the same environmental settings as the inventory model under the forecast evolution. Also, we initiate the same initial inventory position of $I_0 = L \sum_b \mu_{D^b}$ units at the beginning of the planning horizon. The manufacturer observes only one-period ahead order commitments from the buyers in each replenishment decision point. Each initial forecast state F_1^b for $b \in \{1, 2\}$ is initialized to be equivalent to μ_{D^b}. The first replenishment order q_1^{ma} will arrive at the beginning of period $L+1$. During period 1, the realized orders $d_{1,1}^b$ are observed, and the manufacturer starts satisfying orders from the buyers. Period 2 is the first time that the manufacturer has updated forecasts F_2^b.

We shall compare effectiveness of the alternative inventory control models on the basis of the order-up-to positions and the expected total costs. Once the initial inventory position, I_0, has been specified, the order-up-to position of any period s, $TI_s = I_s + q_s^{ma}$, can be obtained from the inventory balance equation

$$I_s = I_0 + \sum_{i=1}^{s-1} q_i^{ma} - \sum_{i=1}^{s-1} \sum_b d_{i,i}^b, \tag{6.22}$$

for $s = 2, 3, \ldots, N+L$. Due to the L-period replenishment lead-time, the costs associated with carrying inventory and backorders depend on the expected net inventory level that will be carried over in period $s+L$, NI_{s+L},

$$NI_{s+L} = \sum_{i=1}^{s} q_i^{ma} - \sum_{i=1}^{s+L-1} \sum_b d_{i,i}^b, \tag{6.23}$$

for $s = 1, 2, \ldots, N+1$ (where $q_i^{ma} = 0$ for $i = N+1$ since we have N replenishment decisions). It denotes total received replenishment order minus total satisfied demand before ordering in period $s+L$. Thus, q_s^{ma} minimizes the following expected current-period cost

$$J_s(TI_s, \mathbf{F_s}) = \min_{q_s^{ma}} \left\{ c_{pi}\, q_s^{ma} + h\, \mathrm{E}[NI_{s+L}]^+ + \mathrm{E}[g(NI_{s+L}, \pi_1, \pi_2)]^- \right\}, \tag{6.24}$$

where the replenishment cost include only the unit cost of in-house production, c_{pi}, as the manufacturer is assumed to have no capacity restrictions. The term $g(\cdot)$ in (6.24) denotes a function of π_b, $b \in \{1, 2\}$.

Consequently, the minimum expected total cost of the policy PUMA can be obtained as

$$V_1(I_1, \mathbf{F_1}) = \sum_{s=1}^{N+1} J_s(TI_s, \mathbf{F_s}). \tag{6.25}$$

The key element in the dynamics of computing the costs [cf. Eq. (6.24) and (6.25)] is the way in which the random disturbances $\boldsymbol{\eta_s}$ are described. The presence of the flexibility limits makes the associated probability distribution censored. As before, we thus use our finite Markov chain approximation introduced in §3.2 to §3.4, and insert it in finding the expectations, $\mathrm{E}[NI_{s+L}]^+$ and $\mathrm{E}[g(NI_{s+L}, \pi_1, \pi_2)]^-$.

CHAPTER 7

ANALYSIS OF COMPUTATIONAL RESULTS

In this chapter, we shall present and discuss the results of an extensive computational investigation described in the previous chapter. As mentioned in §2.2, the focus of our inquiry in this research is on the decision problems that arise for the manufacturer. It has little to do with the buyers, as we do not attempt to make a specification of the buyers' objective function and constraints. Nevertheless, we emphasize three perspectives on the manufacturer's decision problems to evaluate differing concerns of the manufacturer and his contract buyers, as depicted in Figure 7.1. These perspectives surround our research questions [cf. §6.1] addressed by the computational study.

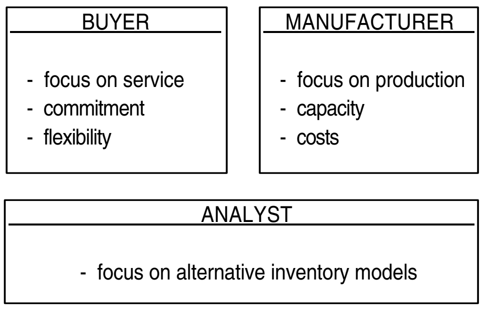


Figure 7.1: Three perspectives on the manufacturer's decision problems

The *buyer perspective* primarily focuses on the service level provided by the manufacturer. Accordingly, it concerns the effects, on the buyers, of the manufacturer's decisions about the QF contracts, relating to our first research question. The length of commitment horizon and the level of revision flexibility to be offered to buyers are taken to be central to this perspective. We propose a menu of various commitment and flexibility arrangements as it may assist the buyers in selecting appropriate scheme of QF contracts. To understand the implications of the buyers' preferences, we also examine the relationship between the cost performance of the manufacturer versus the extent of early order commitments. §7.1 is devoted to an analysis from the buyer's perspective.

The *manufacturer perspective* represents the production viewpoint, and concerns the second and third research question. The production-focus leads the manufacturer to emphasize cost reduction for production and inventory, instead of emphasis on filling periodical buyer orders. Hence, this perspective is primarily concerned with capacity investment decisions, possible ways of supplementing it, and their implications. Central to this perspective is to integrate the capacity decisions with the operational aspects of the QF contracts. We propose a menu of various capacity level choices which vary in the cost differential in production and the backorder-to-holding cost ratio. A menu is a decision aid. It lists possible actions with very close (if not identical) performance outcomes. This may assist the manufacturer in selecting appropriate capacity levels being tuned according to the basic approach in his operations. The implications of the manufacturer's preferences are analyzed based on the cost changes and the service level that can be attained by a particular choice of capacity level. It will be the purpose of §7.2 to give insights from the manufacturer's perspective.

The *analyst perspective* represents an aggregated view of the production and service viewpoints. It centers on the fourth research question. This perspective concerns order information needs of the cooperation between the manufacturer and his contract buyers. There exists an analyst assuming responsibility to evaluate and recommend alternative ways of the production/inventory planning. Specifically, we evaluate the two alternative inventory control models; the model under the forecast evolution (MUFE) and the model under the ARIMA demand process (MUMA), as mentioned before in §6.5. §7.3 is devoted to an analysis from the analyst's perspective.

7.1 Analysis from the Buyer Perspective

This section aims to develop insights into the impact of the QF contracting on the buyer's part in the service collection by our suggested production/inventory planning. The QF contracts are intended to describe mechanisms that align the incentives of the buyers with the manufacturer's costs by way of sharing risks. This is attained by stipulating some cooperative rules primarily related to the commitment and flexibility arrangements. The commitment horizon (H) represents the length of time period for which the buyers submit early order commitments once every period. The flexibility limit per period (FL) quantifies the amount of flexibility the buyers are allowed in updating their order commitments from period to period. The main role of these two contract parameters is to balance buyers' costs that would be incurred as a result of their prediction errors, by the manufacturer's costs in accommodating a range of responses to the buyer's service. Consequently, developing a menu of various commitment and flexibility arrangements may provide valuable insights into the buyers' preferences for the length of commitment horizon and the extent of revision flexibility.

7.1.1 A menu of (H, FL) combinations

Buyers' making early order commitments ensure a certain amount of order information is received by the manufacturer once every period. This is a form of advance order information. Karaesmen et al. (2002) discuss that advance order information can be communicated to upstream partners in a variety of ways. In this research we exercise them via series of commitments for consecutive future periods with limited revision flexibilities.

The commitment horizon H and the flexibility limit per period FL are the two primary leverages in such contracts. In analyzing the computational results in what follows, it is important to keep in mind how they interact with each other. The further out the commitment horizon for a given level of *flexibility limit per period* FL, the less accurate the buyers are required to be. This is reflected by the range $[(1-FL)^H, (1+FL)^H]$ of allowable order quantities $D^b_{s,s}$ cumulatively at the end of update series. Hence the manufacturer must respond to wider range of orders each time period H is extended over. For a given level of *cumulative flexibility* (i.e., for the same allowable range of order quantities) as the commitment horizon H is extended, on the other hand, the buyers are required to be more accurate (i.e., they do not prefer to commit orders in advance). The manufacturer would like to receive commitments for all the periods in the replenishment lead-time. As a compromise, the buyers may be offered more rewards or higher flexibility in modifying their order commitments. Offering more flexibility as an accompaniment to early (but for longer horizons) commitment is to induce the practice of committing orders in advance by making it more appealing to the buyers.

Developing a menu of QF contracts may be an appealing approach to leverage the buyers' preferences. Under this approach, each buyer is offered a menu of various (commitment horizon - flexibility limit) combinations, in turn she can choose among the alternatives. The main incentive for the buyer's choosing among the alternatives is to minimize her risk of experiencing a stockout in servicing the market. Hence, she negotiates for higher flexibility and/or less commitments, passing on some portion of the cost associated with uncertainty to the manufacturer. Buyers expect their costs to go down as they take on lower risk. On the other hand, they may take advantage of rewards (e.g., price discounts, priority etc.) offered by the manufacturer in motivating them to provide the desired level of extended early commitments (i.e. larger H).

To get insights into these implications, we study eleven different levels of the commitment horizon (ranging from $H = 2$ to $H = 12$ $(= L)$), and five different levels of the flexibility limit per period (ranging from $FL = 0.02$ to $FL = 0.10$ by increments 0.02). This amounts to $11 \times 5 = 55$ alternative combinations to be included in a menu of QF contracts. Different $(H,\, FL)$ combinations are evaluated based on the expected loss due to the restricted flexibility, $\daleth^b$, on the buyers side, and the percentage cost saving, CI_H^+, on the manufacturer side. It is important to note that for the purposes of measuring the $\daleth^b$ values, we assume that only upward order modifications are restricted and there is no downward flexibility limit (i.e., $\omega_k^b = 1$, $\alpha_k^b = FL$ for all $b \in \{1,\, 2\}$ and $k \in \{1,\, 2,\, \ldots,\, H\}$).

The $\daleth^b$ value, given by (6.6), is the expected loss due to the limited flexibility the buyer b will suffer from when she accepts a particular $(H,\, FL)$ combination. The CI_H^+ value, given by (6.3), on the other hand, represents the relative cost advantage that could be attained by exercising that particular $(H,\, FL)$ combination against the minimal-commitment base case used as a reference.

Figure 7.2 allows us to see the trade-off between the expected loss of the buyer $b = 1$ and the cost saving for the manufacturer for these 55 alternative (H, FL) combinations. It shows how the $(100 - \daleth^1)\%$ (i.e., the fraction of intended orders being within the cumulative flexibility range) relates to the cost saving as the (H, FL) combinations to be offered to the buyer vary. Table D.1 in Appendix D gives the solution details. A similar menu can be readily developed for the other buyer, but omitted for simplicity.

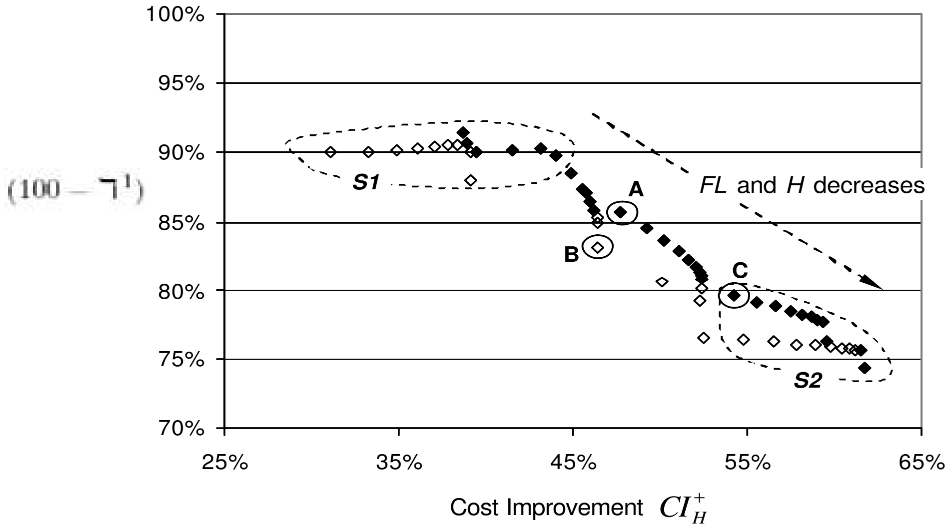


Figure 7.2: A menu of (H, FL) combinations to be offered to buyer $b = 1$

Combinations near the bottom left of the plot have relatively higher expected loss at a lower cost saving. Those towards the northeast corner offer lower expected loss accompanied by higher cost savings. Hence combinations to the upper right (e.g., A) are preferred to points to their lower left (e.g., B). An efficient frontier exists in Figure 7.2, which is shown by the points ♦. It represents those (H, FL) combinations that are considered the most efficient in the sense that they have the greatest relative cost saving given the same level of expected loss or the lowest expected loss given the same cost saving. It begins at the left tail with an expected loss of 8.6% Either the commitment horizon is cut off and/or the flexibility limit is reduced until one reaches the right tail, which represents a 25.6% expected loss with nondecreasing cost savings.

It may also be the case where only some subset of these alternative combinations [cf. Fig. 7.2] is offered to the buyer. For instance, the manufacturer may behave more customer-oriented in the sense that the menu offers those alternatives which imply relatively lower expected buyer losses (e.g., subset $S1$). These alternatives may yield different cost savings when accepted. Hence, the buyer is rewarded (e.g., by providing price discounts, priority etc.) in accordance with the potential cost savings. More reward shall be offered to motivate her to prefer the lowest-cost alternatives (i.e., extended advance commitments and/or stricter flexibility) towards the lower right. This will identify many different but equivalent combinations to be accepted by the buyer. If the manufacturer behaves more conservative, on the other hand, the menu offers those alternatives which imply relatively higher cost savings for the manufacturer (e.g., subset $S2$). They may lead to higher expected buyer losses. From among them all, the buyer selects the ones which have the lowest expected losses (i.e., the highest level from the bottom); e.g., combination C.

7.1.2 Effects of early order commitments and flexibility

In this section, we elaborate further on the effects of various commitment and flexibility arrangements given in Figure 7.2. Specifically, we give more details about how the production/inventory system performance is affected by varying the commitment horizon and/or the flexibility limit. This may assist the manufacturer in evaluating the potential rewards to be offered to the buyers in response to their earlier commitments and/or limited flexibility.

We first examine the relationship between the extent of early order commitments versus the cost performance of the manufacturer from a gain or loss perspective. We study four levels of the commitment horizon from $H = 3$ to $H = 12$ $(= L)$. The buyers are restricted in updating their order commitments up and down, and the levels of upward and downward flexibility per period are taken to be $FL = 0.10$ (i.e., $\omega_k^b = \alpha_k^b = 0.10$ for all $b \in \{1, 2\}$ and $k \in \{1, 2, \ldots, H\}$).

The cost advantage of practicing early order commitments is evaluated against the minimal-commitment base case (i.e., when $H = 1$ and $FL = \infty$), and measured by the percentage cost savings (CI_H^+). The curve above the zero level in Figure 7.3 shows the percentage cost savings CI_H^+ corresponding to the four levels of the commitment horizon. Cost savings range from 28.4% to 16.5% with a higher decrease rate as the horizon length is extended. The solution details are listed in Table D.2 in Appendix D. This reveals that early order commitments are valuable to the manufacturer, but earlier commitments result in lower cost saving against the minimal-commitment base case.

The cost disadvantage of practicing early order commitments for a given level $H = k$ of the commitment horizon, on the other hand, is evaluated against the corresponding minimal-flexibility base case (i.e., when $H = k$ and $FL = 0.01$), and measured by the percentage cost increases (CI_H^-) (i.e., larger losses). Figure 7.3 plots the CI_H^- curve below the zero level corresponding to the four levels of the commitment horizon. Percentage cost increases (i.e., larger loss) range from -27.3% to -35.5% with a higher decrease rate when moving towards the largest horizon of $H = 12$ periods. This indicates the manufacturer's loss that would be experienced by offering higher flexibility than 1% per period.

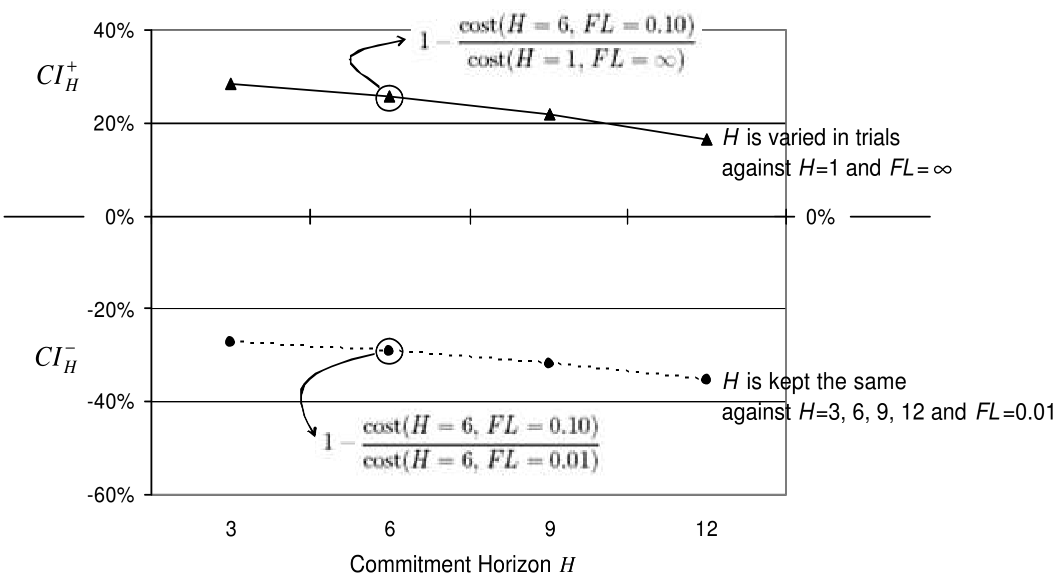


Figure 7.3: Percentage cost savings and increases, CI_H^+ and CI_H^-, by increasing H

Consequently, the cost saving CI_H^+ and cost increase CI_H^- curves serve as upper and lower bound curves, respectively, for alternative (H, FL) combinations to be offered to the buyers. For the largest horizon of $H = 12$ periods and least flexibility of $FL = 0.01$, the cost saving against the minimal-commitment base case (i.e., when $H = 1$ and $FL = \infty$) turns out to be $CI_H^+ = 1 - \frac{\text{cost}(H{=}12, FL{=}0.01)}{\text{cost}(H{=}1, FL{=}\infty)} = 32.1\%$. If the level of flexibility per period FL becomes stricter than 0.10, then both the CI_H^+ and CI_H^- curves shift upward (making the gains larger whereas losses smaller), and they become less steep (making the changes less sensitive and hence indicating a significant interaction between H and FL). The cost changes are inversely related to the level of flexibility. At high flexibility levels, the manufacturer obtains almost no benefit from the buyers' committing early since he must respond to a wider range of orders each time period. This reduces the value of practicing extended advance order commitments. Figure 7.4 illustrates this behavior by plotting the CI_H^+ curve for unrestricted revision flexibility (i.e., $FL = \infty$). It compares the effect of early order commitments only. The solution details are listed in Table D.2 in Appendix D.

