Production-Inventory Systems: Optimal Control and Empirical Analysis

by

Owen Qun Wu

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE STUDIES

(Business Administration)

THE UNIVERSITY OF BRITISH COLUMBIA

June 2006

Abstract

This thesis consists of three independent essays in the area of production and inventory management.

The first essay is concerned with a competitive equilibrium production-inventory model with application to the petroleum refining industry. We first examine an individual firm controlling production and inventories and facing uncertain raw material price, finished goods price and operating costs. The firm maximizes the expected discounted profit over an infinite horizon. We show that the optimal control is of a threshold type. Next, we consider an economy with many raw material suppliers, production firms and consumers. Both the supply and the demand are uncertain and price-sensitive. We establish and characterize the rational expectations equilibrium price process for this economy, and further derive the equilibrium in an explicit form for a special economy. Finally, we simulate the equilibrium model to reproduce some stylized facts of the petroleum refining industry and fit the model with actual data.

The second essay studies an inventory system that supplies price-sensitive demand modeled by Brownian motion. The optimal pricing and inventory replenishment decisions under both long-run average and discounted objectives are derived, and related to or contrasted with previously known results. In addition, we emphasize the interplay between pricing and replenishment decisions, and the ways in which they react to the demand uncertainty. We show that the joint optimization of both decisions may result in significant profit improvement compared to the traditional method of making decisions separately or sequentially. We also show that multiple price changes result in only a limited profit improvement over the optimal single price.

In the third essay, we examine the inventories of publicly traded American manufacturing companies between 1981 and 2000. The median of inventory holding periods were reduced from 96 days to 81 days. The average rate of inventory reduction is about 2% per year. The greatest reduction was found for work-in-process inventory, which declined by about 6% per year. Finished-goods inventories did not decline. Inventory holdings significantly affect firms' long-term stock returns. Firms with abnormally high inventories have abnormally poor long-term stock returns. Firms with slightly lower than average inventories have good stock returns, but firms with the lowest inventories have only ordinary returns.

Contents

List of Tables

List of Figures

Acknowledgements

My years of pursuing Ph.D. at the University of British Columbia have been immensely rewarding. I am deeply grateful to Professor Hong Chen, my research advisor, for his continuous support and encouragement through every step of my research. Hong introduced me into the field, and has spent tremendous effort reviewing my essays and discussing with me. This thesis is immeasurably better as a result of his feedback and comments.

I would also like to thank Professor Derek Atkins and Professor Murray Frank, who served on my thesis committee and provided invaluable comments on my research works. Derek offered me great assistance in shaping the managerial ideas in the thesis. Murray provided me inspiring guidance on empirical research.

My thanks also go to other faculty members at UBC, who have always been willing to offer their assistance of all kinds. Professor Maurice Queyranne and Professor Harish Krishnan deserve special mention.

I also wish to extend my thanks to Professor David Yao at the Columbia University. I am grateful to his tremendous help to my research.

And finally, my deepest thanks go to my family – my parents, my wife Anna, my son Tanwei, and my daughter Beatrice – who were always there when I needed them. Their constant love and support greatly contributed to my research.

Chapter 1

Introduction

1.1 Motivation

Recent decades have seen substantial and growing theoretical research in the area of production and inventory management. Researchers examine real management problems, construct analytical models to address issues, and provide salient insights for better operations management. Empirical research in production and inventory management is significantly different in nature from most theoretical research. Researchers observe a phenomenon, explore existing theory, develop hypotheses, and collect data or conduct experiments to test the hypotheses. Fisher (2005) reviews a few pioneering empirical research works in operations management, and calls for an acceleration of empirical research in the area of production and operations management.

Through three independent essays, this thesis explores the optimal control theory of production-inventory systems and conducts empirical analysis that either supports the theory or describes inventory behavior that is of general interest in the area of production and inventory management.

The first essay (Chapter 2) establishes an equilibrium model for production, inventory and price behavior based on the firm-level optimal control theory for production-inventory systems. This essay then tests the theory empirically using data from the petroleum refining industry.

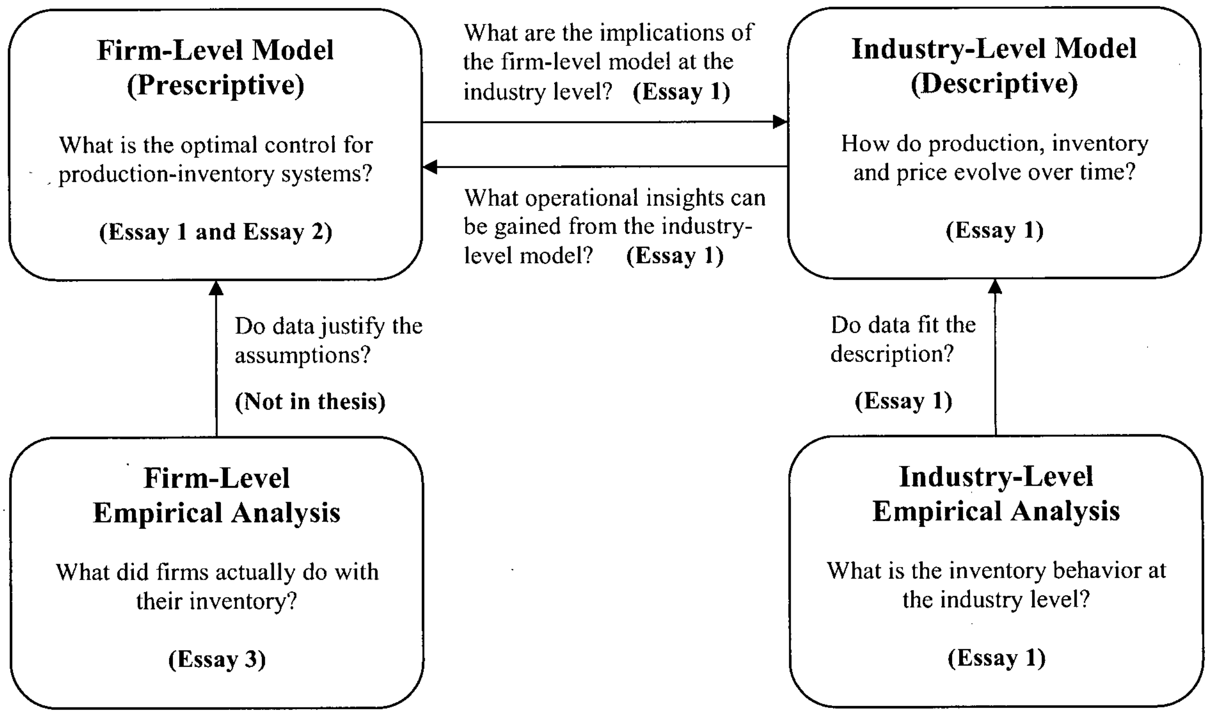
The second essay (Chapter 3) proposes a new demand model and prescribes optimal pricing and replenishment control strategies for an inventory system.

The third essay (Chapter 4) documents some basic empirical findings related to the inventory time trend and the relationship between inventory performance and financial performance of U.S. manufacturing companies over the past two decades.

1.2 Research Framework

The thesis can be summarized using a framework of research subjects, as depicted in Figure 1.1.

Figure 1.1: Research framework



Optimal control of production and inventory at firm level

Theoretical work in the area of production and inventory management generally focuses on questions such as when and how much a firm should produce, how much inventory should be held, what price a firm should charge, and so on. These theories are prescriptive in nature.

The first essay begins with a production-inventory control problem under uncertainty. We examine how a firm should control its production and inventory under stochastically evolving raw material and finished goods prices. This is essentially a continuous-time optimal control problem. Using the classical optimal control theory, we explore the structure of the optimal policy.

The second essay, on inventory control and pricing strategies, also belongs to the prescriptive model category. We consider how a monopoly firm controls its input by managing inventory replenishment, and controls its output by charging varying monopoly prices based on its inventory levels.

Building an industry equilibrium model based on firm-level optimal control theory

It is intriguing to contemplate what industry-level implications can be derived from firm-level production and inventory theories. In other words, based on the firm-level production-inventory control models, can we explain industry-level behavior, such as the evolution and dynamics of production, inventory and price?

In the first essay, based on the firm-level optimal control of production-inventory systems, we develop a competitive rational expectations equilibrium that describes the dynamics of industry inventory level, production level and market prices. We provide a general method for finding equilibrium prices, inventories and production processes. Furthermore, we prove the existence and uniqueness of the equilibrium under specific settings.

The approach used here also exists in other research areas. In finance and economics, many competitive (rational expectations) equilibrium models have been developed based on certain individual decision problems, such as the portfolio choice problem and the income allocation problem.

Understanding firm-level operations based on industry equilibrium

It is also an intriguing question whether additional firm-level operational insights can be gained from the industry equilibrium model.

In the first essay, the industry equilibrium model describes how market prices evolve over time. This helps an individual firm estimate the change of value (appreciation or depreciation) of its assets (i.e., the inventories held), and estimate the opportunity cost of capital tied up in inventory and production investment, thereby making better operations decisions.

Industry-level empirical analysis

Once we have a theory that describes the industry-level production, inventory and price, we can verify whether the data fit the description. A natural question arises: if the descriptive theory is developed under the assumption that all of the individual firms use the optimal control, then how can the data fit the description if most firms in reality deviate from the optimum? Note that deviation could be mitigated by aggregation. Too little inventory in one firm might be compensated by more at another, so that overall inventory levels remain largely unaffected.

In the first essay, we simulate the evolution of the equilibrium based on the descriptive theory, and generate rich inventory and price patterns. Many of these patterns are actually observed in the data from the petroleum refining industry. To explain the actual inventory and price fluctuations, the equilibrium conditions are fitted with the actual data from the petroleum refining industry. The results demonstrate that the estimation captures insights consistent with other known results. For example, significant convenience yield is identified by the model.

Firm-level empirical analysis

Firm-level empirical analysis is more difficult to conduct due to challenges in data gathering. However, the publicly available data can be significantly valuable if exposed to the right empirical research questions. The third essay analyzes the inventory data and the stock return data of all publicly traded U.S. manufacturing companies over two decades. The research documents the inventory time trend and the relationship between inventory performance and long-term financial performance. No theory that currently exists can be perfectly matched to the findings in this essay.

Empirical research at the firm level also provides support and justification for the model assumptions. This thesis does not elaborate on model assumption validation, but we include it in this framework for completeness.

1.3 Summary of Contributions

The first essay makes two main contributions to the production and inventory management area. First, it extends the reach of the traditional production and inventory management theory to the study of industry dynamics and price formation. This approach may be applied to other kinds of production and inventory management theories, thereby extending the impact of the area to other research fields. Second, indirect validation (simulation) and direct validation (empirical test using actual data) of the theory is conducted. The use of empirical analysis to support the theory significantly improves the viability of the theory.

The contributions of the second essay arise from its innovative approach of using the Brownian demand model to study the joint pricing and inventory replenishment problem. First, the Brownian demand model allows us to explicitly and naturally illuminate the impact of demand variability on the optimal pricing and replenishment decisions, whereas results along this line were previously limited to only a few numerical studies. To the best of our knowledge, our study is the first analytical examination of the impact of demand variability. Second, we derive an upper bound for the profit improvement generated from using the dynamic pricing strategy compared to a static strategy. We find that dynamic pricing results in only a limited profit improvement over single price strategy (when both are optimally determined). The relative profit improvement, however, becomes more significant when the profit margin is low. This result is consistent with the numerical results found in the literature.

The third essay establishes two basic empirical points about the inventory holdings of U.S. manufacturing firms over the 1981-2000 period. First, we show that the broad population of manufacturing firms in the U.S. did significantly reduce their inventories. This reduction was particularly marked for work-in-process inventory. Second, we examined the association between abnormal inventory and stock market performance. In the cross-section, abnormal inventory has no effect on the market-to-book ratio or Tobin's q. Over the longer term, inventory does seem to matter. Firms with abnormally high inventory have poor long-term stock market performance. Firms with low, but not extremely low, inventory have unusually good long-term stock market performance; however, firms with the lowest levels of inventory have only ordinary performance. These stock market returns are not accounted for by the conventional financial risk factors.

References

Marshall L. Fisher. 2005. What can we learn about research style from physics, medicine and finance? In *Sixteenth Annual Conference of Production and Operations Management Society*, Chicago, IL.

Chapter 2

Optimal Control and Competitive Equilibrium of Production-Inventory Systems with Application to the Petroleum Refining Industry*

2.1 Introduction

According to the Energy Information Administration, in January 2005, 148 operable refineries were in the U.S. with a total crude distillation capacity of 17.1 million barrels per day. The capacity utilization was 91.3%. The crude oil and petroleum products inventory totaled 1.65 billion barrels in January 2005. In the past a few years, the oil price has become more volatile and extreme movements are also exhibited. Inventory, among many factors, has long been recognized as an important contributor to oil price movements. Zyren (1995) and Ye et al. (2002) statistically identified inventory as being one of the most important variables in explaining short-term crude oil and petroleum product price movements.

This chapter studies the petroleum inventory and price dynamics under a competitive rational expectations equilibrium, and fits the model with actual data. We first consider an individual firm (refinery) making procurement, production, and sales decisions under uncertain raw material (crude oil) price, finished goods (petroleum products) price, and operating costs. The costs include inventory holding cost, production cost, and some other operating costs. The firm maximizes

* An earlier version of this chapter was awarded the First Prize in the 2005 Student Paper Competition of the Manufacturing and Service Operations Management Society at the Institute for Operations Research and the Management Sciences (INFORMS).

the expected long-run discounted profit. Next, we consider an economy with many raw material suppliers, production firms, and consumers. Both the supply and the demand are uncertain and price-sensitive. The rational expectations equilibrium is established. We simulate the evolution of the equilibrium and study its implications for the price and inventory dynamics. We finally fit the theoretical model using the actual data.

Our work is related to several areas of research. The first is the optimal control theory with application to manufacturing systems that range from single machine systems to flexible manufacturing systems. The uncertainties in these systems, including machine failures, repairs, and random demand, are typically modeled as Markov chains or diffusion processes. Various criteria are studied, including discounted costs, long-run average costs, and risk-sensitive criteria, with the objective to find optimal or near-optimal control policies. Fleming, Sethi and Soner (1987) consider a class of discounted optimal control problems with uncertainty governed by a continuous-time Markov chain, and then apply the framework to a production-inventory control problem with demand uncertainty. Akella and Kumar (1986), and Bielecki and Kumar (1988) consider the production rate control for a single failure-prone machine facing constant demand stream. The authors show that the optimal production control is of a threshold type. Sethi et al. (1992) consider a manufacturing system with both capacity uncertainty and demand uncertainty. Presman et al. (1997) further study the optimal control of jobshops. It is also possible to model uncertainties as diffusion processes. Pioneer works include those by Karatzas (1980), Harrison and Taksar (1983), and Harrison et al. (1983). The reader is referred to Sethi et al. (2002) for an extensive survey.