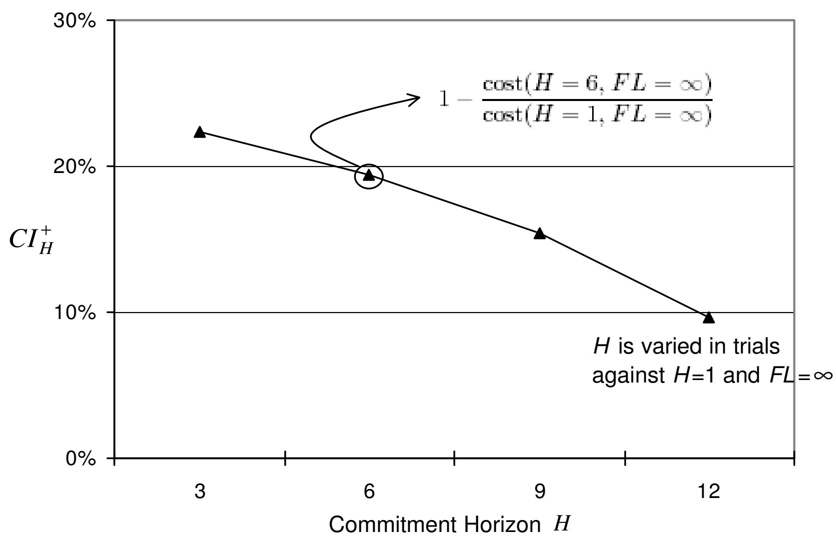


Figure 7.4: Percentage cost savings, CI_H^+, by increasing H under $FL = \infty$

In addition to the cost performance of the production/inventory system, we may also be concerned with how much the order realizations are being matched by the planned target inventory levels. This will give some insights into the buyers' risks of experiencing a stockout. Figure 7.5 provides the plots of the mean order-up-to deviation (TI_{dev}) for various levels of the commitment horizon, where the level of flexibility per period is taken to be $FL = 0.10$. As we have already mentioned in §6.4, a TI_{dev} value gives optimal order-up-to position, as a fraction of the mean order per lead-time, averaged on the N decision periods [cf. Eq. (6.3)]. TI_{dev} increases as H increases (from 1 to 12) but showing a relatively low variation even for $FL = 0.10$, ranging from $TI_{dev} = 1.21$ (when $H = 1$) to 1.37 (when $H = 12$). Table D.2 in Appendix D shows the solution details. The overall level of TI_{dev} is directly related with how much the manufacturer avoids the backorders. This essentially provides the manufacturer with some kind of safety being relative to the coefficient of variation for total orders over the replenishment lead-time. The greater the coefficient of variation, whether in terms of H and FL [1], the higher the value of TI_{dev}. As TI_{dev} becomes greater than one and continues to increase for a certain level of FL, the manufacturer gradually overestimate the buyers' future orders, so the inventory holding cost becomes the dominant cost component while the share of the backorder cost diminishes. Thus, the cost performances of the manufacturer continue to decrease as TI_{dev} increases, [cf. Fig. 7.3].

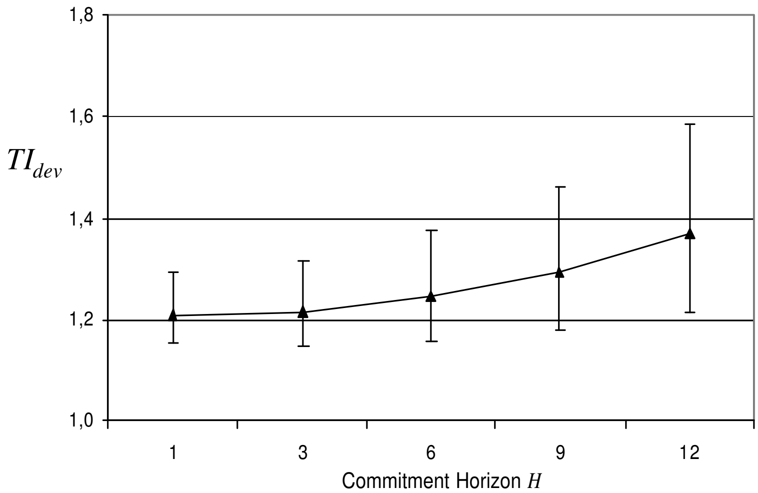


Figure 7.5: Mean order-up-to deviation TI_{dev} by increasing H

[1] Varying the values of H and/or FL relates to different levels for the coefficient of variation since they represent how the discrete values of order realization are dispersed about the mean.

Figure 7.5 also shows intervals around TI_{dev} values. Each interval represents the range of optimal order-up-to positions used in calculating that particular TI_{dev} value. The longer the commitment horizon and/or the less stricter the flexibility, the wider the range of optimal order-up-to positions, which is due to an increasing uncertainty by the extended commitment horizon of $H = 12$.

7.2 Analysis from the Manufacturer Perspective

This section aims to develop insights into the manufacturer's in-house capacity decision integrated with the operational aspects of the QF contracts. The in-house capacity investment needs to be decided quite early in the planning horizon. The decision depends on the terms and conditions of the QF contracts offered by the manufacturer and on the availability of supplementary capacity option. Hence, it is to the manufacturer's benefit to understand the association between his additional capacity costs, QF contract parameters, and subcontracting. We first investigate how various commitment and flexibility arrangements to be stipulated in the QF contracts affect the appropriate level of in-house capacity. This leads to a challenging managerial task when it comes to balance the trade-off between having excess capacity in some periods and insufficient capacity in others while satisfying stochastic orders from contract buyers. Hence we shall also examine how supply-side flexibility through subcontracting can help to more effectively match supply with demand. The manufacturer perspective emphasizes three primary leverages for understanding these associations.

The first is the capacity slack (Δ_K) identifying the amount of excess capacity over the mean order per lead-time. The second leverage is related to manufacturing technology and capability being reflected in the cost differential ($\Delta_c = c_{ps} - c_{pi}$) that the manufacturer incurs between the in-house production and subcontracting. The third leverage is concerned with how much the manufacturer avoids backorders. This is reflected in the ratio ($\Delta_\pi = \pi_b/h$) of backordering cost to holding cost. These may assist the manufacturer in understanding the multifaceted nature of his capacity investment decision.

Consequently, developing a menu of various capacity levels may help in the capacity investment decision for improving the manufacturer's ability to deliver service at acceptable cost during the execution of the QF contracts.

7.2.1 A menu of (Δ_K, Δ_c, Δ_π) combinations

This section proposes a menu of capacity levels which differ in the cost differential in production, Δ_c, and the backorder-to-holding cost ratio, Δ_π. The magnitude of Δ_c gives cost disparities among the manufacturer and subcontractor. The ratio Δ_π reflects how much the manufacturer cares about backorders in executing the QF contracts. We study (*i*) five different levels of the capacity slack ranging from $\Delta_K = 0.5$ to $\Delta_K = 2.5$ by increments 0.5, (*ii*) four levels of the cost differential from $\Delta_c = 0.1$ to $\Delta_c = 1.5$, and (*iii*) two levels of the backorder-to-holding cost ratio Δ_π being low and high. This amounts to $5 \times 4 \times 2 = 40$ alternative combinations to be included in a menu of capacity levels. Different (Δ_K, Δ_c, Δ_π) combinations are evaluated based on the cost improvement (CI_K), on the manufacturer side, and the fill-rate service level (φ), on the buyers side.

The φ value, given by (6.6), represents the expected proportion of total realized order over all the buyers that is satisfied immediately from the manufacturer's finished-goods inventory. The cost improvements CI_K, given by (6.3), are measured relative to the capacity base case, where the maximum amount that can be produced internally is equal to the mean order per lead-time (i.e., with $\Delta_K = 0$).

Figure 7.6 plots these 40 alternative (Δ_K, Δ_c, Δ_π) combinations based on the cost improvements CI_K and the fill-rate service level φ. Several clusters of alternative combinations can be identified. Combinations for a given level of the backorder-to-holding cost ratio Δ_π form a major cluster. We have two of them as we study two different levels of Δ_π. The fill-rate service levels for low Δ_π are significantly smaller than those associated with high Δ_π. Within each Δ_π group, four crescent-shaped clusters exist (denoted I through IV for low Δ_π) corresponding to four different levels of the cost differential Δ_c. Crescent-shaped cluster I corresponds to $\Delta_c = 0.1$ and cluster IV to $\Delta_c = 1.5$. The capacity slack Δ_K increases as one moves from the left to the right. Combinations near the bottom left have relatively low service level and poorer cost improvement. Those near the upper right (in northeast direction) offer higher cost improvements at a higher service level. The solution details are listed in Table D.3 in Appendix D.

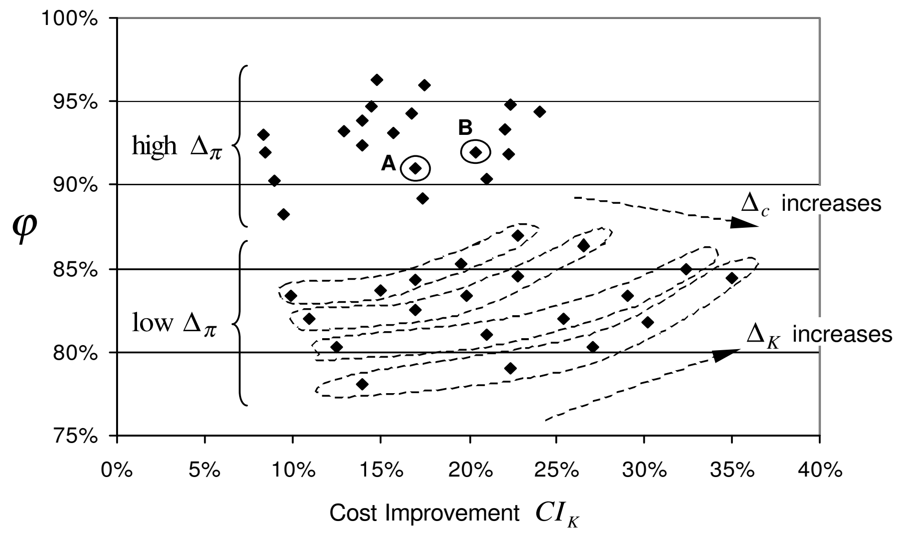


Figure 7.6: A menu of (Δ_K, Δ_c, Δ_π) combinations

The manufacturer generally seeks the lowest-cost operations for a given level of service (i.e., rightmost combination). This requires a larger in-house capacity. However, often some organizational constraints exist, and capacity investment can be restricted such that complete match of the demand may not be feasible or desirable. As we have already mentioned in §6.2, these organizational constraints are assumed to be indirectly reflected in structural and executional cost drivers (i.e., Δ_c and Δ_π, respectively). The attitudes of the manufacturer to these cost drivers affect his capacity investment decisions, as it identifies the perceived relationship between his structural costs and the level of supply and customer service he can provide. Thus, an important methodological feature of this menu approach is that it can determine the performance of a chosen capacity level by comparing it with those of others under various cost structures (i.e., organizational restrictions).

As Δ_c is somehow reduced (e.g., by cooperating with the subcontractor to help them achieve lower costs), it is less likely to experience restrictions on the supply availability. Hence the manufacturer takes advantage of smoothing his releases in the face of non-stationary stochastic order commitments. This is also the reason for the narrowing ranges of CI_K with decreasing Δ_c (e.g., cluster I as compared to IV). That is; the lower the level of Δ_c, the smaller the cost advantage that can be obtained from a unit increase in the capacity slack Δ_K. The range of service level to be provided is essentially chosen by selecting the value of Δ_π. It is critical to follow that the menu contains disjoint sets of Δ_π levels. If the manufacturer desires to provide a higher service level to buyers, hence larger inventories, then he is allowed to those alternatives to the upper left.

Of course, more capacity is always better. But, the choice may be a function of capacity cost. Take two points A and B. We are interested in the marginal cost of investing in additional capacity when we are currently in A. This marginal cost of additional capacity should be compared with the benefits (i.e., slightly better service with a significant cost saving due to backorder savings) that could be obtained. That is the topic in §7.2.2.

7.2.2 Effects of building capacity slack

In this section, we further elaborate on the menu of alternative (Δ_K, Δ_c, Δ_π) combinations, Figure 7.6. By focusing more clearly on the production performance, this elaboration may assist the manufacturer in evaluating the implications of his preferences for the capacity level. The computational investigation in what follows is collected in two categories.

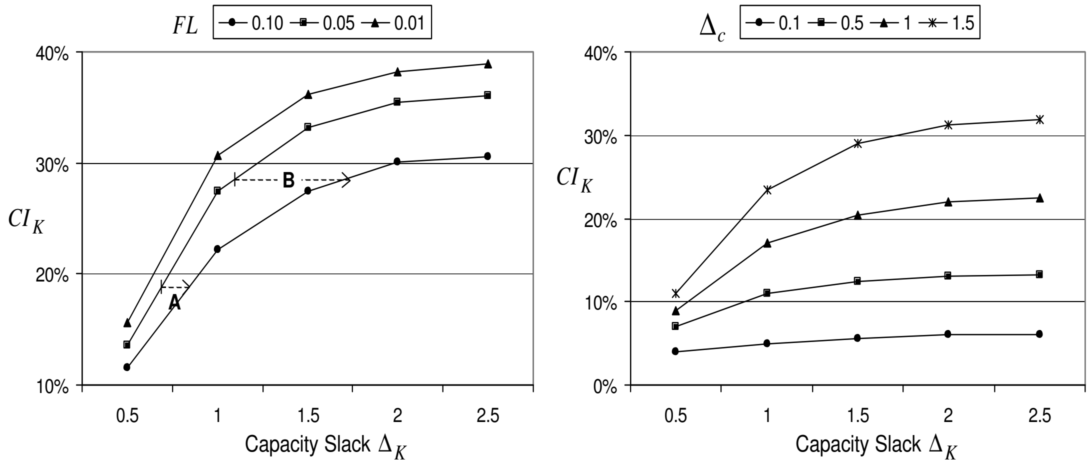
The first category illustrates the relationship between the capacity slack and the extent of flexibility limit per period (FL). The non-stationary stochastic nature of buyers' order commitments and the guaranteed supply up to a certain percentage under QF contracts require the manufacturer to build a certain level of excess capacity. Hence, we study various values of FL to represent different levels of demand uncertainty (risks in realized orders). A higher FL makes the manufacturer subject to a higher demand uncertainty through a wider range of orders. This allows us to explore the effectiveness of capacity slack, as another form (perhaps a complementary feature) of flexibility, under various levels of demand uncertainty.

The second category illustrates the relationship between the capacity slack and the cost structure (Δ_c, Δ_π). As we have already discussed in §6.2, the cost structure identifies the perceived relationship between the manufacturer's structural costs and the level of supply and customer service he can provide. The cost differential between the in-house production and subcontracting, $\Delta_c = c_{ps} - c_{pi}$, represents a major structural cost driver of the manufacturer. It reflects the technology available and the technical complexity of the production system (where the technical complexity is related with the efficiency of production processes and the costs of production factors). The backorder-to-holding cost ratio, $\Delta_\pi = \pi_b/h$, on the other hand, represents executional cost drivers. It concerns the manufacturer's attitude to stockout risk. This allows us to explore the effectiveness of capacity slack under various cost structures.

Figure 7.7 plots the cost improvements (CI_K) that could be attained by building certain levels of the capacity slack; two series of plots are displayed for different values of FL and Δ_c. The backorder-to-holding cost ratio is assumed to be high. As before, the CI_K values are measured relative to the capacity base case (i.e., when $\Delta_K = 0$).

Figure 7.7 (a) shows that the cost improvement is the highest when the in-house capacity is relatively loose ($\Delta_K = 2.5$) and the flexibility limit per period is restrictive ($FL = 0.01$). CI_K increases as Δ_K increases from 0.5 to 2.5 but showing a diminishing return. The higher the flexibility limit, the lower the cost improvement over the capacity base case. These results state that a higher capacity level does decrease the cost of the manufacturer, irrespective of the extent of flexibility limits stipulated in the QF contracts, indicating the value of capacity slack. However, the capacity slack is the most valuable when the buyers are allowed to order from a narrower range (i.e., lower FL). Table D.4 in Appendix D lists the solution details.

Looking at it from the other side, we can ask the question "which one is more effective; a lower FL or a larger Δ_K?". There arises a tradeoff between FL and Δ_K. To keep the same CI_K with a larger FL, Δ_K has to go up by smaller additions when the manufacturer has lower capacity levels (e.g., A), whereas relatively larger additions are needed under higher capacity levels (e.g., B). Consequently, lower FL (i.e., greater information content in advance order information due to smaller order range) is more effective for the cases when the manufacturer has high capacity levels, as a unit increase in FL corresponds to a larger Δ_K additions for the same cost advantage. This makes the effect of FL highly dependent on Δ_K levels. Gavirneni et al. (1999) provide similar results for a model where a manufacturer has limited capacity and uses point of sale data in production decisions. There exists uncertainty about timing and amount of the orders due to the retailer's (s, S) ordering policy. They demonstrate that the benefits of information decrease as capacity becomes tighter.



(a) Categorized by flexibility limit *FL*

(b) Categorized by cost differential Δ_c

Figure 7.7: Cost improvements CI_K by increasing Δ_K, categorized by FL and Δ_c

As shown in Figure 7.7 (b), we found that when Δ_π is high and $\Delta_c \simeq 0$, the CI_K values vary slightly with the capacity slack. This means that the benefits that could be obtained from the capacity flexibility are not so significant, when the manufacturer's subcontracting business is subject to serious competition. Nevertheless, even small gains would be valuable, making the manufacturer operating in such environments better off. On the other hand, when the cost differential Δ_c is relatively larger (e.g., $\Delta_c = 1.5$), the cost improvement does vary greatly with the capacity flexibility. This phenomenon explains that building capacity flexibility under QF contracts is of value only when there is a considerable amount of cost differential between the in-house production and subcontracting. Furthermore, if the backorder-to-holding cost ratio Δ_π becomes lower, then the CI_K curves shift upward (making the gains larger), and they become steeper (making the changes more sensitive). Table D.5 in Appendix D lists the solution details.

Figure 7.8 illustrates the relationship between the instability of optimal order-up-to positions (TI_{ins}) and the capacity slack, categorized by FL and Δ_c. Each TI_{ins} value represents an average absolute deviation between optimal order-up-to positions of consecutive decision periods, in fractions of the mean order per lead-time, (6.4). A lower capacity slack means that the manufacturer has a tendency of subcontracting more to satisfy uncertain orders. This leads to higher order-up-to instability as under tight capacity it is harder for the manufacturer to smooth out the decisions. As Δ_K becomes larger, the manufacturer is expected to require less subcontracting and hence the TI_{ins} value decreases. This implies less cost associated with period-to-period production variation.

Figure 7.8 (a) shows that TI_{ins} is the highest when FL is loose and it decreases as FL gets stricter. Table D.4 in Appendix D lists the solution details. The decrease rate of TI_{ins} over different Δ_K levels is steeper for $FL = 0.01$ than that for $FL = 0.05$, and it is much more steady for $FL = 0.10$. Similarly, we observed that a longer commitment horizon (i.e., larger H) leads to a higher order-up-to instability due to the accumulated variability of buyers' orders. These results reveal that requesting limited flexibility through QF contracts is significantly beneficial to production smoothing. A unit decrease in FL and/or H is more effective for the cases when the manufacturer has high capacity levels, as it amounts to a larger Δ_K reduction (e.g., A compared to B) for the same level of the order-up-to instability.

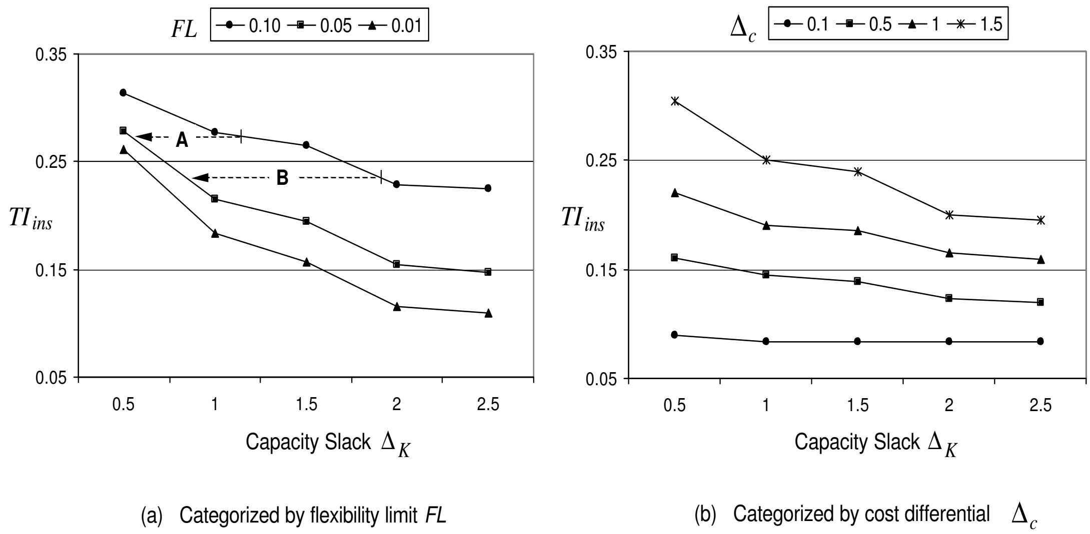


Figure 7.8: Order-up-to instability TI_{ins} by increasing Δ_K, categorized by FL and Δ_c

Figure 7.8 (b) shows that the instability of optimal order-up-to positions is inversely related to the cost differential in the in-house production and subcontracting, Δ_c. Table D.5 in Appendix D lists the solution details. When the cost differential

Δ_c is low, capacity availability is no longer a challenge, and increasing the capacity slack does not generate a significant difference as the manufacturer is not reluctant to substitute subcontracting. Obviously, if there is no restriction on the supply availability then it is less likely to observe successive backorder periods, hence the stability is not expected to vary greatly. Still there is a certain level of instability because the manufacturer's ordering has to be done so much in advance. We should also mention that extending the commitment horizon (i.e., larger H) yielded a higher order-up-to instability. TI_{ins} is increasing in H, with an increasing rate under a higher cost differential Δ_c. A unit decrease in Δ_c is less effective for the cases when the manufacturer has high capacity levels, as it amounts to a smaller Δ_K reduction for the same TI_{ins} level. Moreover, if the backorder-to-holding cost ratio Δ_π becomes lower, then the TI_{ins} curves shift downward, and they become less steep.