The second area that is related is inventory management under price and cost uncertainty. Scheller-Wolf and Tayur (1998) consider a periodic-review inventory system with capacitated order quantity. The cost parameters depend on the exchange rate which is modeled as a Markov chain. For the uncapacitated situation, Gavirneni (2004) shows that the order-up-to policy is optimal when the unit purchasing cost is fluctuating according to a Markov chain, and provides conditions under which the optimal order-up-to level decreases in the unit purchasing cost.

Petroleum inventory and price behavior have been studied extensively in the economics literature. Pindyck (1994) studies the optimal production and inventory control of a price-taking firm. The firm is treated as a representative agent of the industry, and the first-order conditions are estimated using industry-level data for heating oil and other commodities. The results suggest the existence of significant convenience yield (the flow of benefits to inventory holders). Considine (1997) extends Pindyck's model to include a multi-product cost structure and applies it to the petroleum refining industry. Considine and Larson (2001) formulate a firm's problem as a continuous-time production and inventory control problem and further estimate the optimality equations using crude oil data. All of the above works do not study the evolution of the price.

Rational expectations theory (first proposed by Muth, 1961) has been fruitful in studies of price evolution of storable commodities. In this approach, firms make decisions based on their belief about the price process. A rational firm adopts a belief that is consistent with the market equilibrium price process. The goal of this area of research is to explain the market backwardation phenomena (futures prices below spot prices) observed in many commodity markets. Williams and Wright (1991) provide a comprehensive review of the pioneering efforts in this area. In the basic model, producers sell harvest to the storers in each period and then decide how much to invest for the next period's harvest. The storers decide how much to sell and how much to carry over to the next period. The consumers' demand is price-dependent. Both producers and storers are price-taking competitors and hold rational expectations. Williams and Wright (1991) also extend the basic model to include two commodity markets, where one commodity can be transformed to the other via transportation or production. Deaton and Laroque (1992, 1996) simulate the basic model in an attempt to reproduce some of the stylized facts of commodity price behavior, and to test some of its implications. Routledge et al. (2000) further investigate the implications of the competitive storage model to the term structure of forward prices. Cafiero and Wright (2003) provide an excellent review of the gaps between theory and empirical evidence, and call for a consolidated and reliable theory of production, demand, and storage.

Rational expectations equilibrium models have been recently adopted in the supply chain research. Tayur and Yang (2002) study a natural gas supply chain, in which a competitive market and an oligopoly market are connected by a pipeline monopolist. They introduce the rational expectations into an oligopoly game and further study the equilibrium in the supply chain. More recently, Sapra and Jackson (2005) study an equilibrium model of a supply chain in which the buyers purchase capacity in a competitive capacity market and sell to the end consumers. The forecast for consumer demand is modeled using a continuous-time martingale model of forecast evolution. The buyers' decision regarding the rate of purchase depends on their expectation for the future prices. This is where the rational expectations theory is used to determine the price evolution.

In this chapter, we first characterize the firm-level optimal controls under price and cost uncertainties, and then study the rational expectations equilibrium price and inventory behavior. At the firm level, we analyze a continuous-time production and inventory control problem in which raw material price, finished good price, and operating costs are modeled as diffusion processes. By controlling the rates of inflow, production, and outflow, the firm maximizes the expected discounted profit over an infinite horizon. We prove the optimality of certain threshold type control policies. At the economy level, the fluctuations in inventory, price, and production originate from the uncertainties in demand and supply. The dynamics of price, inventory, and production under the rational expectations equilibrium is characterized by a set of equations that describes certain operational trade-offs. We provide a general procedure to determine the inventory and price dynamics for both raw material and finished goods, and then derive the equilibrium in explicit form for a special economy.

To assess the empirical usefulness of the equilibrium theory, we next study its implications for the behavior of the price, inventory and production. We simulate the industry equilibrium of the special economy under various settings. We find that the simulated equilibrium price and inventory processes exhibit some patterns that are observed in the actual petroleum industry data. Finally, we fit the theoretical model using the actual data. The results generally support the theoretical model.

The rest of this chapter is organized as follows. Section 2.2 analyzes the firm level optimal control problem. The competitive rational expectations equilibrium model is developed in section 2.3, and explicitly solved for a special economy section 2.4. Section 2.5 discusses simulations of the equilibrium model, and section 2.6 presents the results from fitting the model with the data. We conclude the chapter by pointing out several possible extensions in section 2.7.

2.2 Individual Firm's Problem

2.2.1 Problem Setup and Optimality Conditions

We begin our analysis by considering a competitive firm's production and inventory control problem. Let $t \in [0, \infty)$ index time, and let $\mathbf{k}_t = [k_{it}] \in \mathcal{K} \subset \Re^n$ denote the vector of n exogenous factors, such as economy growth rate, interest rate, inflation rate, exchange rate, weather conditions, etc. The factors are governed by a stochastic differential equation of the form:

$$d\mathbf{k}_t = \boldsymbol{\mu}_0(\mathbf{k}_t)dt + \boldsymbol{\sigma}_0(\mathbf{k}_t)d\mathbf{w}_t, \tag{2.1}$$

where $\boldsymbol{\mu}_0 = [\mu_{0i}]$ is an n-dimensional vector function, $\boldsymbol{\sigma}_0 = [\sigma_{0ij}]$ is an $n \times m$ $(m \geq n)$ matrix function, and $\mathbf{w}_t$ is an m-dimensional Wiener process.

Let $\mathbf{p}_t = [p_{1t}, p_{2t}]^\top \in \Re^2$, where p_{1t} is the raw material price and p_{2t} is the finished goods price. The prices follow a diffusion process determined by the following stochastic differential equation:

$$d\mathbf{p}_t = \boldsymbol{\mu}(\mathbf{p}_t, \mathbf{k}_t)dt + \boldsymbol{\sigma}(\mathbf{p}_t, \mathbf{k}_t)d\mathbf{w}_t, \tag{2.2}$$

where $\boldsymbol{\mu} = [\mu_1, \mu_2]^\top$ is a two-dimensional vector function and $\boldsymbol{\sigma} = [\sigma_{ij}]$ is a $2 \times m$ $(m \geq 2)$ matrix function. We assume that $\boldsymbol{\mu}_0, \boldsymbol{\sigma}_0, \boldsymbol{\mu}, \boldsymbol{\sigma}$ satisfy growth and Lipschitz conditions such that (2.1)-(2.2) with any given initial data has a pathwise unique solution. (See Fleming and Soner 1993 and the references therein.)