Finally, we may be concerned with the relationship between capacity utilization (CU) and commitment horizon for a given level of the capacity slack. CU relates the actual in-house productions resulting from the optimal policy to the potential amounts that could be produced with a given in-house capacity investment. Figure 7.9 provides the CU plots across various values of H, categorized by FL and Δ_c. In measuring the capacity utilization values, we assume that $\Delta_K = 1$ (i.e., there is an excess capacity being one standard deviation over the mean order per lead-time). Observe that the CU values vary slightly with H. Figure 7.9 (a) shows that CU is higher for a higher FL, and increases slightly as H increases from 3 to 12. The increase rate of CU depends on the level of FL more than it does on H, where more restrictive FL has lower rates. Likewise, Figure 7.9 (b) illustrates the relationship between CU and H, categorized by Δ_c. Tables D.6 and D.7 in Appendix D list the solution details.

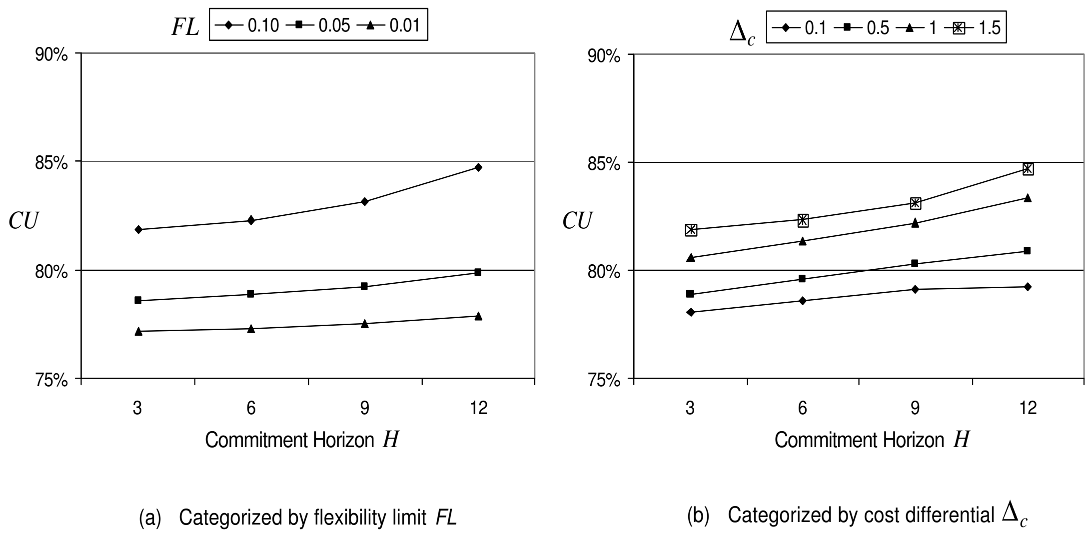


Figure 7.9: Capacity utilization CU by increasing H, categorized by FL and Δ_c

These results give valuable insights into the effects of changing the capacity level, and facilitate the choice of appropriate actions being tuned relative to some internal factors. As a consequence, we can identify some preferable actions. In particular, the manufacturer may prefer the action that provides a high fill-rate service level without sacrificing his cost advantage much and keeping the production pretty smooth. The buyers sacrifice some level of flexibility for that action, but in turn the manufacturer cares more about the backorders (by higher backorder penalty), and builds larger supply-side flexibilities (by higher capacity slack and less costly subcontracting option). These underscore the overwhelming role FL plays in the actions of both parties.

7.3 Analysis from the Analyst Perspective

Differing views of the manufacturer and his contract buyers are usually held by production and sales functions within the manufacturer's business. These may be categorized mainly as *production-focused* concerns versus *service-focused* concerns. This section aims to develop insights into the value of functional harmony that could be attained by an aggregated view of the production and sales functions and its information needs. This harmony is primarily concerned with how well the key information flow from the contract buyers, order commitment information, is utilized in the planning process of the manufacturer. This is articulated by comparatively evaluating the effectiveness of the ways in which the demand process could be modeled as an integral part of the production/inventory planning. Specifically, we evaluate the two alternatives: (*i*) the inventory model under the forecast evolution (MUFE) and (*ii*) the inventory model under the ARIMA process (MUMA), as mentioned before in §6.5.

These comparative alternatives differ in the extent of order commitments and the quality of the buyers' forecasting machinery in generating those order commitments. This relates to the worth of extra information, on the manufacturer side, and the worth of better forecasts (i.e., less variable commitment updates), on the buyers side. The claim laid by the analyst in using MUFE is that the manufacturer benefits significantly from using the model under the forecast evolution, hence the value of associated information gathering from forecast evolution and optimization is considerable. This is conjectured due to a better explanation of the demand uncertainty that is attained by taking into account the correlation and variability of commitment updates.

7.3.1 Comparing the two alternative inventory control models

In this section, we first quantify the ARIMA demand process used in developing the myopic replenishment policy (PUMA) given by (6.21). Then we will investigate how effective the optimal policy under the forecast evolution is as compared to the myopic policy for various system scenarios with different levels of order information.

The demand process was identified as a bivariate ARIMA(0, 1, 1) for two time series being modeled jointly corresponding to two contract buyers, as mentioned in Appendix C. In this bivariate model, the 2×2 matrix $\mathbf{\Theta}$ of moving average parameters and the 2×2 covariance matrix $\Sigma_{\boldsymbol{\eta}}$ of disturbance vector $\boldsymbol{\eta_s}$ are unknown. As we have already mentioned in §6.5.1, to develop estimators for these unknown parameters, we first calculate the target auto- and cross-covariances of the demand process, [cf. Eqs. (6.12) and (6.13)].

For the purposes of calculating these target covariances, suppose the correlation structure inherent in the system has been represented by a 26×26 covariance matrix

$\Sigma_{\tilde{\mathcal{E}}}$ of intended commitment updates [2]. It is better to interpret the covariance matrix $\Sigma_{\tilde{\mathcal{E}}}$ in the form of correlation coefficients. They represent the correlations of intended commitment updates occurring in the same period of time, and are of two classes. (*i*) Correlations for a particular buyer $b \in \{1, 2\}$ are all assumed to be positive. The further apart the commitment updates are, the smaller the correlations between them. They are assumed to take values in $\{0.8, 0.5, 0.3\}$. (*ii*) Correlations across buyers for the commitment updates made for the same future period are taken to be negative, $\{-0.8, -0.5, -0.3\}$. But, we assume that k-period ahead commitment updates from a buyer are uncorrelated with l-period ahead updates from the other buyer for $k \neq l \in \{1, 2, \ldots, H+1\}$. Table D.8 in Appendix D contains the associated 26×26 correlation matrix of intended commitment updates. The variance of these intended commitment updates are taken to be $\sigma^2_{\tilde{\varepsilon}^1_k} = 4$ and $\sigma^2_{\tilde{\varepsilon}^2_k} = 2$ for all $k \in \{1, 2, \ldots, H+1\}$. Therefore, we used the expressions (6.12) and (6.13), and the resultant target auto- and cross-correlation functions for the bivariate ARIMA(0, 1, 1) demand process were calculated as shown in Figure 7.10.

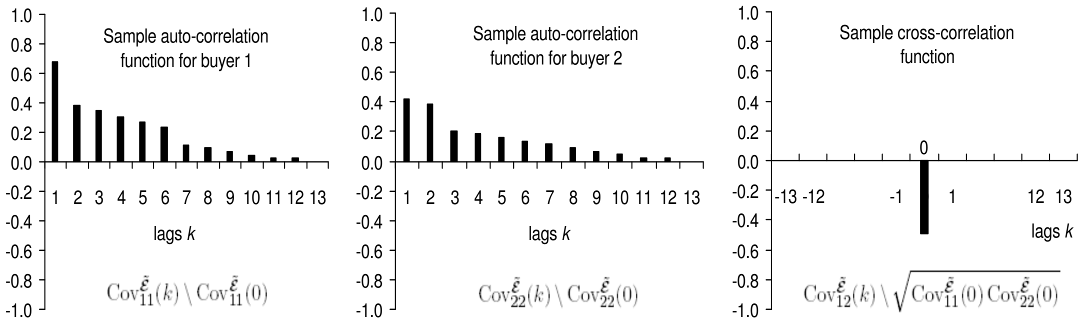


Figure 7.10: The target auto- and cross-correlation functions

Consequently, given the target correlation functions [cf. Fig. 7.10], we solved the optimization program $\text{PAR}_{\boldsymbol{\Theta},\boldsymbol{\eta}}$, given by (6.14) - (6.19), for the unknown moving average parameters $\boldsymbol{\Theta}$ and disturbance covariances $\Sigma_{\boldsymbol{\eta}}$. The optimization program suggested a bivariate ARIMA(0, 1, 1) model of the form

[2] Note that the dimensionality of the covariance matrix $\Sigma_{\tilde{\mathcal{E}}}$ depends on the number of buyers, $B = 2$, and the length of the commitment horizon, $H = 12\ (= L)$ such that $26 = B \times (H+1)$.

$$
\begin{aligned}
D^b_{1,1} &= \mu_{D^b} + \eta^b_1 \qquad \text{for all buyers } b \in \{1,\, 2\}, \quad \text{and} \\
D^1_{s,s} &= D^1_{s-1,s-1} - 0.32\,\eta^1_{s-1} - 1.07\,\eta^2_{s-1} + \eta^1_s \\
D^2_{s,s} &= D^2_{s-1,s-1} + 1.01\,\eta^1_{s-1} - 0.06\,\eta^2_{s-1} + \eta^2_s
\end{aligned} \tag{7.1}
$$

for $s = 2,\, 3,\, \ldots,\, N + L$. We have $\mu_{D^1} = 30$ and $\mu_{D^2} = 20$. The estimated matrix of moving average parameters was found to be $\boldsymbol{\Theta} = \left[\begin{smallmatrix} 0.68 & -1.07 \\ 1.01 & 0.94 \end{smallmatrix}\right]$, and the vector of disturbances $\boldsymbol{\eta_s}$ was characterized by the estimated covariance matrix $\Sigma_{\boldsymbol{\eta}} = \left[\begin{smallmatrix} 14.57 & -8.39 \\ -8.39 & 8.11 \end{smallmatrix}\right]$.

Note that the estimated covariances $\Sigma_{\boldsymbol{\eta}}$ in (7.1) are *uncensored* [cf. Eq. 6.20], as revision limits are not accommodated in the estimation process. As we have already mentioned in §6.5.2, we modify the estimated covariances $\Sigma_{\boldsymbol{\eta}}$ for incorporating those revision limits. To illustrate, suppose MUFE with $H = 3$ and $FL = 0.10$. Then, the estimated covariances in (7.1) turn out to be $\Sigma_{\boldsymbol{\eta}} = \left[\begin{smallmatrix} 9.75 & -0.05 \\ -0.05 & 3.61 \end{smallmatrix}\right]$. The replenishment order quantities under PUMA [cf. Eq. 6.21] are adjusted accordingly.

We now investigate the implications of varying the commitment horizon on the effectiveness of the MUFE as compared to the MUMA. We study five different levels of the commitment horizon from $H = 1$ to $H = 12$ for the MUFE, where the level of flexibility limit per period is taken to be $FL = 0.10$. For the MUMA, on the other hand, we have variations of the estimated ARIMA process (7.1) corresponding to these five different levels of the commitment horizon.

We assume the same environmental settings across the alternative inventory control models. These are summarized in Table 6.1. Also, we initiate the same initial inventory position, I_0. For the purposes of comparability, we assume that the cost differential between the in-house production and subcontracting, Δ_c, is negligible. This makes the inventory system uncapacitated as in Graves (1999).

We evaluate alternative models in terms of four performance measures; the cost improvement (CI_{model}), the mean order-up-to deviation (TI_{dev}), the order-up-to instability (TI_{ins}), and the fill-rate service level (φ). CI_{model} is the relative cost benefit that could be attained by using the MUFE for a certain flexibility arrangement as compared to the corresponding myopic policy. TI_{dev}, TI_{ins} and φ are calculated for both MUFE and MUMA in exactly the same manner as in (6.3), (6.4) and (6.6), respectively.

Figure 7.11 plots the cost improvements CI_{model} for various levels of the commitment horizon H. CI_{model} values are calculated for MUFE with $H = k$ and $FL = 0.10$ against MUMA with $H = 1$ and the same allowable range for buyers' orders (as an accumulation of the flexibility limit per period, FL, over $H = k$). Step 3 in §6.5.2 discussed this in more detail. MUFE has cost advantages since the forecast evolution mechanism allows the buyers to update their commitments in a gradual manner by decomposing total variability. When the total variance is decomposed into smaller terms, the manufacturer is able to react more efficiently to buyers' stochastic orders by adjustments in target inventory in time. This is because he does not have to rely on costly reactions to last-minute order updates, which would be likely to cause inefficiency and extra costs. Early indications and restricted distortions due to the decomposed variance cause this.

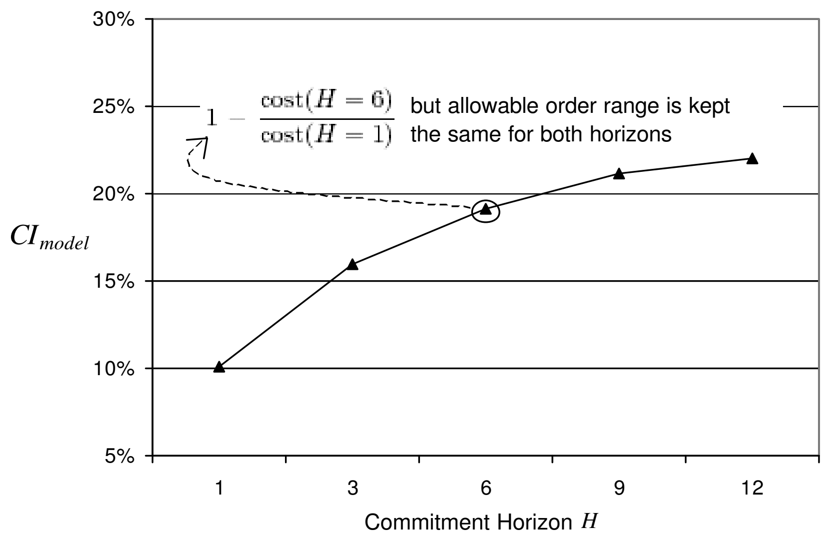


Figure 7.11: Cost improvements CI_{model} of MUFE against MUMA by increasing H

Figure 7.11 shows that the cost improvement is the lowest when the commitment horizon for MUFE is one-period long ($H = 1$). Table D.9 in Appendix D lists the solution details. CI_{model} values range from 10.1% to 22% as moving towards the largest horizon of $H = 12$, but showing a diminishing return. If the level of flexibility limit per period becomes less stricter than $FL = 0.10$, then the CI_{model} curve shifts upward (making the gains larger), and it becomes more steep (making the changes more sensitive). It is also found that the higher the level of flexibility limit per period, FL, the higher the portion of variability captured by the forecast evolution in advance. We observe that this makes the difference between the expected total costs of the optimal and myopic policies larger. These results state that the MUFE definitely results in lower expected total costs, indicating the value of associated information gathering from forecast evolution and optimization. As the buyers somehow commit earlier (i.e., larger H), the difference between the expected total costs gets larger.

The manufacturer often does not have direct control over the environment that his buyers face. Nevertheless, the performance of the production/inventory system may be influenced by variation in the operating environment. The most important uncontrollable but influencing environmental features in our case are the correlation of demand (CO) across buyers and the coefficient of variation (CV) for the buyers' orders. Different levels of them may characterize different industrial contexts in which the manufacturer and his contract buyers operate. For instance, the phase of the product life cycle may account for different values of CV (e.g., maturity phase for low CV and innovation phase for high CV). The degree of seasonality in demand may give rise to radically different values of CO (e.g., negative CO for inverse seasonal demand patterns across buyers).

We experiment with various levels of these environmental factors to understand how the cost improvements CI_{model} are affected by the commitment horizon. It is found that a lower CV value results in a lower cost improvement irrespective of the length of commitment horizon (i.e., a decrease in the benefit from using MUFE). When the commitment horizon is not long relative to the replenishment lead-time, extending the commitment horizon has a significant impact on the improvement of the cost (i.e., relatively greater benefits are obtained from using MUFE). As H extends further, the effect on the improvement of cost shows a diminishing return. These results say that the higher the uncertainty in the operating environment, the more valuable for the manufacturer to exercise the MUFE. This is mainly because FL's restrict realized orders $d^b_{s,s}$ to the same region no matter what CV is. The only exception is the first time $d^b_{s,s+H-1}$ is updated freely from μ_{D^b}. When buyers somehow commit earlier, the manufacturer is able to react more efficiently to their non-stationary stochastic order commitments by using the MUFE. This is because he does not have to rely on costly reactions to last-minute order updates, which would be likely to cause inefficiency and extra costs.

We found that the order-up-to positions under both inventory models increase when the order commitments from buyers are more positively correlated. This leads to an increase in the expected total cost for both models, but CI_{model} has lower values (i.e., the difference between the MUFE and MUMA gets smaller). When CO decreases towards -1, the expected total costs both decrease but the MUFE experience larger reductions, hence the CI_{model} plot shifts upward. This is due to the proliferation of deviations when $CO > 0$ and the substitution effects when $CO < 0$. These results indicate the benefit from integrating the evolution of order commitments with the manufacturer's production and inventory planning. The value of committing earlier increases as orders are more volatile and/or more negatively correlated.

Figure 7.12 (a) and (b) provide the plots of the mean order-up-to deviation, TI_{dev}, and the order-up-to instability, TI_{ins}, respectively, for various levels of the commitment horizon. Each figure shows two curves corresponding to the two alternative inventory control models. For the MUMA curves, $H = 1$ is kept the same in all trials. For the MUFE curves, on the other hand, $H = k$ is varied in trials. As before, the allowable ranges of buyers' orders (i.e., as an accumulation of the flexibility limit per period, FL) are kept the same across the alternatives given a certain level of the commitment horizon.

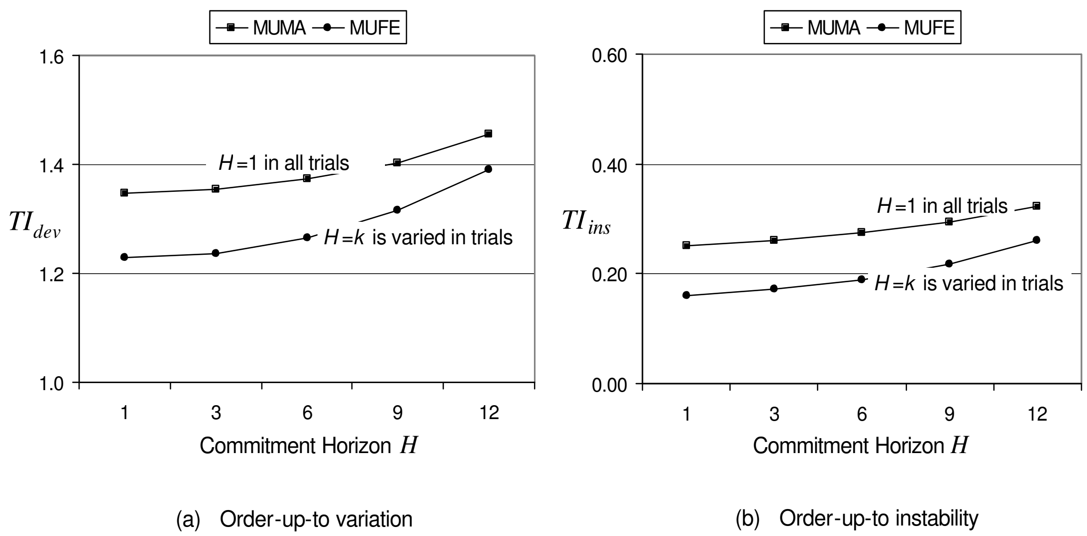


Figure 7.12: Mean order-up-to deviation TI_{dev} and order-up-to instability TI_{ins} by increasing H, categorized by alternative inventory models

Both TI_{dev} and TI_{ins} are increasing in H (i.e., earlier order commitments). However, they take higher values with a relatively low variation for MUMA, irrespective of the length of commitment horizon. In evaluating these results note that the level of flexibility limit per period is taken to be $FL = 0.10$. Hence with a higher H, the manufacturer must respond to wider range of orders each time due to the increased cumulative revision flexibility. Large TI_{dev} and TI_{ins} values for MUMA (i.e., additional inventory with higher variation) are experienced due to insufficient reactions to last-minute updates in realized orders causing abrupt changes in the inventory state. All these results reveal that (i) practicing early order commitments and (ii) modeling their time series through the forecast evolution mechanism help to smooth out the manufacturer's inventory levels.

Figure 7.13 allows us to see the service disparities between the two alternative inventory control models as the commitment horizon varies. The fill-rate service level is given for the aggregated demand being a composite of all the buyers. As before, the allowable ranges of buyers' orders are kept the same across alternative inventory models at a certain H. The figure states that a given level of commitment horizon can result in different levels of customer service under the two inventory models, depending on how the total demand variability is distributed over the commitment horizon. The MUFE is less sensitive to an increased uncertainty (caused by wider range of orders through larger H), whereas the MUMA is not able to respond efficiently to the changes in the demand variability.

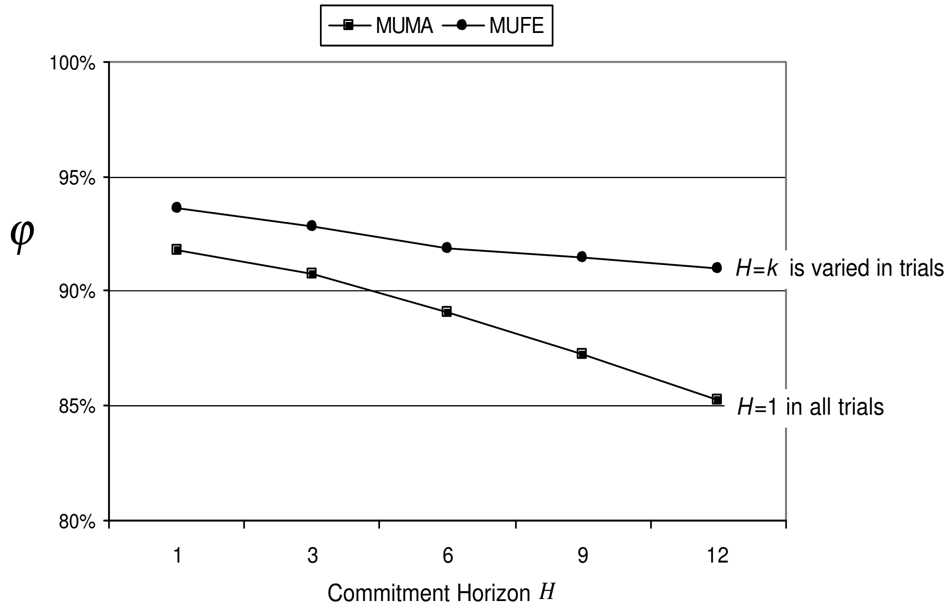


Figure 7.13: Fill-rate service level φ by increasing H, categorized by alternative inventory models

We conclude that the production/inventory planning under the forecast evolution can bring improvements in the inventory investment, and improves the customer service slightly. We demonstrate that the myopic policy under the ARIMA process results in higher order-up-to positions and higher expected total costs.

7.4 Summary

In this chapter we have discussed the computational results and insights to our research questions [cf. §6.1]. This was performed by emphasizing three perspectives [cf. Fig 7.1] that surround these research questions.

The buyer perspective concerns the parameter-setting problem of QF contracts, addressing the first research question. We proposed a menu approach in which each contract buyer is offered various commitment and flexibility arrangements. We demonstrated that extended advance order commitments for a given level of flexibility per period results in lower cost savings against the minimal-commitment case and larger cost increases against the minimal-flexibility case for the manufacturer. This is accompanied by a greater mean order-up-to deviation caused by the manufacturer's overestimating the buyers' future orders in response to a wider range of orders. The inventory holding cost thus becomes the dominant cost component while the share of the backorder cost diminishes. This means improved services to the buyers, and in turn leads to lower risks of the buyers' experiencing a stockout. We also demonstrated that the cost performance is inversely related to the level of revision flexibility. At high flexibility levels, the manufacturer obtains almost no benefit from the buyers' committing early since he must respond to a wider range of orders each time period. This reduces the value of practicing extended advance order commitments.

The manufacturer perspective centers on the second and third research questions. We proposed a menu of various capacity levels which differ in the backorder-to-holding cost ratio, Δ_π, and the cost differential between the in-house production versus subcontracting, Δ_c. We examined the marginal cost of investing in additional capacity as compared to the benefits that could be obtained. We observed slightly better service with a significant cost saving due to backorder savings. When Δ_π is lower, the cost advantage that can be obtained from an additional capacity investment gets larger. The lower the level of cost differential Δ_c, on the other hand, the smaller the cost advantage from an additional capacity investment. This indicates that when the manufacturer's subcontracting business is subject to serious competition, the value of capacity slack gets smaller. We demonstrated that additional capacity is the most valuable when the buyers are allowed to order from a narrower range. But, there arises a tradeoff between revision flexibility and capacity slack. We examined production smoothing and in-house capacity utilization as well. We demonstrated that a unit decrease in Δ_c is less effective for the cases when the manufacturer has high capacity levels, as it amounts to a smaller capacity reduction for the same level of production smoothing. We observed that the capacity utilization increases with extended advance order commitments and larger revision flexibility.

The analyst perspective is related with the fourth research question. As an effort to benchmark, we evaluated two alternative approaches; the evolution-based inventory management and the ARIMA-based inventory management. We demonstrated that the first approach has cost advantages since when the total variance is decomposed into smaller terms by the forecast evolution mechanism, the manufacturer is able to get early indications and restricted distortions.

CHAPTER 8

CONCLUSION

This research has examined the role of early order commitments in providing manufacturing companies with incentives for efficient inventory management under quantity flexibility (QF) contracts. Early order commitments with revision flexibility ensure systematic order information is received by the manufacturer in every period. This is a form of advance order information. In this context, we suggested a general approach for integrating a probabilistic model of the changes in the committed orders with an analytical model of the production and inventory planning under multi-period QF contracts. We demonstrated that this integration permit more effective management of the production and inventory system. Such an integrated use of refined commitment update scheme differs considerably from the literature related to contracting and inventory management.

We introduced our approach in a general problem environment under a decentralized decisional structure. A single capacitated manufacturer with an option of subcontracting produces and sells a single item to multiple buyers under multi-period QF contracts. The buyers face stochastic market demands in every period, and submit unreliable order commitments (caused especially by distortions) to the manufacturer on a range of future periods. The committed orders evolve over time. The associated commitment updates are not independent of one another, implying correlations of the updates across buyers and through time. This constitutes the key information flow in the planing process of the manufacturer. The problem environment is a complicated one, but appears worthwhile in the view of the interplay among factors forming the basis of effective decision making.

In this context, we analyzed interplay of a number of factors regarding the decisions within the aggregate production/inventory planning framework. We assumed a multi-period stochastic production/inventory decision model to underlie the analysis. We adopted a computational approach using an analytical optimization rather than simulation. The model minimizes the expected total cost of the manufacturer, directed toward exploring the relationships between the experimental factors of interest and the measures of system performance. The computational study thus was performed to delineate ways that this model might be elaborated to capture various important features of the production and inventory system.

The research differs significantly from the related literature in several aspects: (*i*) We introduce a refined commitment update scheme as an integral part of the production and inventory planning under QF contracts. The integration permits an enhanced production and inventory planning that is informed of how the order commitments evolve from one period to another. The associated decision model also facilitates determining the QF contract terms and conditions. (*ii*) We develop a modified MMFE as it accommodates revision limits stipulated in the QF contracts. We then introduce a finite Markov chain approximation to this modified martingale forecast evolution process to model the probabilistic framework of our dynamic decision model. It accommodates contract revision limits and correlation structure in the buyers' ordering. By this approximation we provide a novel approach to discretization in stochastic dynamic programming, and an estimation scheme for non-stationary stochastic dynamics. (*iii*) We characterize the manufacturer's optimal replenishment policy as a staircase, state-dependent order-up-to policy using this finite stochastic framework. (*iv*) We implement computational dynamic programming as a solution technique. We develop an efficient approach for reducing state-space dimensionality building upon our forecast evolution structure. This makes the computation associated with the stochastic recurrence relations much less demanding. (*v*) Finally, as an effort to benchmark, comparison of the evolution-based inventory management with an ARIMA-based approach was performed.

We provided results on three perspectives to evaluate concerns of the manufacturer and his contract buyers. These perspectives surround the research questions addressed through the computational study.

First, we investigated the effects, on the buyers, of the manufacturer's decisions about the QF contracts. We proposed a menu of various commitment and flexibility arrangements. Menus are for the manufacturer's indifference among combinations that the buyers are not indifferent. They assist the buyers in selecting appropriate scheme of QF contracts. The main incentive for the buyer's choosing among the alternatives is to minimize her risk of experiencing a stockout in servicing the market. The buyers are offered more rewards (e.g., price discounts, priority etc.) or higher revision flexibility (i.e., passing on some portion of the cost associated with uncertainty to the manufacturer) in motivating them to provide the desired level of extended early commitments.

We examined the implications of the buyers' preferences by means of the manufacturer's cost performance versus the extent of information content in early order commitments. We demonstrated that earlier commitments for a given flexibility per period result in lower cost savings against the minimal-commitment case and larger cost increases against the minimal-flexibility case. A higher revision flexibility reduces the value of practicing extended advance order commitments. The inventory holding cost gets dominance while the share of the backorder cost diminishes caused by the associated increase in the mean order-up-to deviation. This means improved services to the buyers, and in turn leads to lower risks of the buyers' experiencing a stockout.

Second, we are concerned with the manufacturer's capacity investment decisions, and made an integrated analysis with the operational aspects of the QF contracts. We provided results on how effective the level of capacity slack is as compared to the flexibility of the QF contracts and the attractiveness of the subcontracting option. We proposed a menu of various capacity levels which differ in the backorder-to-holding cost ratio and the cost differential between the in-house production and subcontracting, Δ_c. Such a menu serves as a decision aid, and includes possible actions with very close (if not identical) performance outcomes.

We examined the implications of the manufacturer's preferences by means of the cost improvements and the service level that can be attained by a particular choice of capacity level. We evaluated the marginal cost of investing in additional capacity as compared to the benefits that could be obtained, and observed slightly better service with a significant cost saving due to backorder savings. The manufacturer prefers the action that provides a high fill-rate service level without sacrificing his cost advantage much and keeping the production pretty smooth. The buyers sacrifice some level of flexibility for that action, but in turn the manufacturer cares more about the backorders (by higher backorder penalty), and builds larger supply-side flexibilities (by higher capacity slack and less costly subcontracting option). These underscore the overwhelming role FL plays in the actions of both parties. We thus examined the worth of revision flexibility in the face of capacity. The capacity slack is the most valuable when the buyers are allowed to order from a narrower range (i.e., lower revision flexibility). Lower revision flexibility indicates greater information content in advance order commitments due to smaller order range. Reducing the revision flexibility is more effective for the cases when the manufacturer has high capacity levels, as a unit decrease in revision flexibility corresponds to a larger capacity reduction for the same cost advantage.

We also demonstrated that building capacity flexibility under QF contracts is of value only when there is a considerable amount of cost differential between the in-house production and subcontracting. We demonstrated that there is a reciprocal relationship between cost differential Δ_c and capacity slack Δ_K. The lower the level of Δ_c, the smaller the cost advantage that can be obtained from a unit increase in Δ_K. We also provided results on production smoothing and capacity utilization levels. We showed that requesting limited flexibility through QF contracts improves production smoothing. The capacity utilization, on the other hand, depends on the level of revision flexibility more than it does on the extent of order commitments (i.e., commitment horizon), where more restrictive flexibility has lower rates.

Third, we tried to benchmark our refined commitment update scheme. This was articulated by comparatively evaluating two alternative inventory control models; the model under the forecast evolution and the model under the ARIMA demand process. These represent different ways of modeling advance order information as an integral part of production and inventory planning. We demonstrated that the manufacturer benefits significantly from using the forecast evolution framework. The inventory model under the ARIMA process resulted in higher order-up-to positions and higher expected total costs without an improvement in the service level. The evolution-based model has cost advantages since the forecast evolution mechanism allows the buyers to update their commitments in a gradual manner by decomposing the total variability. When the total variance is decomposed into smaller terms, the manufacturer is able to react more efficiently to buyers' stochastic orders by adjustments in target inventory in time. Furthermore, the higher the portion of variability captured, the greater the benefits of information gathering from forecast evolution and optimization. This is because the sensitivity to larger variation reduces by spreading it over periods. The benefits from the decomposed variance show the adjusting role of the flexibility level coupled with early indications and restricted distortions.

An interesting direction for further research would be the development of a buyer's side modeling. This would allow us to perform incentive calculation for the buyers' cooperation. We in turn would look for a centralized mechanism for allocating system benefits among the manufacturer and his contract buyers, and study the implications of the evolution-based inventory management at the system level. Also, it would be desirable to study contractual relation with the subcontractor. This would permit more comprehensive treatment of capacity issues. The major difficulty that lie in these is the representation of stochastic parameters. This may also defy the use of optimization based-approach.

Another interesting avenue for further research would be to study some type of inventory rationing policy (instead of using our simple allocation policy where all order information are handled in the same way). Accordingly, the manufacturer would hold back inventory for future needs of the highest-priority buyer. This may allow him to make better use of available order information to allocate inventory to contract buyers. The resultant inventory model would permit to handle different demand classes, by defining different allocation functions.

As for computational analysis, we would study how much revision flexibility the manufacturer can tolerate in negotiating with contract buyers for the commitment horizon. The manufacturer may concern how much increase he can endure in revision flexibility as the commitment horizon is shortened for a fixed cost improvement. We would also study how the accumulation of censored commitment updates affects the computational results as it alters the associated distributional forms (spreading probability mass in between the two censoring points over allowable range of the associated random sum).

REFERENCES

[1] Anupindi, R. and Bassok, Y. (1998). Supply Contracts with Quantity Commitments and Stochastic Demand. In Tayur, S., Magazine, M. and Ganeshan, R., editors, *Quantitative models for supply chain management* (chapter 7), Kluwer Academic Publishers.

[2] Aviv, Y. (2003). A Time-Series Framework for Supply-Chain Inventory Management. *Operations Research*, 51(2), 210-227.

[3] Azoury, K.S. (1985). Bayes Solution to Dynamic Inventory Models under Unknown Demand Distribution. *Management Science*, 31(9), 1150-1160.

[4] Barr, D.M. and Davidson, T.A. (1973). Kolmogorov-Smirnov Test for Censored Samples. *Technometrics*, 15(4).

[5] Box, G.E.P., Jenkins, G.M. and Reinsel, G.C. (1994). *Time Series Analysis Forecasting and Control*, 3rd Ed. Holden-Day, San Francisco, CA.

[6] Bradley, J.R. (2004). A Brownian Approximation of a Production-Inventory System with a Manufacturer that Subcontracts. *Operations Research*, 52(5), 765-784.

[7] Chen, F. and Song, J.S. (2001). Optimal Policies for Multiechelon Inventory Problems with Markov-Modulated Demand. *Operations Research*, 49(2), 226-234.

[8] Çakanyıldırım, M. and Roundy, R. (2002). SeDFAM: Semiconductor Demand Forecast Accuracy Model. *IIE Transactions*, 34(5), 449-465.

[9] Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1952a). The Inventory Problem. I. Case of Known Distribution of Demand. *Econometrica*, 20, 187-222.

[10] Dvoretzky, A., Kiefer, J., and Wolfowitz, J. (1952b). The Inventory Problem. II. Case of Unknown Distributions of Demand. *Econometrica*, 20, 450-466.

[11] Eppen, G. and Iyer, A. (1997). Backup Agreements in Fashion Buying: The Value of Upstream Flexibility. *Management Science*, 43(11), 1469-1484.

[12] Feng, Q., Sethi, S.P., Yan, H. and Zhang, H. (2006). Are Base-stock Policies Optimal in Inventory Problems with Multiple Delivery Modes? *Operations Research*, 54(4), 801-807.

[13] Gallego, G. and Özer, Ö. (2001). Integrating Replenishment Decisions with Advance Demand Information. *Management Science*, 47(10), 1344-1360.

[14] Gavirneni, S., Kapuscinski, R. and Tayur, S. (1999). Value of Information in Capacitated Supply Chains. *Management Science*, 45(1), 16-24.

[15] Gerchak, Y. and Wang, Y. (2004). Revenue-sharing vs. Wholesale-price Contracts in Assembly Systems with Random Demand. *Production and Operations Management*, 13(1), 23-33.

[16] Graves, S.C., Meal, H.C., Dasu, S. and Qui, Y. (1986). Two-stage Production Planning in a Dynamic Environment. *In Lecture Notes in Economics and Mathematical Systems, Multi-Stage Production Planning and Inventory Control*, edited by S. Axsater, C. Schneeweiss, E. Silver, Springer-Verlag, Berlin, 266, 9-43.

[17] Graves, S.C. (1999). A Single-Item Inventory Model for a Nonstationary Demand Process. *Manufacturing & Service Operations Management*, 1(1), 50-61.

[18] Greene, W.H. (2000). *Econometric Analysis.* 4th ed., Prentice Hall, Upper Saddle River, New Jersey.

[19] Güllü, R. (1993). Analysis of the Production/Inventory Policies under the Martingale Model of Forecast Evolution. Unpublished Ph.D. dissertation, Cornell University, Ithaca, NY.

[20] Güllü, R. (1996). On the Value of Information in Dynamic Production Inventory Problems under Forecast Evolution. *Naval Research Logistics*, 43, 289-303.

[21] Güllü, R. (1997). A Two-echelon Allocation Model and the Value of Information under Correlated Forecasts and Demands. *European Journal of Operational Research*, 99, 386-400.

[22] Hausman, W.H. (1969). Sequential Decision Problems: A Model to Exploit Existing Forecasters. *Management Science*, 16(2), 93-111.

[23] Hausman, W.H. and Peterson, R. (1972). Multi-product Production Scheduling for Style Goods with Limited Capacity, Forecast Revisions and Terminal Delivery. *Management Science*, 18(7), 370-383.

[24] Heath, C.D. and Jackson, L.P. (1994). Modeling the Evolution of Demand Forecasts with Application to Safety Stock Analysis in Production/Distribution Systems. *IIE Transactions*, 26(3), 17-30.

[25] Iida, T. and Zipkin, P.H. (2006). Approximate Solutions of a Dynamic Forecast-Inventory Model. *Manufacturing & Service Operations Management*, 8(4), 407-425.

[26] Jenkins, G.M. and Alavi, A.S. (1981). Some Aspects of Modeling and Forecasting Multivariate Time Series. *Journal of Time Series Analysis*, 2(1), - .

[27] Kaminsky, P. and Swaminathan, J.M. (2004). Effective Heuristics for Capacitated Production Planning with Multi-period Production and Demand with Forecast Band Refinement. *Manufacturing & Service Operations Management*, 6, 184-194.

[28] Karaesmen, F., Buzacott, J.A. and Dallery, Y. (2002). Integrating Advance Order Information in Make-to-stock Production Systems. *IIE Transactions*, 34, 649-662.

[29] Karlin, S. (1958). Optimal Inventory Policies for the Arrow-Harris-Marschak Dynamic Model. In K.J. Arrow, S. Karlin, and H. Scarf (eds.), *Studies in the Mathematical Theory of Inventory and Production.* Stanford University Press, Stanford, CA.

[30] Kayhan, M., Erkip, N. and Güllü, R. (2005). Integrating Forecast Evolution with Production-Inventory Planning: An Implementation Framework for a Fast Moving Consumer Goods Manufacturer. Technical Report, Industrial Engineering Department, Middle East Technical University.