The firm takes (2.1)-(2.2) as exogenously given. At any time t, the firm chooses a control $\pi_t = (\lambda_t, q_t, s_t)$ from a compact set $\mathcal{U} = [\underline{\lambda}, \overline{\lambda}] \times [\underline{q}, \overline{q}] \times [\underline{s}, \overline{s}] \subset \Re^3_+$, where λ_t is the rate of procuring raw material, s_t is the rate of selling finished goods, and q_t is the production rate. Let $\mathbf{x}_t = [x_{1t}, x_{2t}]^\top \in \Re^2$ be the firm's inventory level, where x_{1t} is the raw material inventory level and x_{2t} is the finished goods inventory level. We assume without loss of generality that one unit of raw material yields one unit of finished goods. Then, the flows must satisfy the following balance equations:

$$dx_{1t} = (\lambda_t - q_t)dt, \qquad dx_{2t} = (q_t - s_t)dt. \tag{2.3}$$

Figure 2.1 illustrates the production-inventory system under consideration.

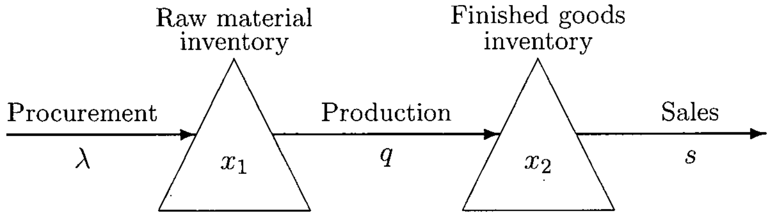


Figure 2.1: A production-inventory system

Let $h(\mathbf{x}, \mathbf{k})$ denote the physical cost of holding $\mathbf{x}$ units of inventory per unit of time under factor

$\mathbf{k}$. This includes the cost of maintaining the storage facilities, but does not include the opportunity cost of capital tied up in the inventory. The latter is accounted for by the firm's discount rate, and will become explicit in the equilibrium conditions later in this chapter (see equation (2.31)). Let $g(q, \mathbf{x}, \mathbf{k})$ denote all the other operating costs per unit of time when the production rate is q, the inventory level is $\mathbf{x}$, and the factor is $\mathbf{k}$. Thus, $h + g$ is the total operating cost rate function. We assume that the operations adjustment cost is zero.

Inventory levels are often required to be within a certain range. This can be modeled into the cost structures: $(g + h)(\mathbf{x}, \mathbf{k}) \to \infty$ as $\mathbf{x}$ approaches to certain boundary. Let $O = \big\{(\mathbf{x}, \mathbf{k}) \in \Re^{n+2} : g(\mathbf{x}, \mathbf{k}) < \infty, h(\mathbf{x}, \mathbf{k}) < \infty\big\}$. Let τ denote the exit time of $(\mathbf{x}_t, \mathbf{k}_t)$ from O, or $\tau = \infty$ if $(\mathbf{x}_t, \mathbf{k}_t) \in O$ for all $t \geq 0$.

The firm maximizes its discounted profit. Let $\rho(\mathbf{k}_t)$ be the firm's discount rate at time t. For example, $\rho(\mathbf{k}_t)$ can be just the interest rate, a component of $\mathbf{k}_t$. The cash flows occurring at time t should be discounted by the following factor:

$$R_t = \int_0^t \rho(\mathbf{k}_u) du. \tag{2.4}$$

We further assume that R_t satisfies the growth condition: $\liminf_{t \to \infty} R_t/t = C_1$, *a.s.* for some $C_1 > 0$. This ensures that the long-run expected discounted profit is finite.

Let us restate the problem in a more systematic way. Let $(\Omega, \mathcal{F}, \mathsf{P})$ be a probability space, and $\{\mathcal{F}_t : t \in [0, \infty)\}$ be a collection of σ-algebras with $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}, \forall s \leq t$. The m-dimensional Wiener process $\mathbf{w}_t$ is $\mathcal{F}_t$-adapted on $[0, \infty)$. The state $(\mathbf{x}_t, \mathbf{p}_t, \mathbf{k}_t)$ evolves according to an $\Re^{n+4}$-valued process given by (2.1)-(2.3). Let $\boldsymbol{\pi} = \{\pi_t : t \in [0, \infty)\}$ denote a control process, and let $\mathcal{A}$ denote the set of admissible controls, i.e., the set of all $\mathcal{F}_t$-progressively measurable, $\mathcal{U}$-valued processes $\boldsymbol{\pi}$ on $[0, \infty)$ with absolutely integrable discounted profit. (The reader is referred to Fleming and Soner 1993 for theoretic background.)

The firm's problem is to choose $\boldsymbol{\pi} \in \mathcal{A}$ to maximize the expected discounted profit:

$$\begin{aligned} V(\mathbf{x}, \mathbf{p}, \mathbf{k}) &= \sup_{\boldsymbol{\pi} \in \mathcal{A}} \mathsf{E}_0^{\pi} \int_0^{\tau} e^{-R_t} \Big(s_t p_{2t} - \lambda_t p_{1t} - g(q_t, \mathbf{x}_t, \mathbf{k}_t) - h(\mathbf{x}_t, \mathbf{k}_t) \Big) dt \\ &\quad \text{subject to (2.1)-(2.4),} \end{aligned} \tag{2.5}$$

where E_0^{π} denotes the expectation with respect to the state process starting at $(\mathbf{x}_0, \mathbf{p}_0, \mathbf{k}_0) = (\mathbf{x}, \mathbf{p}, \mathbf{k})$ and evolving under the control $\boldsymbol{\pi}$.

Note that the optimal control problem (2.5) has a state-dependent discount rate. Its Hamilton-

Jacobi-Bellman (HJB) equation, which is derived in the appendix, is given by

$$\rho(\mathbf{k})V(\mathbf{x},\mathbf{p},\mathbf{k}) = \sup_{\pi\in\mathcal{U}}\Big\{sp_2 - \lambda p_1 - g(q,\mathbf{x},\mathbf{k}) - h(\mathbf{x},\mathbf{k}) + (L^\pi V)(\mathbf{x},\mathbf{p},\mathbf{k})\Big\}, \quad \mathbf{x}\in O, \tag{2.6}$$

where the operator L^π is defined as

$$\begin{aligned} L^\pi &= (\lambda-q)\frac{\partial}{\partial x_1} + (q-s)\frac{\partial}{\partial x_2} + \sum_{i=1}^{2}\mu_i\frac{\partial}{\partial p_i} + \sum_{i=1}^{n}\mu_{0i}\frac{\partial}{\partial k_i} \\ &\quad + \frac{1}{2}\sum_{i,j=1}^{n} b_{ij}\frac{\partial^2}{\partial k_i\partial k_j} + \frac{1}{2}\sum_{i,j=1}^{2} b_{n+i,n+j}\frac{\partial^2}{\partial p_i\partial p_j} + \sum_{i=1,2,\ j=1,\dots,n} b_{n+i,j}\frac{\partial^2}{\partial p_i\partial k_j}, \end{aligned} \tag{2.7}$$

where $b_{ij} = (\mathbf{\Sigma}\mathbf{\Sigma}^\top)_{ij}$ and $\mathbf{\Sigma}^\top = [\boldsymbol{\sigma}_0^\top, \boldsymbol{\sigma}^\top]$. The HJB equation is considered with the growth condition:

$$\lim_{t\to\infty} e^{-R_t}\mathsf{E}_0\big[\chi_{t\le\tau}V(\mathbf{x}_t^*,\mathbf{p}_t,\mathbf{k}_t)\big] = 0.$$