[31] Kendall, M. and Stuart, A. (1979). *The Advanced Theory of Statistics, Volume 2*. 4th ed., Griffin, London.

[32] Lariviere, M.A. and Porteus, E.L. (2001). Selling to the Newsvendor: An Analysis of Price-only Contracts. *Manufacturing & Service Operations Management*, 3(4), 293-305.

[33] Law, M.A. and Kelton, D.W. (2000). *Simulation Modeling and Analysis*. McGraw-Hill, New York.

[34] Lee, H.L., Padmanabhan, V. and Whang, S. (1997). Information Distortion in a Supply Chain: The Bullwhip Effect. *Management Science*, 43(4), 546-558.

[35] Lee, H.L., So, K.C. and Tang, C.S. (2000). The Value of Information Sharing in a Two-level Supply Chain. *Management Science*, 46(5), 626-643.

[36] Loukas, S. and Kemp, C.D. (1986). On the Chi-square Goodness-of-fit Statistic for Bivariate Discrete Distributions. *The Statistician*, 35(5), 525-529.

[37] Muth, J.F. (1960). Optimal Properties of Exponentially Weighted Forecasts. *American Statistical Association Journal*, 55, 299-306.

[38] Rose, C. (1994). A Statistical Identity Linking Folded and Censored Distributions. *Journal of Economic Dynamics and Control*, 19(1995), 1391-1403.

[39] Ross, S.M. (1997). *Introduction to Probability Models*. 6th ed., Academic Press, San Diego.

[40] Sánchez, A.M. and Pérez, M.P. (2005). Supply Chain Flexibility and Firm Performance: A Conceptual Model and Empirical Study in the Automotive Industry. *International Journal of Operations & Production Management*, 25(7), 681-700.

[41] Scarf, H. (1959). Bayes Solution of the Statistical Inventory Problem. *Annals of Mathematical Statistics*, 30, 490-508.

[42] Sethi, S.P., and Cheng, F. (1997). Optimality of (s,S) Policies in Inventory Models with Markovian Demand. *Operations Research*, 45, 931-939.

[43] Sethi, S.P., Yan, H. and Zhang, H. (2004). Quantity Flexibility Contracts: Optimal Decisions with Information Updates. *Decision Sciences*, 35(4), 691-712.

[44] Stevenson, M. and Spring, M. (2007). Flexibility from a Supply Chain Perspective: Definition and Review. *International Journal of Operations & Production Management*, 27(7), 685-713(29).

[45] Tan, T. and Alp, O. (2008). An Integrated Approach to Inventory and Flexible Capacity Management subject to Fixed Costs and Non-stationary Stochastic Demand. *OR Spectrum*.

[46] Toktay, L.B. and Wein, L.M. (2001). Analysis of a Forecasting-Production-Inventory System with Stationary Demand. *Management Science*, 47(9), 1268-1281.

[47] Tsay, A. (1999). The Quantity Flexibility Contract and Supplier-Customer Incentives. *Management Science*, 45, 1339-1358.

[48] Tsay, A. and Lovejoy, W.S. (1999). Quantity Flexibility Contracts and Supply Chain Performance. *Manufacturing & Service Operations Management*, 1(2), 89-111.

[49] Tsay, A., Nahmias, S., and Agrawal, N. (1999). Modeling Supply Chain Contracts: A Review. In Tayur, S., Ganeshan, R., & Magazine, M. (editors), *Quantitative Models for Supply Chain Management*. Norwell, MA, Kluwer Academic Publishers, 299-336.

[50] Vickery, S., Calantone, R. and Dröge, C. (1999). Supply Chain Flexibility: An Empirical Study. *Journal of Supply Chain Management*, 35(3), 16-24.

[51] Williams, D. (1991). *Probability with Martingales*. Cambridge University Press.

[52] Wu, J. (2005). Quantity Flexibility Contracts under Bayesian Updating. *Computers & Operations Research*, 32, 1267-1288.

APPENDIX A

PROOFS

A.1 Proof of Proposition 4.1

We have for $s = 1, 2, \ldots, N$

$$\begin{aligned} J_s(TI_s,\, I_s,\, \mathbf{d_s}) &= PC_s(TI_s,\, I_s) + L_s(TI_s,\, \mathbf{d_s}) \\ &= c_{pi}(TI_s - I_s) + (c_{ps} - c_{pi})(TI_s - I_s - K)\mathbf{1}(A_s) \\ &+ h \sum_{z_1=0}^{TI_s} \sum_{z_2=0}^{TI_s - z_1} (TI_s - z_1 - z_2)\, f_{Z_L^2}(z_2|\,\mathbf{d_s})\, f_{Z_L^1}(z_1|\,\mathbf{d_s}) \\ &+ \pi_1 \sum_{z_1=TI_s}^{\infty} (z_1 - TI_s)\, f_{Z_L^1}(z_1|\,\mathbf{d_s}) + \pi_2 \sum_{z_1=TI_s}^{\infty} \sum_{z_2=0}^{\infty} z_2\, f_{Z_L^2}(z_2|\,\mathbf{d_s})\, f_{Z_L^1}(z_1|\,\mathbf{d_s}) \\ &+ \pi_2 \sum_{z_1=0}^{TI_s} \sum_{z_2=TI_s - z_1}^{\infty} (z_2 - (TI_s - z_1))\, f_{Z_L^2}(z_2|\,\mathbf{d_s})\, f_{Z_L^1}(z_1|\,\mathbf{d_s}). \end{aligned} \tag{A.1}$$

Note that $L_s(TI_s,\, \mathbf{d_s})$ is known to be continuous [cf. Eq. 4.5] although we apply our finite Markov approximation introduced in §3.2 to §3.4 for this cost function [cf. Eq. 4.6]. Differentiating (A.1) and simplifying, we get the partial derivative of $J_s(TI_s, I_s,\, \mathbf{d_s})$ with respect to TI_s,

$$\begin{aligned} \partial J_s(TI_s,\, I_s,\, \mathbf{d_s})/\partial TI_s &= c_{pi} + (c_{ps} - c_{pi})\mathbf{1}(A_s) \\ &+ h \sum_{z_1=0}^{TI_s} \sum_{z_2=0}^{TI_s - z_1} f_{Z_L^2}(z_2|\,\mathbf{d_s})\, f_{Z_L^1}(z_1|\,\mathbf{d_s}) \\ &- \pi_1 \sum_{z_1=TI_s}^{\infty} f_{Z_L^1}(z_1|\,\mathbf{d_s}) - \pi_2 \sum_{z_2=0}^{\infty} z_2\, f_{Z_L^2}(z_2|\,\mathbf{d_s})\, f_{Z_L^1}(TI_s|\,\mathbf{d_s}) \\ &- \pi_2 \sum_{z_1=0}^{TI_s} \sum_{z_2=TI_s - z_1}^{\infty} f_{Z_L^2}(z_2|\,\mathbf{d_s})\, f_{Z_L^1}(z_1|\,\mathbf{d_s}) + \pi_2 \sum_{z_2=0}^{\infty} z_2\, f_{Z_L^2}(z_2|\,\mathbf{d_s})\, f_{Z_L^1}(TI_s|\,\mathbf{d_s}). \end{aligned}$$

And, simplifying we have

$$\begin{aligned}\partial J_s(TI_s,\, I_s,\, \mathbf{d_s})/\partial TI_s \;&=\; c_{pi} + (c_{ps} - c_{pi})\mathbf{1}(A_s) \;+\; h\, F_{Z_L}(TI_s|\,\mathbf{d_s}) \\ &- \pi_1 \left[1 - F_{Z_L^1}(TI_s|\,\mathbf{d_s})\right] \;-\; \pi_2 \left[F_{Z_L^1}(TI_s|\,\mathbf{d_s}) - F_{Z_L}(TI_s|\,\mathbf{d_s})\right]. \quad \text{(A.2)}\end{aligned}$$

The convexity of $J_s(TI_s,\, I_s,\, \mathbf{d_s})$ is easily shown by recognizing that its partial derivative of second order taken with respect to TI_s is always nonnegative, given by

$$\begin{aligned}\partial^2 J_s(TI_s,\, I_s,\, \mathbf{d_s})/\partial TI_s^2 \;&= \\ h\, f_{Z_L}(TI_s|\,\mathbf{d_s}) \;&+\; [\pi_1 - \pi_2]\, f_{Z_L^1}(TI_s|\,\mathbf{d_s}) \;+\; \pi_2 f_{Z_L}(TI_s|\,\mathbf{d_s}), \qquad \text{(A.3)}\end{aligned}$$

where $f_{Z_L^1}(TI_s|\,\mathbf{d_s})$ and $f_{Z_L}(TI_s|\,\mathbf{d_s})$ are the corresponding probability functions. The right-hand side of (A.3) is clearly non-negative since we have $\pi_1 \geqslant \pi_2$.

□

A.2 Proof of Theorem 4.1

The proof is by induction on s.

Step 1: Verify the base case:

The base case is for period $s = N$, which is the last decision period for replenishment. We have

$$\begin{aligned}G_N(TI_N,\, I_N,\, \mathbf{d_N}) \;&=\; J_N(TI_N,\, I_N,\, \mathbf{d_N}) + E_{\boldsymbol{\mathcal{E}_N}} \left[V_{N+1}(TI_N - \sum_b D^b_{N,N},\, \mathbf{D_{N+1}})\right] \\ &=\; J_N(TI_N,\, I_N,\, \mathbf{d_N}) - c_{ps} E_{\boldsymbol{\mathcal{E}_N}} \left[TI_N - Z_{[N,N+L)}\right].\end{aligned}$$

Since $J_N(TI_N,\, I_N,\, \mathbf{d_N})$ and $c_{ps} E_{\boldsymbol{\mathcal{E}_N}} \left[TI_N - Z_{[N,N+L)}\right]$ are individually convex in TI_N, the statement (i) is true. The presence of a finite in-house capacity level K allows us to write

$$G_N(TI_N,\, I_N,\, \mathbf{d_N}) \;=\; \begin{cases} G_N^{sub}(TI_N,\, I_N,\, \mathbf{d_N}) & \text{if } I_N + K \leqslant TI_N \\ G_N^{inh}(TI_N,\, I_N,\, \mathbf{d_N}) & \text{if } TI_N \leqslant I_N + K \end{cases} \qquad \text{(A.4)}$$

where

$$\begin{aligned}G_N^{inh}(TI_N,\, I_N,\, \mathbf{d_N}) \;&=\; c_{pi}(TI_N - I_N) + L_N(TI_N,\, \mathbf{d_N}) - c_{ps} E_{\boldsymbol{\mathcal{E}_N}} \left[TI_N - Z_{[N,N+L)}\right], \\ G_N^{sub}(TI_N,\, I_N,\, \mathbf{d_N}) \;&=\; G_N^{inh}(TI_N,\, I_N,\, \mathbf{d_N}) + (c_{ps} - c_{pi})(TI_N - I_N - K),\end{aligned}$$

which are convex in TI_N. Note that

$$\begin{aligned}\partial G_N^{inh}(TI_N,\, I_N,\, \mathbf{d_N})/\partial TI_N &= \partial L_N(TI_N,\, \mathbf{d_N})/\partial TI_N - (c_{ps} - c_{pi}),\\ \partial G_N^{sub}(TI_N,\, I_N,\, \mathbf{d_N})/\partial TI_N &= \partial L_N(TI_N,\, \mathbf{d_N})/\partial TI_N.\end{aligned}$$

Hence the solution of the corresponding first-order conditions (TI_N^{inh} and TI_N^{sub}, respectively) has the property that $TI_N^{sub} \leqslant TI_N^{inh}$ since $c_{ps} \geqslant c_{pi}$. Thus, we now need to consider the following four regions on the value of $I_N + K$, instead of the two given in (A.4).

Regions of $(I_N + K)$	Minimizer
$I_N + K \leqslant TI_N^{sub}$	TI_N^{sub}
$TI_N^{sub} \leqslant I_N + K \leqslant TI_N^{inh}$	$I_N + K$
$TI_N^{inh} \leqslant I_N + K \leqslant TI_N^{inh} + K$	TI_N^{inh}
$TI_N^{inh} + K \leqslant I_N + K$	I_N

The function $G_N(TI_N,\, I_N,\, \mathbf{d_N})$ for a given $\mathbf{d_N}$ value takes different forms in these regions, with different minimizers. Thus, the statement (iii) is true. For the statement (ii), since $V_N(I_N,\, \mathbf{d_N})$ is determined by minimizing $G_N(TI_N,\, I_N,\, \mathbf{d_N})$ over $I_N \leqslant TI_N$, we have

$$V_N(I_N,\, \mathbf{d_N}) = \begin{cases} G_N(TI_N^{sub},\, I_N,\, \mathbf{d_N}) & I_N \leqslant TI_N^{sub} - K \\ G_N(I_N + K,\, I_N,\, \mathbf{d_N}) & TI_N^{sub} - K \leqslant I_N \leqslant TI_N^{inh} - K \\ G_N(TI_N^{inh},\, I_N,\, \mathbf{d_N}) & TI_N^{inh} - K \leq I_N \leqslant TI_N^{inh} \\ G_N(I_N,\, I_N,\, \mathbf{d_N}) & TI_N^{inh} \leqslant I_N. \end{cases}$$

Hence given the convexity of $G_N(\cdot\,,\, I_N,\, \mathbf{d_N})$ which we have just shown above for the statement (i), $V_N(I_N,\, \mathbf{d_N})$ is convex in I_N. This proves the statement (ii).

Step 2: Formulate the inductive hypothesis:

As the induction hypothesis, suppose the statements (i)-(iii) are true for a particular period $s + 1$, where $2 \leqslant s + 1 < N$.

Step 3: Prove the inductive step:

Given the above hypothesis, we shall prove that the statements (i)-(iii) are true for period s. We have

$$G_s(TI_s,\, I_s,\, \mathbf{d_s}) = J_s(TI_s,\, I_s,\, \mathbf{d_s}) + E_{\boldsymbol{\mathcal{E}_s}}\left[V_{s+1}(TI_s - \sum_b D_{s,s}^b,\, \mathbf{D_{s+1}})\right],$$

which is convex in TI_s since $J_s(TI_s,\, I_s,\, \mathbf{d_s})$ and $E_{\boldsymbol{\mathcal{E}_s}}[V_{s+1}(TI_s - \sum_b D^b_{s,s},\, \mathbf{D_{s+1}})]$ are convex in TI_s due to the induction hypothesis for the statement (ii) and the fact that the expectation of convex function $V_{s+1}(x,\, \mathbf{D_{s+1}})$, $E_{\boldsymbol{\mathcal{E}_s}}[V_{s+1}(x,\, \mathbf{D_{s+1}})]$, is convex in x. This proves the statement (i). For the statement (ii), we have

$$V_s(I_s,\, \mathbf{d_s}) \;=\; \begin{cases} G_s(TI_s^{sub},\, I_s,\, \mathbf{d_s}) & I_s \leqslant TI_s^{sub} - K \\ G_s(I_s + K,\, I_s,\, \mathbf{d_s}) & TI_s^{sub} - K \leqslant I_s \leqslant TI_s^{inh} - K \\ G_s(TI_s^{inh},\, I_s,\, \mathbf{d_s}) & TI_s^{inh} - K \leqslant I_s \leqslant TI_s^{inh} \\ G_s(I_s,\, I_s,\, \mathbf{d_s}) & TI_s^{inh} \leqslant I_s, \end{cases}$$

since $V_s(I_s,\, \mathbf{d_s})$ is determined by minimizing $G_s(TI_s,\, I_s,\, \mathbf{d_s})$ over $I_s \leqslant TI_s$. Due to the convexity of $G_s(\cdot\,,\, I_s,\, \mathbf{d_s})$, $V_s(I_s,\, \mathbf{d_s})$ is convex in I_s as well. This proves the statement (ii). The statement (iii) follows from statement (i). The function $G_s(\cdot\,,\, I_s,\, \mathbf{d_s})$ takes different forms in regions on the value of I_s with the following minimizers (as was shown for $G_N(\cdot\,,\, I_N,\, \mathbf{d_N})$) in each of these regions

$$TI_s^\star(I_s,\, \mathbf{d_s}) = \begin{cases} TI_s^{sub} & I_s \leqslant TI_s^{sub} - K \\ I_s + K & TI_s^{sub} - K \leqslant I_s \leqslant TI_s^{inh} - K \\ TI_s^{inh} & TI_s^{inh} - K \leqslant I_s \leqslant TI_s^{inh} \\ I_s & TI_s^{inh} \leqslant I_s. \end{cases}$$

Thus, the statement (iii) is true. This completes the proof.

□

A.3 Proof of Theorem 4.2

The proof is by induction on s.

Step 1: Verify the base case:

The base case is for two consecutive periods $s = N$ and $s = N + 1$, where period $s = N$ is the last decision period for replenishment. For the statement (i), we have, from the end-of-horizon condition and the optimal replenishment policy of period N,

$$
\begin{aligned}
V_{N+1}(x,\,\mathbf{d}) &= -\,c_{ps}\,E_{\boldsymbol{\mathcal{E}_N}}\left[x - Z_{[N+1,N+L)}\right]\\
V_N(x,\,\mathbf{d}) &= \begin{cases}
G_N^{sub}(TI_N^{sub},\,x,\,\mathbf{d}), & x \leqslant TI_N^{sub} - K\\
G_N^{inh}(x+K,\,x,\,\mathbf{d}), & TI_N^{sub} - K \leqslant\; x \leqslant TI_N^{inh} - K\\
G_N^{inh}(TI_N^{inh},\,x,\,\mathbf{d}), & TI_N^{inh} - K \leqslant\; x \leqslant TI_N^{inh}\\
G_N^{inh}(x,\,x,\,\mathbf{d}), & TI_N^{inh} \leqslant\; x,
\end{cases}
\end{aligned}
$$

since $V_N(x,\,\mathbf{d})$ is determined by minimizing $G_N(TI_N,\,x,\,\mathbf{d})$ over $x \leqslant TI_N$. Taking the partial derivative with respect to x,

$$
\begin{aligned}
\partial V_{N+1}(x,\,\mathbf{d})/\partial x &= -\,c_{ps}\\
\partial V_N(x,\,\mathbf{d})/\partial x &= \begin{cases}
-\,c_{ps} & x \leqslant TI_N^{sub} - K\\
\partial G_N^{inh}(x+K,\,x,\,\mathbf{d})/\partial x & TI_N^{sub} - K \leqslant\; x \leqslant TI_N^{inh} - K\\
-\,c_{pi} & TI_N^{inh} - K \leqslant\; x \leqslant TI_N^{inh}\\
\partial G_N^{inh}(x,\,x,\,\mathbf{d})/\partial x & TI_N^{inh} \leqslant\; x,
\end{cases}
\end{aligned}
$$

where

$$
\begin{aligned}
\partial G_N^{inh}(x+K,\,x,\,\mathbf{d})/\partial x \;=\; & h\,F_{Z_L}(x+K|\,\mathbf{d}) \;-\; \pi_1\left[1 - F_{Z_L^1}(x+K|\,\mathbf{d})\right]\\
& -\,\pi_2\left[F_{Z_L^1}(x+K|\,\mathbf{d}) - F_{Z_L}(x+K|\,\mathbf{d})\right] \;-\; c_{ps}.
\end{aligned}
\tag{A.5}
$$

We conclude that $\partial G_N^{inh}(x+K,\,x,\,\mathbf{d})/\partial x \;\geqslant\; -c_{ps}$. This can be shown as follows. At $x = TI_N^{sub} - K$ (lower bound of the second region), we have $\partial G_N^{inh}(TI_N^{sub},\,x,\,\mathbf{d})/\partial x = \partial G_N^{sub}(TI_N^{sub},\,x,\,\mathbf{d})/\partial x \,=\, -\,c_{ps}$. This requires the term

$$
\begin{aligned}
& h\,F_{Z_L}(x+K|\,\mathbf{d}) \;-\; \pi_1\left[1 - F_{Z_L^1}(x+K|\,\mathbf{d})\right]\\
& -\,\pi_2\left[F_{Z_L^1}(x+K|\,\mathbf{d}) - F_{Z_L}(x+K|\,\mathbf{d})\right]
\end{aligned}
\tag{A.6}
$$

in the right-hand side of (A.5) to vanish at $x = TI_N^{sub} - K$. As x approaches the upper bound of the second region, the cumulative probabilities $F_{Z_L^1}(\cdot\,|\,\mathbf{d})$ and $F_{Z_L}(\cdot\,|\,\mathbf{d})$ can not decrease. This will make the term (A.6) non-decreasing as well, and hence it is always true that $\partial G_N^{inh}(x+K,\,x,\,\mathbf{d})/\partial x \;\geqslant\; -\,c_{ps}$ over the second region. In a similar manner, we can show that $\partial G_N^{inh}(x,\,x,\,\mathbf{d})/\partial x \;\geqslant\; -\,c_{ps}$ over the fourth region. For the other two cases, we know that $-c_{pi} \;\geqslant\; -c_{ps}$. So, we can conclude that

$\partial V_N(x,\,\mathbf{d})/\partial x \geqslant \partial V_{N+1}(x,\,\mathbf{d})/\partial x$ as given in the statement (i).