The maximization problem is partially separable (this is partial because the value function is unknown):

$$\lambda^*(\mathbf{x},\mathbf{p},\mathbf{k}) = \arg\sup_{\lambda\in[\underline{\lambda},\overline{\lambda}]}\Big\{\Big(\frac{\partial V(\mathbf{x},\mathbf{p},\mathbf{k})}{\partial x_1} - p_1\Big)\lambda\Big\}, \tag{2.8}$$

$$q^*(\mathbf{x},\mathbf{p},\mathbf{k}) = \arg\sup_{q\in[\underline{q},\overline{q}]}\Big\{\Big(\frac{\partial V(\mathbf{x},\mathbf{p},\mathbf{k})}{\partial x_2} - \frac{\partial V(\mathbf{x},\mathbf{p},\mathbf{k})}{\partial x_1}\Big)q - g(q,\mathbf{x},\mathbf{k})\Big\}, \tag{2.9}$$

$$s^*(\mathbf{x},\mathbf{p},\mathbf{k}) = \arg\sup_{s\in[\underline{s},\overline{s}]}\Big\{\Big(p_2 - \frac{\partial V(\mathbf{x},\mathbf{p},\mathbf{k})}{\partial x_2}\Big)s\Big\}. \tag{2.10}$$

The optimal control for the procurement rate has the following properties. Whenever the input price p_1 falls below its marginal profit $\partial V/\partial x_1$, the firm purchases input to fill its inventory at the maximum rate $\overline{\lambda}$. If the input price is higher than its marginal profit, the firm operates under the minimum input rate $\underline{\lambda}$ (stops input if $\underline{\lambda} = 0$). A similar property holds for the finished goods.

2.2.2 Properties of the Value Function

We state some assumptions on the cost functions, thereby obtaining more properties of the value function and a more structured optimal control.

Assumption 2.1 *Inventory holding cost $h(\mathbf{x},\mathbf{k})$ is strictly convex and increasing in $\mathbf{x}$. The other operating costs function $g(q,\mathbf{x},\mathbf{k})$ is strictly convex in $(q,\mathbf{x})$, decreasing in $\mathbf{x}$ and increasing in q.*

The function $g(q, \mathbf{x}, \mathbf{k})$ represents all operating costs except the holding cost. It includes production cost, production scheduling cost, product delivery cost, etc. These costs are particularly high when inventory levels are very low, as raw material inventory that is too low can cause glitches in production and scheduling, and finished goods inventory that is too low would necessitate overtime shifts and/or expedited delivery. Brennan (1958) and Pindyck (1994) document the evidence on the convenience yield in various commodity markets; the latter explicitly defines the marginal convenience yield as the reduced cost of a marginal unit of inventory. In our model, similar to Pindyck (1994), the marginal convenience yield of inventory x_i is $-\partial g(q, \mathbf{x}, \mathbf{k})/\partial x_i$.

Proposition 2.1 *Under Assumption 2.1, $V(\mathbf{x}, \mathbf{p}, \mathbf{k})$ is strictly concave in $\mathbf{x}$ for any initial price $\mathbf{p}$ and factor $\mathbf{k}$.*

Proof. For initial inventory levels $\mathbf{x}^a \neq \mathbf{x}^b$, let $\boldsymbol{\pi}^a = \{\pi_t^a = (\lambda_t^a, q_t^a, s_t^a) : t \geq 0\}$ and $\boldsymbol{\pi}^b = \{\pi_t^b = (\lambda_t^b, q_t^b, s_t^b) : t \geq 0\}$ be the corresponding optimal controls, and let $\{\mathbf{x}_t^a : t \geq 0\}$ and $\{\mathbf{x}_t^b : t \geq 0\}$ be the corresponding optimal inventory processes.

Now consider the initial inventory $\mathbf{x}^c = (\mathbf{x}^a + \mathbf{x}^b)/2$, and apply a policy $\boldsymbol{\pi}^c = \{\pi_t^c = (\lambda_t^c, q_t^c, s_t^c) : t \geq 0\}$ with $\lambda_t^c = (\lambda_t^a + \lambda_t^b)/2$, $q_t^c = (q_t^a + q_t^b)/2$ and $s_t^c = (s_t^a + s_t^b)/2$. It is clear that $\boldsymbol{\pi}^c$ is an admissible control. From the balance equations in (2.3), the inventory process starting from $\mathbf{x}^c$ controlled by policy $\boldsymbol{\pi}^c$ is $\mathbf{x}_t^c = (\mathbf{x}_t^a + \mathbf{x}_t^b)/2$. Thus, we have

$$
\begin{aligned}
V(\mathbf{x}^c, \mathbf{p}, \mathbf{k}) &\geq \mathsf{E}\int_0^T e^{-R_t}\Big[s_t^c p_{2t} - \lambda_t^c p_{1t} - g(q_t^c, \mathbf{x}^c, \mathbf{k}_t) - h(\mathbf{x}_t^c, \mathbf{k}_t)\Big]dt \\
&> \frac{1}{2}\mathsf{E}\int_0^T e^{-R_t}\Big[(s_t^a + s_t^b)p_{2t} - (\lambda_t^a + \lambda_t^b)p_{1t} - g(q_t^a, \mathbf{x}_t^a, \mathbf{k}_t) - g(q_t^b, \mathbf{x}_t^b, \mathbf{k}_t) \\
&\quad -h(\mathbf{x}_t^a, \mathbf{k}_t) - h(\mathbf{x}_t^b, \mathbf{k}_t)\Big]dt \\
&= \frac{1}{2}\Big[V(\mathbf{x}^a, \mathbf{p}, \mathbf{k}) + V(\mathbf{x}^b, \mathbf{p}, \mathbf{k})\Big],
\end{aligned}
$$

where the first inequality is due to the fact that $\boldsymbol{\pi}^c$ is admissible but not necessarily optimal, and the second inequality is from the definition of $\boldsymbol{\pi}^c$, $\mathbf{x}^c$ and the strict convexity of g and h. This proves the strict concavity of $V(\mathbf{x}, \mathbf{p}, \mathbf{k})$ in $\mathbf{x}$. ∎

Assumption 2.2 $g(q, \mathbf{x}, \mathbf{k}) = g_1(q, \mathbf{k}) + g_2(\mathbf{x}, \mathbf{k})$. *Both $g_2(\mathbf{x}, \mathbf{k})$ and $h(\mathbf{x}, \mathbf{k})$ are supermodular in $\mathbf{x}$.*

Intuitively, the supermodularity implies $\frac{\partial}{\partial x_2}(-\frac{\partial g_2}{\partial x_1}) < 0$ (assuming differentiability), which means that the marginal convenience yield of holding raw material will decrease if more finished goods are available. The supermodularity of h means that the marginal holding cost of raw material increases as more finished goods are in storage.

Proposition 2.2 *Under Assumption 2.1 and 2.2, $V(\mathbf{x}, \mathbf{p}, \mathbf{k})$ is submodular in $\mathbf{x}$ for any initial price $\mathbf{p}$ and factor $\mathbf{k}$.*

The proof is in the appendix. The submodularity of the value function implies that the marginal value of one type of inventory decreases when the other type of inventory accumulates.

The last property of the value function is based on a concept defined as follows.