For the statement (ii),

$$\begin{aligned}
\partial G_N(y,\,x,\,\mathbf{d})/\partial y &= \partial J_N(y,\,x,\,\mathbf{d})/\partial y - c_{ps} \\
\partial G_{N-1}(y,\,x,\,\mathbf{d})/\partial y &= \partial J_{N-1}(y,\,x,\,\mathbf{d})/\partial y + E_{\boldsymbol{\mathcal{E}_{N-1}}}[\partial V_N(y - \sum_b D^b_{N-1,N-1}, \mathbf{D_N})/\partial y].
\end{aligned}$$

It was already shown in the above that $\partial V_N(x,\,\mathbf{d})/\partial x \geqslant -c_{ps}$ for all x and $\mathbf{d}$. Therefore, we have $E_{\boldsymbol{\mathcal{E}_{N-1}}}[\partial V_N(y-\sum_b D^b_{N-1,N-1}, \mathbf{D_N})/\partial y] \geqslant -c_{ps}$. Also, $\partial J_{N-1}(y,\,x,\,\mathbf{d})/\partial y$ is equivalent to $\partial J_N(y,\,x,\,\mathbf{d})/\partial y$ for a given value of $\mathbf{d}$ since we have that

$$\begin{aligned}
\partial J_{N-1}(y,\,x,\,\mathbf{d})/\partial y &= \partial J_N(y,\,x,\,\mathbf{d})/\partial y \\
&= c_{pi} + (c_{ps} - c_{pi})\mathbf{1}(A_{N-1}) + hF_{Z_L}(y|\,\mathbf{d}) \\
&\quad - \pi_1 \left[1 - F_{Z_L^1}(y|\,\mathbf{d})\right] - \pi_2 \left[F_{Z_L^1}(y|\,\mathbf{d}) - F_{Z_L}(y|\,\mathbf{d})\right].
\end{aligned}$$

Hence $\partial G_{N-1}(y,\,x,\,\mathbf{d})/\partial y \geqslant \partial G_N(y,\,x,\,\mathbf{d})/\partial y$, which proves the statement (ii). Due to the convexity of $G_s(\,.\,,\,.\,,\,.)$ and statement (ii), statement (iii) is trivial.

Step 2: Formulate the inductive hypothesis:

Suppose statements (i)-(iii) are true for a particular combination of two consecutive periods s and $s+1$, where $2 \leqslant s \leqslant N$.

Step 3: Prove the inductive step:

Given the above hypothesis, we shall prove that the statements (i)-(iii) are true for two consecutive periods $s-1$ and s. Let us examine the cases on the value of x, the inventory position before ordering. Let $\mathcal{I}^i_s$ for $i = 1,\,2,\,3,\,4$ be the event that inventory position before ordering in period s, x, falls in the ith region. We number the regions in an increasing order, where the first case are given the number $i = 1$.

Label	Regions on x	Minimizer
$\mathcal{I}^1_s$	$x \leqslant TI^{sub}_s - K$	TI^{sub}_s
$\mathcal{I}^2_s$	$TI^{sub}_s - K \leqslant x \leqslant TI^{inh}_s - K$	$x + K$
$\mathcal{I}^3_s$	$TI^{inh}_s - K \leqslant x \leqslant TI^{inh}_s$	TI^{inh}_s
$\mathcal{I}^4_s$	$TI^{inh}_s \leqslant x$	x

For the statement (i), we need to consider the following seven cases on the value of x, being possible for the two consecutive periods $s-1$ and s. The cases are determined according to statement (iii) of the induction hypothesis and the condition that $TI_s^{inh}(\mathbf{d}) - TI_s^{sub}(\mathbf{d}) \geqslant K$. For the sake of brevity in the following, define δ_s as the difference $\delta_s = \partial V_{s-1}(x,\,\mathbf{d})/\partial x - \partial V_s(x,\,\mathbf{d})/\partial x$.

- **Case 1:** $\mathbf{1}(\mathcal{I}_{s-1}^1\mathcal{I}_s^1) = 1$. We have

$$\begin{aligned}\partial V_{s-1}(x,\,\mathbf{d})/\partial x &= \partial G_{s-1}^{sub}(TI_{s-1}^{sub},\,x,\,\mathbf{d})/\partial x = -\,c_{ps}\\ \partial V_s(x,\,\mathbf{d})/\partial x &= \partial G_s^{sub}(TI_s^{sub},\,x,\,\mathbf{d})/\partial x = -\,c_{ps}.\end{aligned}$$

  Thus, $\delta_s = 0$ and we conclude that the statement (i) is true for this case.

- **Case 2:** $\mathbf{1}(\mathcal{I}_{s-1}^2\mathcal{I}_s^1) = 1$. We have

$$\begin{aligned}\partial V_{s-1}(x,\,\mathbf{d})/\partial x &= \partial G_{s-1}^{inh}(x+K,\,x,\,\mathbf{d})/\partial x\\ &= h\,F_{Z_L}(x+K|\,\mathbf{d}) - \pi_1\left[1 - F_{Z_L^1}(x+K|\,\mathbf{d})\right]\\ &\quad - \pi_2\left[F_{Z_L^1}(x+K|\,\mathbf{d}) - F_{Z_L}(x+K|\,\mathbf{d})\right]\\ &\quad + E_{\boldsymbol{\mathcal{E}}_{s-1}}\left[\partial V_s(x+K-\sum_b D_{s-1,s-1}^b,\,\mathbf{D_s})/\partial x\right],\\ \partial V_s(x,\,\mathbf{d})/\partial x &= \partial G_s^{sub}(TI_s^{sub},\,x,\,\mathbf{d})/\partial x = -\,c_{ps}.\end{aligned}$$

  δ_s is nonnegative due to the similar reasoning made for (A.6). Thus, we conclude that the statement (i) is true for this case.

- **Case 3:** $\mathbf{1}(I_{s-1}^2\mathcal{I}_s^2) = 1$. We have

$$\begin{aligned}\partial V_{s-1}(x,\,\mathbf{d})/\partial x &= \partial G_{s-1}^{inh}(x+K,\,x,\,\mathbf{d})/\partial x\\ &= h\,F_{Z_L}(x+K|\,\mathbf{d}) - \pi_1\left[1 - F_{Z_L^1}(x+K|\,\mathbf{d})\right]\\ &\quad - \pi_2\left[F_{Z_L^1}(x+K|\,\mathbf{d}) - F_{Z_L}(x+K|\,\mathbf{d})\right]\\ &\quad + E_{\boldsymbol{\mathcal{E}}_{s-1}}\left[\partial V_s(x+K-\sum_b D_{s-1,s-1}^b,\,\mathbf{D_s})/\partial x\right],\\ \partial V_s(x,\,\mathbf{d})/\partial x &= \partial G_s^{inh}(x+K,\,x,\,\mathbf{d})/\partial x\\ &= h\,F_{Z_L}(x+K|\,\mathbf{d}) - \pi_1\left[1 - F_{Z_L^1}(x+K|\,\mathbf{d})\right]\\ &\quad - \pi_2\left[F_{Z_L^1}(x+K|\,\mathbf{d}) - F_{Z_L}(x+K|\,\mathbf{d})\right]\\ &\quad + E_{\boldsymbol{\mathcal{E}}_s}\left[\partial V_{s+1}(x+K-\sum_b D_{s,s}^b,\,\mathbf{D_{s+1}})/\partial x\right].\end{aligned}$$

δ_s is nonnegative since $\partial V_s(x + K - \sum_b D^b_{s-1,s-1},\, \mathbf{D_s})/\partial x \;\geqslant\; \partial V_{s+1}(x + K - \sum_b D^b_{s,s},\, \mathbf{D_{s+1}})/\partial x$ due to statement (i) of the induction hypothesis. To conclude, statement (i) is true.

- **Case 4:** $\mathbf{1}(\mathcal{I}^3_{s-1}\mathcal{I}^1_s) = 1$. We have

$$\begin{aligned}
\partial V_{s-1}(x,\, \mathbf{d})/\partial x &= \partial G^{inh}_{s-1}(TI^{inh}_{s-1},\, x,\, \mathbf{d})/\partial x \;=\; -\,c_{pi}, \\
\partial V_s(x,\, \mathbf{d})/\partial x &= \partial G^{sub}_s(TI^{sub}_s,\, x,\, \mathbf{d})/\partial x \;=\; -\,c_{ps}.
\end{aligned}$$

$\delta_s = c_{ps} \,-\, c_{pi} \geqslant 0$ by definition. Thus, the statement (i) is true.

- **Case 5:** $\mathbf{1}(\mathcal{I}^3_{s-1}\mathcal{I}^2_s) = 1$. We have

$$\begin{aligned}
\partial V_{s-1}(x,\, \mathbf{d})/\partial x &= \partial G^{inh}_{s-1}(TI^{inh}_{s-1},\, x,\, \mathbf{d})/\partial x \;=\; -\,c_{pi}, \\
\partial V_s(x,\, \mathbf{d})/\partial x &= \partial G^{inh}_s(x + K,\, x,\, \mathbf{d})/\partial x \\
&= h\, F_{Z_L}(x + K|\, \mathbf{d}) \;-\; \pi_1 \left[1 - F_{Z^1_L}(x + K|\, \mathbf{d})\right] \\
&\quad - \pi_2 \left[F_{Z^1_L}(x + K|\, \mathbf{d}) - F_{Z_L}(x + K|\, \mathbf{d})\right] \\
&\quad + E_{\boldsymbol{\mathcal{E}_s}} \left[\partial V_{s+1}(x + K - \sum_b D^b_{s,s},\, \mathbf{D_{s+1}})/\partial x\right].
\end{aligned}$$

δ_s is nonnegative since $x + K \leqslant TI^{inh}_s$ for $\mathcal{I}^2_s$ and $V_s(\cdot\,,\cdot)$ is convex and these make $\partial V_s(\cdot\,,\cdot)/\partial x \leqslant -\,c_{pi}$.

- **Case 6:** $\mathbf{1}(\mathcal{I}^3_{s-1}\mathcal{I}^3_s) = 1$. We have

$$\begin{aligned}
\partial V_{s-1}(x,\, \mathbf{d})/\partial x &= \partial G^{inh}_{s-1}(TI^{inh}_{s-1},\, x,\, \mathbf{d})/\partial x \;=\; -\,c_{pi}, \\
\partial V_s(x,\, \mathbf{d})/\partial x &= \partial G^{inh}_s(TI^{inh}_s,\, x,\, \mathbf{d})/\partial x \;=\; -\,c_{pi}.
\end{aligned}$$

Thus, $\delta_s = 0$ and we conclude that the statement (i) is true for this case.

- **Case 7:** $\mathbf{1}(\mathcal{I}^4_{s-1}\mathcal{I}^4_s) = 1$. We have

$$\begin{aligned}
\partial V_{s-1}(x,\, \mathbf{d})/\partial x &= \partial G^{inh}_{s-1}(x,\, x,\, \mathbf{d})/\partial x \\
&= h\, F_{Z_L}(x|\, \mathbf{d}) \;-\; \pi_1 \left[1 - F_{Z^1_L}(x|\, \mathbf{d})\right] \\
&\quad - \pi_2 \left[F_{Z^1_L}(x|\, \mathbf{d}) - F_{Z_L}(x|\, \mathbf{d})\right] \\
&\quad + E_{\boldsymbol{\mathcal{E}_{s-1}}} \left[\partial V_s(x - \sum_b D^b_{s-1,s-1},\, \mathbf{D_s})/\partial x\right],
\end{aligned}$$

$$\begin{aligned}
\partial V_s(x,\,\mathbf{d})/\partial x \;&=\; \partial G_s^{inh}(x,\,x,\,\mathbf{d})/\partial x \\
&=\; h\,F_{Z_L}(x|\,\mathbf{d}) \;-\; \pi_1\left[1 - F_{Z_L^1}(x|\,\mathbf{d})\right] \\
&\quad -\pi_2\left[F_{Z_L^1}(x|\,\mathbf{d}) - F_{Z_L}(x|\,\mathbf{d})\right] \\
&\quad + E_{\boldsymbol{\mathcal{E}_s}}\left[\partial V_{s+1}(x - \sum_b D_{s,s}^b,\,\mathbf{D_{s+1}})/\partial x\right].
\end{aligned}$$

δ_s is nonnegative since $\partial V_s(x - \sum_b D^b_{s-1,s-1},\,\mathbf{D_s})/\partial x \geqslant \partial V_{s+1}(x - \sum_b D^b_{s,s},\,\mathbf{D_{s+1}})/\partial x$ due to the statement (i) of the induction hypothesis. To conclude, the statement (i) is true.

To conclude, the statement (i) is true for all seven cases. For the statement (ii), on the other hand, we have

$$\begin{aligned}
\partial G_{s-1}(y,\,x,\,\mathbf{d})/\partial y \;&=\; \partial J_{s-1}(y,\,x,\,\mathbf{d})/\partial y \,+\, E_{\boldsymbol{\mathcal{E}_{s-1}}}\left[\partial V_s(y - \sum_b D^b_{s-1,s-1},\,\mathbf{D_s})/\partial y\right], \\
\partial G_s(y,\,x,\,\mathbf{d})/\partial y \;&=\; \partial J_s(y,\,x,\,\mathbf{d})/\partial y \,+\, E_{\boldsymbol{\mathcal{E}_s}}\left[\partial V_{s+1}(y - \sum_b D^b_{s,s},\,\mathbf{D_{s+1}})/\partial y\right].
\end{aligned}$$

where $\partial J_{s-1}(y,\,x,\,\mathbf{d})/\partial y = \partial J_s(y,\,x,\,\mathbf{d})/\partial y$ for a given value of y and $\mathbf{d}$ due to constant parameters, and $\partial V_s(y - \sum_b D^b_{s-1,s-1},\,\mathbf{D_s})/\partial y \geqslant \partial V_{s+1}(y - \sum_b D^b_{s,s},\,\mathbf{D_{s+1}})/\partial y$ due to statement (i) of the induction hypothesis. Consequently we have $\partial G_{s-1}(y,\,x,\,\mathbf{d})/\partial y \geqslant \partial G_s(y,\,x,\,\mathbf{d})/\partial y$, which proves that statement (ii) is true. Statement (iii) holds due to the statement (ii). This completes the proof.

□

A.4 Proof of Theorem 4.3

The proof is by induction on s.

Step 1: Verify the base case:

The base case is for period $s = N$, which is the last decision period for replenishment. We have two distinct values $\bar{\mathbf{d}}$ and $\mathbf{d}$ of the order commitment vector available at the beginning of period s, with $\sum_b \sum_{k=1}^{L} \bar{d}^b_{s-1,s-1+k} \leqslant \sum_b \sum_{k=1}^{L} d^b_{s-1,s-1+k}$. For statement ($i$), by applying the optimal replenishment policy of period N and then taking the partial derivative of first order with respect to x, we get

$$\partial V_N(x,\,\mathbf{d})/\partial x \;=\; \begin{cases} -\,c_{ps} & x \leqslant TI_N^{sub} - K \\ \partial G_N^{inh}(x+K,\,x,\,\mathbf{d})/\partial x & TI_N^{sub} - K \leqslant\; x \leqslant TI_N^{inh} - K \\ -\,c_{pi} & TI_N^{inh} - K \leqslant\; x \leqslant TI_N^{inh} \\ \partial G_N^{inh}(x,\,x,\,\mathbf{d})/\partial x & TI_N^{inh} \leqslant\; x, \end{cases}$$

where

$$\begin{aligned} \partial G_N^{inh}(x+K,\,x,\,\mathbf{d})/\partial x \;&=\; h\,F_{Z_L}(x+K|\,\mathbf{d}) \;-\; \pi_1\left[1 - F_{Z_L^1}(x+K|\,\mathbf{d})\right] \\ &\quad -\,\pi_2\left[F_{Z_L^1}(x+K|\,\mathbf{d}) - F_{Z_L}(x+K|\,\mathbf{d})\right] \;-\; c_{ps}, \\ \partial G_N^{inh}(x,\,x,\,\mathbf{d})/\partial x \;&=\; h\,F_{Z_L}(x|\,\mathbf{d}) \;-\; \pi_1\left[1 - F_{Z_L^1}(x|\,\mathbf{d})\right] \\ &\quad -\,\pi_2\left[F_{Z_L^1}(x|\,\mathbf{d}) - F_{Z_L}(x|\,\mathbf{d})\right] \;-\; c_{ps}. \end{aligned}$$

$\partial V_N(x,\,\bar{\mathbf{d}})/\partial x$ for the value of $\bar{\mathbf{d}}$ can be readily written, and hence is omitted for the sake of brevity. Note that we have $F_{Z_L^1}(x+K|\,\mathbf{d}) \leqslant F_{Z_L^1}(x+K|\,\bar{\mathbf{d}})$ and $F_{Z_L}(x+K|\,\mathbf{d}) \leqslant F_{Z_L}(x+K|\,\bar{\mathbf{d}})$ since $\mathbf{d} \geqslant_{st} \bar{\mathbf{d}}$ (where $\geqslant_{st}$ means that *stochastically larger than*). This argument is also valid for $F_{Z_L^1}(x|\,\cdot\,)$ and $F_{Z_L}(x|\,\cdot\,)$. Consequently, it is true that $\partial V_N(x,\,\bar{\mathbf{d}})/\partial x \geqslant \partial V_N(x,\,\mathbf{d})/\partial x$, which proves the statement (i).

For the statement (ii), we have

$$\partial G_N(y,\,x,\,\mathbf{d})/\partial y \;=\; \partial J_N(y,\,x,\,\mathbf{d})/\partial y \;-\; c_{ps}.$$

$\partial G_N(y,\,x,\,\bar{\mathbf{d}})/\partial y \geqslant \partial G_N(y,\,x,\,\mathbf{d})/\partial y$ since $\partial J_N(y,\,x,\,\bar{\mathbf{d}})/\partial y \geqslant \partial J_N(y,\,x,\,\mathbf{d})/\partial y$ due to the stochastic order relations discussed above. This proves the statement (ii) for period N. The statement (iii) follows from statement (ii).

Step 2: Formulate the inductive hypothesis:

As the induction hypothesis, suppose the statements (i)-(iii) are true for a particular period $s+1$, where $1 \leqslant s+1 < N$.

Step 3: Prove the inductive step:

Given the above hypothesis, we shall prove that the statements (i)-(iii) are true for period s. For the statement (i), we have, by applying the optimal replenishment policy of period s for the value of $\mathbf{d}$ and then taking the partial derivative of first order with respect to x,

$$\partial V_s(x,\,\mathbf{d})/\partial x \;=\; \begin{cases} -\,c_{ps} & x \leqslant TI_s^{sub} - K \\ \partial G_s^{inh}(x+K,\,x,\,\mathbf{d})/\partial x & TI_s^{sub} - K \leqslant\; x \leqslant TI_s^{inh} - K \\ -\,c_{pi} & TI_s^{inh} - K \leqslant\; x \leqslant TI_s^{inh} \\ \partial G_s^{inh}(x,\,x,\,\mathbf{d})/\partial x & TI_s^{inh} \leqslant\; x, \end{cases}$$

$\partial V_s(x,\,\bar{\mathbf{d}})/\partial x$ for the value of $\bar{\mathbf{d}}$ can be readily written, and hence is omitted for the sake of brevity. We have

$$\partial G_s^{inh}(x,\,x,\,\bar{\mathbf{d}})/\partial x \;=\; \partial J_s(x,\,x,\,\bar{\mathbf{d}})/\partial x \,+\, E_{\boldsymbol{\mathcal{E}}_s}\left[\partial V_{s+1}(x - \sum_b \bar{D}^b_{s,s},\,\bar{\mathbf{D}}_{\mathbf{s+1}})/\partial x\right],$$

$$\partial G_s^{inh}(x,\,x,\,\mathbf{d})/\partial x \;=\; \partial J_s(x,\,x,\,\mathbf{d})/\partial x \,+\, E_{\boldsymbol{\mathcal{E}}_s}\left[\partial V_{s+1}(x - \sum_b D^b_{s,s},\,\mathbf{D}_{\mathbf{s+1}})/\partial x\right],$$

where $\partial J_s(x,\,x,\,\bar{\mathbf{d}})/\partial x \;\geqslant\; \partial J_s(x,\,x,\,\mathbf{d})/\partial x$ as discussed above. Since we know that $D_{s+1} \;\geqslant_{st}\; \bar{D}_{s+1}$ and that $\partial V_{s+1}(x,\,\mathbf{d})/\partial x$ is non-decreasing in x due to the convexity of $V_{s+1}(x,\,\mathbf{d})$ in x, we have $\partial V_{s+1}(x - \sum_b \bar{D}^b_{s,s},\,\bar{\mathbf{D}}_{\mathbf{s+1}})/\partial x \;\geqslant\; \partial V_{s+1}(x - \sum_b D^b_{s,s},\,\mathbf{D}_{\mathbf{s+1}})/\partial x$, which follows from the statement (i) of the induction hypothesis. Consequently, this proves the statement (i).