Definition 2.1 *A real-valued function $f(x_1, x_2)$, $(x_1, x_2) \in \Re^2$, is said to have increasing substitution in x_1 if*

$$f(x_1^a - \delta, x_2 + \delta) - f(x_1^a, x_2) \quad \leq \quad f(x_1^b - \delta, x_2 + \delta) - f(x_1^b, x_2), \qquad \forall x_1^a \leq x_1^b, \delta \geq 0.$$

f is said to have decreasing substitution in x_1 if $-f$ has increasing substitution in x_1. Let $\widetilde{f}(x_2, x_1) := f(x_1, x_2)$. f is said to have increasing substitution in x_2 if $\widetilde{f}$ has increasing substitution in x_2.

Assumption 2.3 $g(q, \mathbf{x}, \mathbf{k}) = g_1(q, \mathbf{k}) + g_2(\mathbf{x}, \mathbf{k})$. *Both $g_2(\mathbf{x}, \mathbf{k})$ and $h(\mathbf{x}, \mathbf{k})$ have decreasing substitution in both x_1 and x_2.*

Notice that if we define $\widehat{f}(y, x_2) := f(y - x_2, x_2)$, then the inequality in Definition 1 implies that $\widehat{f}(y, x_2)$ is supermodular in (y, x_2). Hence, an equivalent definition of increasing substitution in x_1 is that $\widehat{f}(y, x_2)$ is supermodular in (y, x_2). But the notion of substitution becomes useful in the context where production is considered as a transformation process that substitutes one type of inventory for the other. If δ amount of raw material is substituted for (produced into) the same amount of finished goods, then the operating cost changes by $(g_2+h)(x_1-\delta, x_2+\delta)-(g_2+h)(x_1, x_2)$. Assumption 2.3 implies that this change in the operating cost is decreasing in x_1 and increasing in x_2, which means that production is more desirable when the firm has more raw material and less finished goods.

The decreasing substitution property of the cost functions leads to increasing substitution of the value function, as stated in the following proposition. The proof is in the appendix.

Proposition 2.3 *Under Assumption 2.1 and 2.3, $V(\mathbf{x}, \mathbf{p}, \mathbf{k})$ has increasing substitution in both x_1 and x_2 for any initial price $\mathbf{p}$ and factor $\mathbf{k}$.*

2.2.3 Structural Optimal Control Policy

Concavity of the value function gives rise to an optimal control policy of threshold type. Submodularity and increasing substitution properties further characterize the thresholds.

Theorem 2.1 *Under Assumption 2.1, there exist unique thresholds $\widehat{x}_1(x_2, \mathbf{p}, \mathbf{k})$ and $\widehat{x}_2(x_1, \mathbf{p}, \mathbf{k})$ such that the optimal procurement rate and sales rate are*

$$\lambda^*(\mathbf{x}, \mathbf{p}, \mathbf{k}) = \begin{cases} \overline{\lambda}, & \text{if } x_1 < \widehat{x}_1(x_2, \mathbf{p}, \mathbf{k}), \\ \lambda \in [\underline{\lambda}, \overline{\lambda}], & \text{if } x_1 = \widehat{x}_1(x_2, \mathbf{p}, \mathbf{k}), \\ \underline{\lambda}, & \text{otherwise}, \end{cases} \tag{2.11}$$

$$s^*(\mathbf{x}, \mathbf{p}, \mathbf{k}) = \begin{cases} \overline{s}, & \text{if } x_2 > \widehat{x}_2(x_1, \mathbf{p}, \mathbf{k}), \\ s \in [\underline{s}, \overline{s}], & \text{if } x_2 = \widehat{x}_2(x_1, \mathbf{p}, \mathbf{k}), \\ \underline{s}, & \text{otherwise}. \end{cases} \tag{2.12}$$

If the thresholds are finite, they are uniquely determined by

$$\frac{\partial V(\widehat{x}_1, x_2, \mathbf{p}, \mathbf{k})}{\partial x_1} = p_1, \tag{2.13}$$

$$\frac{\partial V(x_1, \widehat{x}_2, \mathbf{p}, \mathbf{k})}{\partial x_2} = p_2. \tag{2.14}$$

Furthermore, under Assumption 2.2 and 2.3, and assuming $\widehat{x}_1(x_2, \mathbf{p}, \mathbf{k})$ and $\widehat{x}_2(x_1, \mathbf{p}, \mathbf{k})$ are differentiable in x_2 and x_1, respectively, then

$$-1 \leq \frac{\partial \widehat{x}_1(x_2, \mathbf{p}, \mathbf{k})}{\partial x_2} \leq 0, \qquad -1 \leq \frac{\partial \widehat{x}_2(x_1, \mathbf{p}, \mathbf{k})}{\partial x_1} \leq 0.$$

Proof. From (2.8), $\lambda = \overline{\lambda}$ is optimal when $\frac{\partial V(\mathbf{x}, \mathbf{p}, \mathbf{k})}{\partial x_1} > p_1$, and $\lambda = \underline{\lambda}$ is optimal when $\frac{\partial V(\mathbf{x}, \mathbf{p}, \mathbf{k})}{\partial x_1} < p_1$. Due to the strict concavity of V, there exists a unique (possibly infinite) threshold $\widehat{x}_1(x_2, \mathbf{p}, \mathbf{k})$ such that the optimal procurement control is of threshold type (2.11). Moreover, if it is finite, the threshold $\widehat{x}_1(x_2, \mathbf{p}, \mathbf{k})$ is determined by equating marginal profit to price, which is just (2.13). Similarly, we can show that the optimal sales rate and its threshold are given by (2.12) and (2.14), respectively.

To prove the trends of the thresholds, first differentiate (2.13) with respect to x_2,

$$\frac{\partial^2 V(\widehat{x}_1, x_2, \mathbf{p}, \mathbf{k})}{\partial x_1^2} \frac{\partial \widehat{x}_1}{\partial x_2} + \frac{\partial^2 V(\widehat{x}_1, x_2, \mathbf{p}, \mathbf{k})}{\partial x_1 \partial x_2} = 0.$$

Under Assumption 2.2 and 2.3, we have

$$\frac{\partial^2 V(\widehat{x}_1, x_2, \mathbf{p}, \mathbf{k})}{\partial x_1^2} \leq \frac{\partial^2 V(\widehat{x}_1, x_2, \mathbf{p}, \mathbf{k})}{\partial x_1 \partial x_2} \leq 0,$$

where the first inequality follows from the increasing substitution property of V, and the second inequality is from the submodularity of V. Hence, $\partial \widehat{x}_1 / \partial x_2$ must be within $[-1, 0]$. The proof for $\partial \widehat{x}_2 / \partial x_1 \in [-1, 0]$ is completely analogous. ∎

The strict concavity of the value function (Proposition 2.1) implies that V has different gradient (with respect to $\mathbf{x}$) at different inventory levels, which in turn implies that there is at most one intersection of the two thresholds $\widehat{x}_1(x_2, \mathbf{p}, \mathbf{k})$ and $\widehat{x}_2(x_1, \mathbf{p}, \mathbf{k})$. The last part of Theorem 2.1 further depicts the trend of the thresholds. Figure 2.2 illustrates the thresholds and the optimal procurement and sales decisions for fixed $\mathbf{p}, \mathbf{k}$.

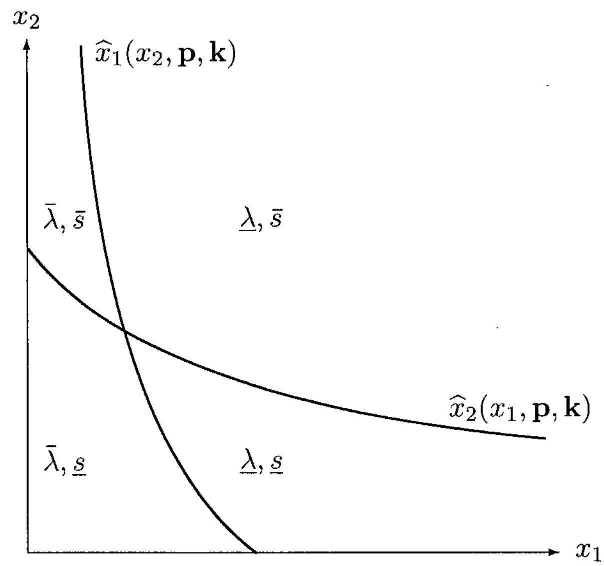


Figure 2.2: Optimal procurement and sales
(Thresholds are not necessarily convex as shown.)

If the optimal production rate q^* is in the interior of $[\underline{q}, \overline{q}]$, then it follows from (2.9) and the strict convexity of g that q^* is uniquely determined by

$$\frac{\partial V(\mathbf{x},\mathbf{p},\mathbf{k})}{\partial x_2} - \frac{\partial V(\mathbf{x},\mathbf{p},\mathbf{k})}{\partial x_1} = \frac{\partial g(q^*,\mathbf{x},\mathbf{k})}{\partial q}. \tag{2.15}$$

If q^* determined above is outside $[\underline{q}, \overline{q}]$, then the nearest boundary point is optimal because the objective function in (2.9) is concave in q. Note that under Assumption 2.1, there need not exist an inventory threshold for the optimal production policy.

To conclude this section, we comment that the thresholds are not easy to find because this requires knowledge of the value function, which is typically difficult to compute. In the next section, we will show that, under the competitive equilibrium, we can find the threshold without even knowing the value function.

2.3 Competitive Rational Expectations Equilibrium

In this section, we consider the competitive equilibrium of an economy where the price $\mathbf{p}$ (taken to be exogenously given by each individual firm in the previous section) is *endogenously* determined. Equilibrium inventory and price dynamics are derived.

2.3.1 The Economy

We consider an economy with three types of individuals: raw material suppliers, producers (production firms), and consumers. Two markets are present: a raw material market and a finished goods market. The suppliers sell raw materials to the producers, who produce finished goods and then sell them to the consumers.

A sufficiently large number of suppliers, producers and consumers exist such that they all behave as price takers. Exactly how many individuals there are in this economy is not crucial. The crucial assumption is that each individual's decisions have negligible influence on the market prices and factors. We also assume that neither suppliers nor consumers keep inventories, nor do they incur cost of supply adjustment or consumption adjustment.

We adopt the representative or aggregate individual approach that is commonly used in the finance and economics literature. The problems of the aggregate individuals are presented below with discussions about how the aggregate individuals are constructed.

Let $A(p_1, \mathbf{k})$ be the aggregate supply function, which is strictly increasing in p_1. The underlying aggregate supplier's problem[1] is

$$\sup_{\{A_t\}} \mathsf{E}_0 \int_0^\infty e^{-R_t}\Big(A_t p_{1t} - \int_0^{A_t} A^{-1}(q, \mathbf{k}_t)dq\Big)dt, \tag{2.16}$$

where $A^{-1}(\cdot, \mathbf{k})$ is the inverse function of $A(\cdot, \mathbf{k})$, interpreted as the aggregate marginal cost function.

Similarly, consumers' utility maximization generates consumers demand. Suppose the aggregate demand function is $D(p_2, \mathbf{k})$, which is strictly decreasing in p_2. Then, the aggregate consumer's problem can be written as

$$\sup_{\{D_t\}} \mathsf{E}_0 \int_0^\infty e^{-R_t}\Big(\int_0^{D_t} D^{-1}(q, \mathbf{k}_t)dq - D_t p_{2t}\Big)dt, \tag{2.17}$$

where $D^{-1}(\cdot, \mathbf{k})$ is the inverse function of $D(\cdot, \mathbf{k})$, interpreted as the marginal utility function.

Let $\mathbf{X}_t = (X_{1t}, X_{2t})$ denote the industry aggregate inventory level, and let $\Pi_t = (\Lambda_t, Q_t, S_t)$ denote the industry aggregate rates of inflow, production, and outflow, respectively. Similar to (2.3), we have the balance equations at the industry level:

$$dX_{1t} = (\Lambda_t - Q_t)dt, \qquad dX_{2t} = (Q_t - S_t)dt. \tag{2.18}$$

The aggregate production firm solves the following problem:

$$\begin{aligned}\widetilde{V}(\mathbf{X}, \mathbf{p}, \mathbf{k}) \;=\; & \sup_{\mathbf{\Pi}\in\widetilde{\mathcal{A}}} \mathsf{E}_0^{\mathbf{\Pi}} \int_0^T e^{-R_t}\Big(S_t p_{2t} - \Lambda_t p_{1t} - G(Q_t, \mathbf{X}_t, \mathbf{k}_t) - H(\mathbf{X}_t, \mathbf{k}_t)\Big)dt \\ & \text{subject to (2.1), (2.2), (2.4), (2.18),}\end{aligned} \tag{2.19}$$

where $\mathsf{E}_0^{\mathbf{\Pi}}$ denotes the expectation with respect to the state process starting at $(\mathbf{X}_0, \mathbf{p}_0, \mathbf{k}_0) = (\mathbf{X}, \mathbf{p}, \mathbf{k})$ and evolving under the control $\mathbf{\Pi} = \{\Pi_t : t \geq 0\}$, $G(Q, \mathbf{X}, \mathbf{k})$ and $H(\mathbf{X}, \mathbf{k})$ are the aggregate operating cost functions, and $\widetilde{\mathcal{A}}$ is the set of admissible controls with control space $\widetilde{\mathcal{U}} := [\underline{\Lambda}, \overline{\Lambda}] \times [\underline{Q}, \overline{Q}] \times [\underline{S}, \overline{S}]$.

Similar to (2.6), the HJB equation for problem (2.19) is

$$\rho(\mathbf{k})\widetilde{V}(\mathbf{X}, \mathbf{p}, \mathbf{k}) = \sup_{\Pi\in\widetilde{\mathcal{U}}} \Big\{S p_2 - \Lambda p_1 - G(Q, \mathbf{X}, \mathbf{k}) - H(\mathbf{X}, \mathbf{k}) + (L^{\Pi}\widetilde{V})(\mathbf{X}, \mathbf{p}, \mathbf{k})\Big\}, \tag{2.20}$$

[1] The aggregate raw material supplier can be constructed as follows. Let j index suppliers, and let A_{jt} denote supplier j's supply rate at time t. Supplier j controls the supply rate to maximize the expected discounted profit: $\sup_{\{A_{jt}\}} \mathsf{E}_0 \int_0^\infty e^{-R_t}(A_{jt}p_{1t} - C_j(A_{jt}, \mathbf{k}_t))dt$, where $C_j(A, \mathbf{k})$ is the cost per unit of time of supplying raw material at rate A under factor $\mathbf{k}$. Assuming $C_j(A, \mathbf{k})$ is strictly increasing and convex in A, and assuming interior solution, the optimal supply rate at time t is determined by $A_{jt} = C_j'^{-1}(p_{1t}, \mathbf{k}_t)$, where $C_j'^{-1}(\cdot, \mathbf{k})$ is the inverse function of $C_j'(\cdot, \mathbf{k})$. Then the aggregate supply function is $A(p_1, \mathbf{k}) = \sum_j C_j'^{-1}(p_1, \mathbf{k})$, which is strictly increasing in p_1 due to the strict convexity of $C_j(\cdot, \mathbf{k})$.

where L^{Π} is defined as in (2.7) with π replaced by Π and $\mathbf{x}$ replaced by $\mathbf{X}$.