From the discussion above, it is obvious that $\partial G_s(y,\,x,\,\bar{\mathbf{d}})/\partial y \;\geqslant\; \partial G_s(y,\,x,\,\mathbf{d})/\partial y$, which proves the statement (ii). The statement (iii) follows from the statement (ii). This completes the proof.

□

A.5 Proof of Proposition 5.1

Due to the forecast evolution structure we know that

$$\begin{aligned} Z^b_{[s,s+L)} &= \sum_{k=1}^{L} D^b_{s+k-1,s+k-1} \\ &= \sum_{k=1}^{L} d^b_{s-1,s+k-1}\; e^{(\varepsilon^b_{s+k-1,1}+\varepsilon^b_{s+k-2,2}+\cdots+\varepsilon^b_{s,k})}. \end{aligned} \qquad \text{(A.7)}$$

Figure A.1 shows the accumulating nature of $e^{\varepsilon^b_{s+k-m,m}}$'s.

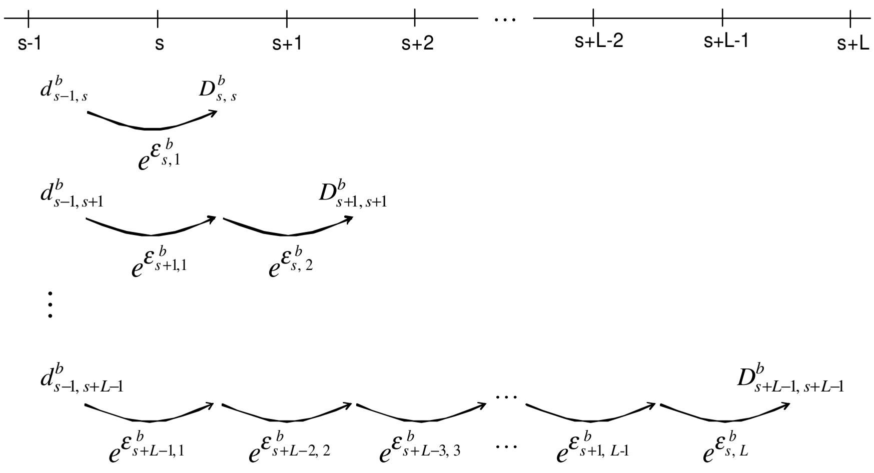


Figure A.1: Accumulation of $e^{\varepsilon^b_{s+k-m,m}}$'s

Assuming $E^b_k = d^b_{s-1,s+k-1}\ e^{(\varepsilon^b_{s+k-1,1}+\varepsilon^b_{s+k-2,2}+\cdots+\varepsilon^b_{s,k})}$ for convenience in what follows, taking logarithms on both sides of (A.7) and arranging we get

$$\ln\ Z^b_{[s,s+L)} = \ln \sum_{k=1}^{L} E^b_k = \ln \sum_{k=1}^{L} e^{\ln E^b_k} = g(\ln \mathbf{E^b}),$$

where $\mathbf{E^b}$ is the vector $[\,E^b_1,\, E^b_2,\, \ldots,\, E^b_L\,]$ and $g(\ln \mathbf{E^b})$ is a function of $\ln \mathbf{E^b}$. As an approximation to the logarithm of the sum of a series $\{E^b_k,\ 1 \leqslant k \leqslant L\}$, consider the first-order Taylor series expansion for $g(\ln \mathbf{E^b})$ around $E^b_k = \beta^b_k,\ k = 1,\, 2,\, \ldots,\, L$. Using the forecast evolution mechanism, we may take β^b_k as

$$\beta^b_k = \mu_{D^b}\ e^{(\mu_{\varepsilon^b_1}+\mu_{\varepsilon^b_2}+\cdots+\mu_{\varepsilon^b_k})}.$$

μ_{D^b} and $\mu_{\varepsilon^b_m}$ are known constants. More specifically, μ_{D^b} is the mean order size of stationary series and $\mu_{\varepsilon^b_m}$ is the mean non-stationary update factor m periods ahead. Thus β^b_k is a known constant, which estimates $(s+k-1)$th period's order realization from buyer b and does not depend on a given $d^b_{s-1,s+k-1}$ value. The first-order Taylor series approximation turns out to be

$$\begin{aligned}
\ln Z^b_{[s,s+L)} \;=\; g(\ln \mathbf{E^b}) \;&\simeq\; g(\ln \boldsymbol{\beta^b}) + \sum_{k=1}^{L} \frac{\partial g(\ln \mathbf{E^b})}{\partial \ln E^b_k}\Bigg|_{\boldsymbol{E^b}=\boldsymbol{\beta^b}} \left[\ln E^b_k - \ln \beta^b_k\right] \\
&= \ln \sum_{k=1}^{L} \beta^b_k + \sum_{k=1}^{L} \frac{\beta^b_k}{\sum\limits_{k=1}^{L} \beta^b_k} \left[\ln E^b_k - \ln \beta^b_k\right] \\
&= \ln \sum_{k=1}^{L} \beta^b_k + \sum_{k=1}^{L} \theta^b_k \left[\ln E^b_k - \ln \beta^b_k\right], \qquad \text{(A.8)}
\end{aligned}$$

where $\boldsymbol{\beta^b}$ is the vector $[\beta^b_1, \beta^b_2, \ldots, \beta^b_L]$, and θ^b_k is a dimensionless, known constant given by

$$\theta^b_k = \beta^b_k / \sum_{m=1}^{L} \beta^b_m.$$

Note that θ^b_k denotes the share of β^b_k (i.e., the expected order after k updates) in the total expected realized order from buyer b on the L-period horizon. Continuing with (A.8) by substituting for E^b_k, we have

$$\begin{aligned}
\ln Z^b_{[s,s+L)} \;&=\; g(\ln \mathbf{E^b}) \\
\simeq \;& \ln \sum_{k=1}^{L} \beta^b_k + \sum_{k=1}^{L} \theta^b_k \left[\ln d^b_{s-1,s+k-1} + \sum_{m=1}^{k} \varepsilon^b_{s+k-m,m} - \ln \beta^b_k\right]. \qquad \text{(A.9)}
\end{aligned}$$

Expanding the only random term $\sum\limits_{k=1}^{L} \theta^b_k \sum\limits_{m=1}^{k} \varepsilon^b_{s+k-m,m}$ on the right-hand side of (A.9) and following the MMFE assumption that commitment update vectors $\boldsymbol{\mathcal{E}^b_s}$ form a stationary and independent sequence through time ($\mathrm{Cov}(\varepsilon^b_s, \varepsilon^b_k) = 0,\; s \neq k$), we have

$$\begin{aligned}
&\sum_{k=1}^{L} \theta^b_k \sum_{m=1}^{k} \varepsilon^b_{s+k-m,m} = \\
&\qquad \theta^b_1 \varepsilon^b_{s,1} + \theta^b_2 \left[\varepsilon^b_{s+1,1} + \varepsilon^b_{s,2}\right] + \ldots + \theta^b_L \left[\varepsilon^b_{s+L-1,1} + \varepsilon^b_{s+L-2,2} + \cdots + \varepsilon^b_{s,L}\right] \\
&= \left[\theta^b_1 + \theta^b_2 + \cdots + \theta^b_L\right] \varepsilon^b_{s,1} + \ldots + \left[\theta^b_{L-1} + \theta^b_L\right] \varepsilon^b_{s,L-1} + \theta^b_L \varepsilon^b_{s,L} \\
&= \sum_{k=1}^{L} \lambda^b_k \, \varepsilon^b_{s,k} \\
&\triangleq \vartheta^b_L.
\end{aligned}$$

ϑ^b_L corresponds to a weighted sum of logarithmic commitment updates to be received from buyer b over the time interval $[s, s+L)$. This corresponds to the only random element in the total realized order. We call $\lambda^b_k = \sum\limits_{m=k}^{L} \theta^b_m \in (0,1]$ the weight corresponding to the k-step commitment updates $\varepsilon^b_{s,k}$ from buyer b. It decreases from

$\lambda_1^b = 1$ as period k increases from 1 to L, implying larger weight towards the period of order realization since update coefficients of earlier periods occur in more of the commitment terms.

As a consequence, (A.9) may be written as a random term (ϑ_L^b) plus a series of constant terms

$$\ln Z_{[s,s+L)}^b \simeq \vartheta_L^b + \ln \sum_{k=1}^{L} \beta_k^b + \sum_{k=1}^{L} \theta_k^b \left[\ln d_{s-1,s+k-1}^b - \ln \beta_k^b\right] \qquad \text{(A.10)}$$

Applying the approximation (A.10) to the probability distribution $F_{Z_L^b}(y \,|\, \mathbf{d_s})$, we have

$$\begin{aligned}
F_{Z_L^b}(y \,|\, \mathbf{d_s}) &= Pr\{Z_{[s,s+L)}^b \leqslant y \,|\, \mathbf{d_s}\} \\
&= Pr\{\ln Z_{[s,s+L)}^b \leqslant \ln y \,|\, \mathbf{d_s}\} \\
&\simeq Pr\{\vartheta_L^b + \ln \sum_{k=1}^{L} \beta_k^b + \sum_{k=1}^{L} \theta_k^b [\ln d_{s-1,s+k-1}^b - \ln \beta_k^b] \leqslant \ln y \,|\, \mathbf{d_s}\} \\
&= Pr\{\vartheta_L^b \leqslant \ln \frac{y}{\sum_{k=1}^{L} \beta_k^b} + \sum_{k=1}^{L} \theta_k^b \ln \frac{\beta_k^b}{d_{s-1,s+k-1}^b}\} \\
&= F_{\vartheta_L^b}(\ln \frac{y}{\sum_{k=1}^{L} \beta_k^b} + \sum_{k=1}^{L} \theta_k^b \ln \frac{\beta_k^b}{d_{s-1,s+k-1}^b}),
\end{aligned}$$

which completes the proof. Similar proof can be made for $F_{Z_L}(\cdot \,|\, \mathbf{d_s})$.

□

A.6 Proof of Theorem 5.1

Consider the approximation (A.10) that we introduced in the proof of Proposition 5.1.

$$\ln Z_{[s,s+L)}^b \simeq \vartheta_L^b + \ln \sum_{k=1}^{L} \beta_k^b + \sum_{k=1}^{L} \theta_k^b \left[\ln d_{s-1,s+k-1}^b - \ln \beta_k^b\right]$$

Continuing with this approximation, then

$$\begin{aligned}
\vartheta_L^b &\simeq \ln \frac{Z_{[s,s+L)}^b}{\sum_{k=1}^{L} \beta_k^b} + \sum_{k=1}^{L} \theta_k^b \ln \frac{\beta_k^b}{d_{s-1,s+k-1}^b} \\
&\simeq \frac{Z_{[s,s+L)}^b}{\sum_{k=1}^{L} \beta_k^b} - 1 + \sum_{k=1}^{L} \theta_k^b \ln \frac{\beta_k^b}{d_{s-1,s+k-1}^b}, \qquad \text{(A.11)}
\end{aligned}$$

which follows from the fact that $\ln x \simeq x - 1$ for small values of x. This holds for the ratio $Z^b_{[s,s+L)} / \sum_{k=1}^{L} \beta^b_k$ in (A.11). It takes small values since $\sum_{k=1}^{L} \beta^b_k$ is an estimate to $Z^b_{[s,s+L)}$ (which makes the ratio around 1) and $Z^b_{[s,s+L)}$ is interval censored (which prevents the ratio from taking large values towards the left and right extremes).

Consider the current L-period cost of carrying inventory and backorders associated with period s.

$$\begin{aligned} L_s(TI_s,\, \mathbf{d_s}) \;=\; & \\ & h \sum_{z_1=0}^{TI_s} \sum_{z_2=0}^{TI_s - z_1} (TI_s - z_1 - z_2)\, f_{Z^2_L}(z_2 |\, \mathbf{d_s})\, f_{Z^1_L}(z_1 |\, \mathbf{d_s}) \\ & + \pi_1 \sum_{z_1=TI_s}^{\infty} (z_1 - TI_s)\, f_{Z^1_L}(z_1 |\, \mathbf{d_s}) \;+\; \pi_2 \sum_{z_1=TI_s}^{\infty} \sum_{z_2=0}^{\infty} z_2\, f_{Z^2_L}(z_2 |\, \mathbf{d_s})\, f_{Z^1_L}(z_1 |\, \mathbf{d_s}) \\ & + \pi_2 \sum_{z_1=0}^{TI_s} \sum_{z_2=TI_s - z_1}^{\infty} (z_2 - (TI_s - z_1))\, f_{Z^2_L}(z_2 |\, \mathbf{d_s})\, f_{Z^1_L}(z_1 |\, \mathbf{d_s}), \end{aligned} \qquad \text{(A.12)}$$

Using the expression (A.11) to make the substitution $Z^b_{[s,s+L)} = \vartheta^b_L \sum_{k=1}^{L} \beta^b_k + \sum_{k=1}^{L} \beta^b_k - (\sum_{k=1}^{L} \theta^b_k \ln \frac{\beta^b_k}{d^b_{s-1,s+k-1}}) \sum_{k=1}^{L} \beta^b_k$, we change the basic variable of the integrands in (A.12) from $(z_1,\, z_2)$ to $(v_1,\, v_2)$. This change of variables results in a transformed integral over a domain in the $(\vartheta^1_L,\, \vartheta^2_L)$-space.

$$\begin{aligned} \widehat{L}_s(TI_s,\, \mathbf{d_s}) \;=\; & \\ & h \sum_{v_1=-\infty}^{\widehat{TI}_s} \sum_{v_2=-\infty}^{\widehat{TI}_s - \zeta^1 v_1} (\widehat{TI}_s - \zeta^1 v_1 - \zeta^2 v_2)\, f_{\vartheta^2_L}(v_2)\, f_{\vartheta^1_L}(v_1) \\ & + \pi_1 \sum_{v_1=\widehat{TI}_s}^{\infty} (\zeta^1 v_1 - \widehat{TI}_s)\, f_{\vartheta^1_L}(v_1 |\, \mathbf{d_s}) \;+\; \pi_2 \sum_{v_1=\widehat{TI}_s}^{\infty} \sum_{v_2=-\infty}^{\infty} \zeta^2 v_2\, f_{\vartheta^2_L}(v_2)\, f_{\vartheta^1_L}(v_1) \\ & + \pi_2 \sum_{v_1=-\infty}^{\widehat{TI}_s} \sum_{v_2=\widehat{TI}_s - \zeta^1 v_1}^{\infty} (\zeta^2 v_2 - (\widehat{TI}_s - \zeta^1 v_1))\, f_{\vartheta^2_L}(v_2)\, f_{\vartheta^1_L}(v_1), \end{aligned}$$

where $\zeta^b = \sum_{k=1}^{L} \beta^b_k$, and

$$\widehat{TI}_s \;\triangleq\; TI_s - \sum_b (1 - \sum_{k=1}^{L} \theta^b_k \ln \frac{\beta^b_k}{d^b_{s-1,s+k-1}}) \sum_{k=1}^{L} \beta^b_k, \qquad \text{(A.13)}$$

The dimension of $\widehat{TI}_s$ has the same unit with TI_s. Since $\widehat{TI}_s$ is a linear function in TI_s and $L_s(TI_s,\, \mathbf{d_s})$ is convex in TI_s, the modified cost function $\widehat{L}_s(TI_s,\, \mathbf{d_s})$ is also convex in TI_s. Its minimum occurs at $\widehat{TI}^*_s$, given by

$$\widehat{TI}^*_s \;\triangleq\; TI^*_s - \sum_b (1 - \sum_{k=1}^{L} \theta^b_k \ln \frac{\beta^b_k}{d^b_{s-1,s+k-1}}) \sum_{k=1}^{L} \beta^b_k, \qquad \text{(A.14)}$$

for the minimizer TI_s^* of $L_s(TI_s,\, \mathbf{d_s})$ for a given system state $(I_s,\, \mathbf{d_s})$.

The manufacturer's replenishment amount can be written as

$$\begin{aligned} q_s &= TI_s - I_s \\ &= \widehat{TI}_s + \sum_b (1 - \sum_{k=1}^{L} \theta_k^b \ln \frac{\beta_k^b}{d^b_{s-1,s+k-1}}) \sum_{k=1}^{L} \beta_k^b - I_s \\ &= \widehat{TI}_s - \widehat{I}_s, \end{aligned}$$

where

$$\widehat{I}_s \triangleq I_s - \sum_b (1 - \sum_{k=1}^{L} \theta_k^b \ln \frac{\beta_k^b}{d^b_{s-1,s+k-1}}) \sum_{k=1}^{L} \beta_k^b, \tag{A.15}$$

which obviously accommodates $\mathfrak{D}_{s,H} = \sum_b \sum_{k=1}^{H} \theta_k^b \ln d^b_{s-1,s+k-1}$ as

$$\widehat{I}_s = I_s - \sum_b \left[(1 - \sum_{k=1}^{L} \theta_k^b \ln \beta_k^b + \mathfrak{D}_{s,H}) \sum_{k=1}^{L} \beta_k^b \right].$$

This completes the proof.

□

A.7 Proof of Corollary 5.1

Suppose a certain replenishment decision TI_s is made upon observing the system state $(I_s,\, \mathbf{d_s})$. Using the expression (A.11), we may state inventory position before replenishment decision in period $s+1$ as

$$\begin{aligned} I_{s+1} &= TI_s - \sum_b D^b_{s,s} \\ &\simeq TI_s - \sum_b \left(\beta_1^b \varepsilon^b_{s,1} - \beta_1^b \ln \frac{\beta_1^b}{d^b_{s-1,s}} + \beta_1^b \right), \end{aligned} \tag{A.16}$$

where β_1^b is the expected order after one update from buyer b in any period s (i.e., updated $s-1$ commitment for period s). Then, applying the definition (A.15) to I_{s+1}, and using the forecast evolution equation, we have

$$
\begin{aligned}
\widehat{I}_{s+1} \;=\;& \\
& I_{s+1} - \sum_b (1 - \sum_{k=2}^{L+1} \theta_k^b \ln \frac{\beta_k^b}{D_{s,s+k-1}^b}) \sum_{k=2}^{L+1} \beta_k^b \\
=\;& I_{s+1} - \sum_b (1 - \sum_{k=2}^{L} \theta_k^b \ln \frac{\beta_k^b}{d_{s-1,s+k-1}^b \; e^{\varepsilon_{s,k}^b}} - \theta_{L+1}^b \ln \frac{\beta_{L+1}^b}{\mu_{D^b} \; e^{\varepsilon_{s,L+1}^b}}) \sum_{k=2}^{L+1} \beta_k^b \\
=\;& I_{s+1} - \sum_b (1 - \sum_{k=2}^{L} \theta_k^b \ln \frac{\beta_k^b}{d_{s-1,s+k-1}^b} - \theta_{L+1}^b \ln \frac{\beta_{L+1}^b}{\mu_{D^b}} + \sum_{k=2}^{L+1} \theta_k^b \varepsilon_{s,k}^b) \sum_{k=2}^{L+1} \beta_k^b.
\end{aligned}
\tag{A.17}
$$

By substituting (A.16) into (A.17) and rearranging terms, it follows that

$$
\begin{aligned}
\widehat{I}_{s+1} \;=\;& \\
& TI_s - \sum_b \beta_1^b \varepsilon_{s,1}^b - \beta_1^b \ln \frac{\beta_1^b}{d_{s-1,s}^b} + \beta_1^b \\
& - \sum_b (1 - \sum_{k=2}^{L} \theta_k^b \ln \frac{\beta_k^b}{d_{s-1,s+k-1}^b} - \theta_{L+1}^b \ln \frac{\beta_{L+1}^b}{\mu_{D^b}} + \sum_{k=2}^{L+1} \theta_k^b \varepsilon_{s,k}^b) \sum_{k=2}^{L+1} \beta_k^b \\
=\;& TI_s - \sum_b (1 - \sum_{k=1}^{L} \theta_k^b \ln \frac{\beta_k^b}{d_{s-1,s+k-1}^b}) \sum_{k=1}^{L} \beta_k^b \\
& - \sum_b \beta_1^b \varepsilon_{s,1}^b - \sum_{k=2}^{L+1} \beta_k^b \varepsilon_{s,k}^b + \beta_1^b + \beta_{L+1}^b \ln \frac{\beta_{L+1}^b}{\mu_{D^b}} \\
=\;& \widehat{TI}_s - \sum_b (\sum_{k=1}^{L+1} \beta_k^b \varepsilon_{s,k}^b - \beta_1^b - \beta_{L+1}^b \ln \frac{\beta_{L+1}^b}{\mu_{D^b}})
\end{aligned}
$$

where $\varepsilon_{s,k}^b$ is the random variable for the k-period ahead commitment update made in period s. For $\mathfrak{D}_{s+1,H}$, we have

$$
\begin{aligned}
\mathfrak{D}_{s+1,H} \;=\;& \sum_b^B \sum_{k=2}^{L+1} \theta_k^b \, \ln D_{s,s+k-1}^b \\
=\;& \sum_b^B \left(\sum_{k=2}^{L} \theta_k^b \, \ln d_{s,s+k-1}^b \; + \; \sum_{k=2}^{L+1} \theta_k^b \, \varepsilon_{s,k}^b \; + \; \theta_{H+1}^b \ln \mu_{D^b} \right) \\
=\;& \mathfrak{D}_{s,H} \; + \sum_b (\sum_{k=2}^{H+1} \beta_k^b \varepsilon_{s,k}^b + \beta_{H+1}^b \ln \mu_{D^b} - \beta_1^b \ln d_{s-1,s}^b) \, / \sum_{k=2}^{H+1} \beta_k^b \, ,
\end{aligned}
$$

which completes the proof.

□

APPENDIX B

ON THE SOLUTION OF NONLINEAR MODELS GOF_k

Solving our highly nonlinear problems, which involve discontinuities and large degrees of freedom, required some modifications to the programs. First, we reformulated the nonlinear model, having the discontinuous derivatives due to the absolute functions in the constraints, as an approximately equivalent smooth nonlinear model. The reformulation is necessary since we got termination messages like *Convergence too slow* (saying that the solution process is very slow) or *No change in objective although the reduced gradient is greater than the tolerance* (saying that there is no progress at all). These messages say that the solver stops with a feasible solution but unsuccessful termination where the optimality criteria have not been satisfied. The problem can be caused by discontinuous derivatives of the constraints involving absolute function, leading to inaccurate approximations to the marginal improvements around the current point. The reformulation approach taken for the absolute functions is to introduce positive and negative deviations as extra variables. This reformulation enlarges the feasible space, where it is likely to have multiple local optima depending on the objective function form.

Second, we modify some of the default tolerances and options in CONOPT3 solver since we have experienced solution difficulties. The approach taken is to adjust the algorithmic parameters dynamically as information about the behavior of the model is collected by experimenting. The main algorithmic parameters that we modified involve the minimum feasibility tolerance (increased), the maximum number of stalled iterations (increased), the triangular feasibility tolerance (increased), the relative accuracy of -one-dimensional search (decreased), the optimality tolerance (increased), and the method for finding the maximal step while searching for a feasible solution (switched to bending procedure) [1].

[1] More information on algorithmic parameters can be found in GAMS Reference Manual for CONOPT solver.

APPENDIX C

THE RELATED BIVARIATE ARIMA(0, 1, 1) PROCESS AND ITS FORECAST MODEL

We assume that the order quantities submitted by contract buyers follow a bivariate ARIMA(0, 1, 1) process (two time series being modeled jointly corresponding to two buyers). Since the differencing operator has an order of $d = 1$, the process models the differences between consecutive observations on order quantity (rather than the observed values directly), and is given as follows.

$$\begin{aligned} D^b_{1,1} &= \mu_{D^b} + \eta^b_1 \qquad \text{for all buyers } b \in \{1, 2\}, \quad \text{and} \\ D^1_{s,s} &= D^1_{s-1,s-1} - (1-\theta_{11})\,\eta^1_{s-1} + \theta_{12}\,\eta^2_{s-1} + \eta^1_s \\ D^2_{s,s} &= D^2_{s-1,s-1} + \theta_{21}\,\eta^1_{s-1} - (1-\theta_{22})\,\eta^2_{s-1} + \eta^2_s \end{aligned}$$

for $s = 2, 3, \ldots, N + L$. It may be written compactly in matrix form as

$$\begin{aligned} \mathbf{D_{1,1}} &= \boldsymbol{\mu}_\mathbf{D} + \boldsymbol{\eta}_\mathbf{1} \\ \mathbf{D_{s,s}} &= \mathbf{D_{s-1,s-1}} - (\mathbf{I} - \boldsymbol{\Theta})\,\boldsymbol{\eta}_{\boldsymbol{s}-\mathbf{1}} + \boldsymbol{\eta}_{\boldsymbol{s}} \end{aligned} \tag{C.1}$$

for $s = 2, 3, \ldots, N + L$, where

(*i*) $\mathbf{D_{s,s}} = \left[\begin{smallmatrix} D^1_{s,s} \\ D^2_{s,s} \end{smallmatrix}\right]$ is the 2×1 vector of order realizations submitted by the buyers in period s,

(*ii*) $\boldsymbol{\mu}_\mathbf{D} = \left[\begin{smallmatrix} \mu_{D^1} \\ \mu_{D^2} \end{smallmatrix}\right]$ is the 2×1 known vector of expected order realizations,

(*iii*) $\boldsymbol{\Theta} = \left[\begin{smallmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{smallmatrix}\right]$ is the 2×2 unknown matrix of moving average parameters. We assume that $0 < \theta_{bb} \leqslant 1$, for all $b \in \{1, 2\}$. This assumption is essential for the non-stationarity of the demand process. $\mathbf{I}$ is the 2×2 identity matrix such that only the diagonal elements of the matrix $\mathbf{I} - \mathbf{\Theta}$ have leading terms which are unity,

(*iv*) $\boldsymbol{\eta_s} = \left[\begin{smallmatrix} \eta_s^1 \\ \eta_s^2 \end{smallmatrix}\right]$ is the 2×1 random vector of disturbances in period s. We assume that $\boldsymbol{\eta_s}$ is independent and identically distributed vector with Normal density such that $\mathrm{E}[\boldsymbol{\eta_s}] = \mathbf{0}$ and a covariance matrix given by $\Sigma_{\boldsymbol{\eta}}$. This allows the disturbances at a given point in time to be correlated across the buyers. The variance of the disturbances corresponding to the bth buyer and the covariance between the disturbances corresponding to the bth and rth buyers are denoted by $\sigma^2_{\eta^b} = \mathrm{Var}(\eta_s^b)$ and $\sigma_{\eta^b,\eta^r} = \mathrm{Cov}(\eta_s^b,\, \eta_s^r)$ for $b \neq r \in \{1,\, 2\}$, respectively.

See Jenkins and Alavi (1981) for some aspects of modeling multivariate time series. The unknown parameters in this model are the matrix $\mathbf{\Theta}$ of moving average parameters and the covariance matrix $\Sigma_{\boldsymbol{\eta}}$ of disturbances. We will develop estimators for them later on in §6.5.1. Note also that the parameters are such that $D^b_{s,s}$ can be considered to be *nonnegative* for all practical purposes.

Each component in the demand model (C.1) depends not only on lagged values of itself but also on lagged values of the component corresponding to the other buyer. It has flexibility for representing a variety of time series data conveying trends (by the differencing operator) and correlations (by varying the values of $\mathbf{\Theta}$). This will enable us to investigate the system performance relative to a wide variety of demand characterizations addressed in our forecast evolution modeling.

This ARIMA model is simpler in its representation of demand uncertainty than the forecast evolution model. In particular, it does not utilize knowledge of order commitments and their evolution from period to period. It is based only on the previous observations of the predicted quantities and their errors.

Muth (1960) shows that the minimum-mean-squared-error forecast for this ARIMA(0, 1, 1) demand process is the first-order exponential-weighted moving average (EWMA). Let F^b_{s+j} be the forecast for the order quantity to be submitted by buyer b in period $s{+}j$, made after observing the order of period s. Also, $\mathbf{F_{s+j}} = \left[\begin{smallmatrix} F^1_{s+j} \\ F^2_{s+j} \end{smallmatrix}\right]$ denotes the bivariate forecast vector. If s is the last time period in which the buyers' orders are observed, then the first-order EWMA forecast vector for a future time period $s + j$ is given by

$$\mathbf{F_1} = \mu_{\mathbf{D}}$$

$$\mathbf{F_{s+j}} = \mathbf{\Theta}\,\mathbf{D_{s,s}} + (\mathbf{I} - \mathbf{\Theta})\,\mathbf{F_s} \qquad \text{(C.2)}$$

for $s = 1, 2, \ldots, N + L - 1$ and $j = 1, 2, \ldots, N + L - s$. The exponential smoothing constants are given by the elements of the 2×2 matrix $\mathbf{\Theta}$ of moving average parameters in the demand model (C.1). Hence, the first-order EWMA forecast corresponding to a buyer is a weighted average of the current time series values (i.e., $D^1_{s,s}$ and $D^2_{s,s}$) and the forecasts at the previous time period (i.e., F^1_s and F^2_s).

The right-hand side of the forecast function (C.2) does not depend on j, as it corresponds to a first-order model and thus $\mathbf{F_{s+j}}$ is used to forecast all future values of $\mathbf{D_{t,t}}$, $t > s$. This points out one disadvantage of the EWMA forecasting (as compared to the forecast evolution in MUFE). Any anticipated changes in the demand during the replenishment lead-time are not taken into account in this first-order EWMA forecasting. This is caused by the ARIMA(0, 1, 1) model assumption.

We can also express these first-order EWMA forecasts in terms of the random disturbances, $\boldsymbol{\eta_s}$,

$$\begin{aligned} \mathbf{F_{s+j}} &= \mathbf{F_s} + \mathbf{\Theta}\,\boldsymbol{\eta_s} \\ &= \mathbf{\Theta}\,\boldsymbol{\eta_s} + \mathbf{\Theta}\,\boldsymbol{\eta_{s-1}} + \ldots + \mathbf{\Theta}\,\boldsymbol{\eta_1} + \mu_{\mathbf{D}}, \end{aligned}$$

for all $j > 0$, where it can be readily shown that the forecast error is

$$\mathbf{D_{s,s}} - \mathbf{F_s} = \boldsymbol{\eta_s},$$

for $s = 1, 2, \ldots, N + L$. This implies that the first-order EWMA forecasts are unbiased.

APPENDIX D

TABLES FOR SOLUTION DETAILS

Table D.1: Solution details for Figure 7.2

H	FL	$100 - \daleth^1$ (%)	CI_H^+ (%)
	0.02	88.0	39.1
	0.04	83.1	46.5
2	0.06	79.2	52.3
	0.08	76.3 *	59.6
	0.10	74.4 *	61.7
	0.02	90.1	39.1
	0.04	85.0	46.5
3	0.06	80.8 *	52.4
	0.08	77.7 *	59.4
	0.10	75.6 *	61.5
	0.02	90.7 *	38.9
	0.04	85.3	46.5
4	0.06	81.1 *	52.4
	0.08	77.8 *	59.1
	0.10	75.7	61.3
	0.02	91.4 *	38.7
	0.04	85.9 *	46.3
5	0.06	81.4 *	52.3
	0.08	78.0 *	58.7
	0.10	75.7	60.9
	0.02	90.6	38.3
	0.04	86.5 *	46.1
6	0.06	81.8 *	52.1
	0.08	78.2 *	58.2
	0.10	75.8	60.4
	0.02	90.5	37.8
	0.04	87.4 *	45.6
7	0.06	82.3 *	51.7
	0.08	78.5 *	57.5
	0.10	75.8	59.8
	0.02	90.4	37.0
	0.04	88.5 *	45.0
8	0.06	82.9 *	51.1
	0.08	78.8 *	56.6
	0.10	75.9	58.9
	0.02	90.3	36.1
	0.04	89.9 *	44.1
9	0.06	83.7 *	50.3
	0.08	79.1 *	55.5
	0.10	76.1	57.8

Table D.1 continued.

H	FL	$100 - \daleth^{1}$ (%)	CI_H^+ (%)
	0.02	90.2	34.9
	0.04	90.3 *	43.1
10	0.06	84.6 *	49.3
	0.08	79.5 *	54.2
	0.10	76.2	56.5
	0.02	90.1	33.3
	0.04	90.2 *	41.5
11	0.06	85.7 *	47.8
	0.08	80.1	52.4
	0.10	76.4	54.8
	0.02	90.0	31.1
	0.04	90.1 *	39.5
12	0.06	87.1 *	45.8
	0.08	80.7	50.2
	0.10	76.5	52.5

The dominated (H, FL) combinations are marked by *.

Table D.2: Solution details for Figures 7.3, 7.4 and 7.5

H	CI_H^+ (%) for $FL = 0.10$	CI_H^- (%)	CI_H^+ (%) for $FL = \infty$	TI_{dev}	Downward deviation	Upward deviation
1	-	-	-	1.208	0.06	0.09
3	28.4	-27.3	22.3	1.216	0.07	0.10
6	25.7	-29.2	19.4	1.246	0.09	0.13
9	22.1	-31.8	15.4	1.295	0.12	0.17
12	16.5	-35.5	9.6	1.370	0.16	0.22

Table D.3: Solution details for Figure 7.6

		low Δ_π		high Δ_π	
Δ_c	Δ_K	φ (%)	CI_K (%)	φ (%)	CI_K (%)
	0.5	86.9	22.8	96.3	14.8
	1	85.3	19.5	94.7	14.5
0.1	1.5	84.3	17.0	93.9	14.0
	2	83.7	15.0	93.2	13.0
	2.5	83.4	9.9	93.0	8.4
	0.5	86.4	26.5	96.0	17.5
	1	84.6	22.8	94.3	16.8
0.5	1.5	83.4	19.8	93.1	15.8
	2	82.6	17.0	92.4	14.0
	2.5	82.1	11.0	91.9	8.5
	0.5	85.0	32.4	94.8	22.4
	1	83.4	29.0	93.3	22.0
1	1.5	82.0	25.4	91.9	20.4
	2	81.0	21.0	90.9	17.0
	2.5	80.3	12.5	90.2	9.0
	0.5	84.4	35.0	94.4	24.0
	1	81.8	30.2	91.8	22.2
1.5	1.5	80.3	27.0	90.4	21.0
	2	79.1	22.4	89.1	17.4
	2.5	78.1	14.0	88.2	9.5

Table D.4: Solution details for Figures 7.7 (a) and 7.8 (a)

FL	Δ_K	CI_K (%)	TI_{ins}
0.01	0.5	15.6	0.261
	1	30.7	0.183
	1.5	36.2	0.157
	2	38.2	0.115
	2.5	38.9	0.109
0.05	0.5	13.6	0.278
	1	27.5	0.215
	1.5	33.2	0.195
	2	35.4	0.154
	2.5	36.1	0.147
0.10	0.5	11.5	0.314
	1	22.2	0.277
	1.5	27.5	0.265
	2	30.1	0.229
	2.5	30.5	0.225

Table D.5: Solution details for Figures 7.7 (b) and 7.8 (b)

Δ_c	Δ_K	CI_K (%)	TI_{ins}
0.1	0.5	4.0	0.09
	1	5.0	0.08
	1.5	5.6	0.08
	2	6.0	0.08
	2.5	6.0	0.08
0.5	0.5	7.0	0.16
	1	11.0	0.15
	1.5	12.4	0.14
	2	13.0	0.12
	2.5	13.2	0.12
1	0.5	9.0	0.22
	1	17.0	0.19
	1.5	20.4	0.19
	2	22.0	0.17
	2.5	22.4	0.16
1.5	0.5	11.0	0.304
	1	23.5	0.250
	1.5	29.0	0.240
	2	31.3	0.200
	2.5	31.8	0.195

Table D.6: Solution details for Figure 7.9 (a)

H	CU		
	$FL = 0.01$	$FL = 0.05$	$FL = 0.1$
3	0.772	0.786	0.819
6	0.773	0.789	0.823
9	0.775	0.793	0.831
12	0.779	0.799	0.847

Table D.7: Solution details for Figure 7.9 (b)

H	CU			
	$\Delta_c = 0.1$	$\Delta_c = 0.5$	$\Delta_c = 1$	$\Delta_c = 1.5$
3	0.781	0.789	0.806	0.819
6	0.786	0.796	0.814	0.823
9	0.792	0.803	0.822	0.831
12	0.793	0.809	0.833	0.847

Table D.8: Correlation matrix used for Figure 7.10

		$\tilde{\varepsilon}_k^1$													$\tilde{\varepsilon}_k^2$												
		k													*k*												
		1	2	3	4	5	6	7	8	9	10	11	12	13	1	2	3	4	5	6	7	8	9	10	11	12	13
$\tilde{\varepsilon}_k^1$	*k* 1	1	0.8	0.5	0.5	0.5	0.5	0.5	0.3	0.3	0.3	0.3	0.3	0.3	-0.8												
	2		1	0.8	0.5	0.5	0.5	0.5	0.5	0.3	0.3	0.3	0.3	0.3		-0.8											
	3			1	0.8	0.5	0.5	0.5	0.5	0.5	0.3	0.3	0.3	0.3			-0.8										
	4				1	0.8	0.5	0.5	0.5	0.5	0.5	0.3	0.3	0.3				-0.5									
	5					1	0.8	0.5	0.5	0.5	0.5	0.5	0.3	0.3					-0.5								
	6						1	0.8	0.5	0.5	0.5	0.5	0.5	0.3						-0.5							
	7							1	0.8	0.5	0.5	0.5	0.5	0.5							-0.5						
	8								1	0.8	0.5	0.5	0.5	0.5								-0.5					
	9									1	0.8	0.5	0.5	0.5									-0.3				
	10										1	0.8	0.5	0.5										-0.3			
	11											1	0.8	0.5											-0.3		
	12												1	0.8												-0.3	
	13													1													-0.3
$\tilde{\varepsilon}_k^2$	*k* 1														1	0.5	0.5	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
	2															1	0.5	0.5	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
	3																1	0.5	0.5	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
	4																	1	0.5	0.5	0.3	0.3	0.3	0.3	0.3	0.3	0.3
	5																		1	0.5	0.5	0.3	0.3	0.3	0.3	0.3	0.3
	6																			1	0.5	0.5	0.3	0.3	0.3	0.3	0.3
	7																				1	0.5	0.5	0.3	0.3	0.3	0.3
	8																					1	0.5	0.5	0.3	0.3	0.3
	9																						1	0.5	0.5	0.3	0.3
	10																							1	0.5	0.5	0.3
	11																								1	0.5	0.5
	12																									1	0.5
	13																										1

Table D.9: Solution details for Figures 7.11 to 7.13

H	CI_{model}(%)	TI_{dev}		TI_{ins}		φ (%)	
	MUFE vs. MUMA	MUFE	MUMA	MUFE	MUMA	MUFE	MUMA
1	10.1	1.228	1.347	0.160	0.250	93.6	91.8
3	16.0	1.236	1.354	0.171	0.261	92.8	90.8
6	19.1	1.266	1.373	0.188	0.274	91.9	89.1
9	21.1	1.315	1.403	0.218	0.294	91.5	87.2
12	22.0	1.390	1.455	0.260	0.322	91.0	85.3

VITA

Mehmet Kayhan was born in Karaman, Türkiye, on January 02, 1979. He received his B.S. and M.S. degrees in Industrial Engineering from the Middle East Technical University in Ankara, earned in July 2001 and August 2003, respectively. He has worked as an engineer in ETI Group of Companies from August 2002 to September 2003, and currently been working for ASELSAN, Inc. His main areas of research are stochastic models of production and inventory systems, operations management and supply chain performance issues.