The above aggregate production firm can be constructed as follows. We assume that N *identical* production firms exist in this economy, where N is fixed and sufficiently large. All of the firms have homogenous beliefs about the price and factor processes, and have the same cost structures and initial inventory levels. Thus, they essentially solve the same problem (2.5) and choose the same optimal control. Suppose an individual firm's optimal control is $\boldsymbol{\pi}^*$ and the optimal inventory level is $\mathbf{x}^*$. Then, the aggregate inventory levels and the flows can be written as $\mathbf{X}^* = N\mathbf{x}^*$ and $\boldsymbol{\Pi}^* = N\boldsymbol{\pi}^*$. Define the aggregate operating cost functions as $G(Q, \mathbf{X}, \mathbf{k}) := Ng(\frac{Q}{N}, \frac{\mathbf{X}}{N}, \mathbf{k})$ and $H(\mathbf{X}, \mathbf{k}) := Nh(\frac{\mathbf{X}}{N}, \mathbf{k})$. The control space is $\widetilde{\mathcal{U}} := [\underline{\Lambda}, \overline{\Lambda}] \times [\underline{Q}, \overline{Q}] \times [\underline{S}, \overline{S}] := [N\underline{\lambda}, N\overline{\lambda}] \times [N\underline{q}, N\overline{q}] \times [N\underline{s}, N\overline{s}]$.

It can be shown that $\boldsymbol{\Pi}^*$ is optimal for the problem (2.19). Furthermore, the aggregate value function and the individual value function are related as follows:

$$\widetilde{V}(\mathbf{X}, \mathbf{p}, \mathbf{k}) = NV(\tfrac{\mathbf{X}}{N}, \mathbf{p}, \mathbf{k}). \tag{2.21}$$

This aggregation result can be proved by scaling the controls and inventory levels in problem (2.5). To save space, the proof is not shown.

2.3.2 Competitive Rational Expectations Equilibrium

We study the simultaneous equilibria in both raw material market and finished goods market. Under an equilibrium, the total supply of the raw material equals the total industry demand (or total procurement Λ_t^* decided by the firms), and simultaneously, the total supply of the finished goods (or total sales S_t^* decided by the firms) equals the total consumer demand. In general, the equilibrium price that clears the markets is different from the firms' belief about the price. To determine the evolution of the equilibrium price, we invoke the rational expectations hypothesis and formally define the rational expectations equilibrium as follows:

Definition 2.2 *Let $\Pi^*(\mathbf{X}^*, \mathbf{p}, \mathbf{k}; \boldsymbol{\mu}, \boldsymbol{\sigma})$ denote the optimal solution to (2.19), where the dependence on the functional forms of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ (i.e., firms' belief about the price) is made explicit. A competitive rational expectations equilibrium price process $\mathbf{p}^*$ is a price process that evolves according to:*

$$d\mathbf{p}_t^* \quad = \quad \boldsymbol{\mu}^*(\mathbf{p}_t^*, \mathbf{k}_t)dt + \boldsymbol{\sigma}^*(\mathbf{p}_t^*, \mathbf{k}_t)d\mathbf{w}_t, \tag{2.22}$$

and satisfies the following conditions for all $t \geq 0$:

$$A(p_{1t}^*, \mathbf{k}_t) = \Lambda_t^*(\mathbf{X}_t^*, \mathbf{p}_t^*, \mathbf{k}_t; \boldsymbol{\mu}^*, \boldsymbol{\sigma}^*), \tag{2.23}$$

$$D(p_{2t}^*, \mathbf{k}_t) = S_t^*(\mathbf{X}_t^*, \mathbf{p}_t^*, \mathbf{k}_t; \boldsymbol{\mu}^*, \boldsymbol{\sigma}^*). \tag{2.24}$$

Equations (2.23)-(2.24) are the market clearing conditions. Note that the right side of (2.23)-(2.24) is the firms' collective behavior under their belief about the price in (2.22). This definition means that the rational firms' belief about the price is consistent with the market equilibrium price.

The definition of the equilibrium does not allow us to find the equilibrium easily, as the aggregate controls depend on the value function, which is typically difficult to compute. Next, we investigate several important equilibrium properties and provide an effective way of finding the equilibrium.

First, we make the following industry capacity assumption, which ensures that the industry is able to absorb any possible levels of supply and demand (e.g., by building up or drawing down inventories):

Assumption 2.4 $A(\cdot,\cdot) \in (\underline{\Lambda}, \overline{\Lambda})$ *and* $D(\cdot,\cdot) \in (\underline{S}, \overline{S})$.

Similar capacity assumptions have been *implicitly* made by most of the works in the competitive storage theory (Williams and Wright 1991, Routledge et al. 2000, among others) and in the economics literature (Pindyck 1994, among others). In a discrete time situation, their models basically assume that there is no limit on how much the firms can buy or sell in every period.

Assumption 2.4 implies that the equilibrium must not have extreme procurement rates ($\underline{\Lambda}$ or $\overline{\Lambda}$) or extreme sales rate ($\underline{S}$ or $\overline{S}$), otherwise (2.23) and (2.24) cannot hold simultaneously. As all firms are identical, this further implies that no firm will take extreme procurement rates ($\underline{\lambda}$ or $\overline{\lambda}$) or extreme sales rates ($\underline{s}$ or $\overline{s}$) under the equilibrium. From Theorem 2.1, this can be the firm's optimal control only when the inventory levels are *on* the thresholds given by (2.13)-(2.14). That is:

$$p_{it}^* = \frac{\partial V(\mathbf{x}_t^*, \mathbf{p}_t^*, \mathbf{k}_t)}{\partial x_i}, \qquad i = 1, 2, \ \forall t \geq 0.$$

In other words, the firm controls its inventory such that it always stays at the intersection point of the two thresholds depicted in Figure 2.2. An alternative interpretation of the above conditions is as follows. If the raw material price falls below its marginal profit, then all firms will purchase raw material at the maximum rate, which immediately drives up the raw material price until it equals the marginal profit. On the other hand, if the raw material price rises above its marginal profit, all firms will purchase raw materials at the minimum rate, driving down the price until it equals the marginal profit. This type of interpretation is often seen in the above-mentioned works. (See, for example, Williams and Wright 1991, page 26).

The relation (2.21) implies that $\partial V(\mathbf{x}_t^*, \mathbf{p}_t^*, \mathbf{k}_t)/\partial x_i = \partial \widetilde{V}(\mathbf{X}_t^*, \mathbf{p}_t^*, \mathbf{k}_t)/\partial X_i$ for $i = 1, 2$. Thus, in each market, the equilibrium price must equal the marginal profit of the goods to the industry:

$$p_{it}^* = \frac{\partial \widetilde{V}(\mathbf{X}_t^*, \mathbf{p}_t^*, \mathbf{k}_t)}{\partial X_i}, \qquad i = 1, 2, \ \forall t \geq 0. \tag{2.25}$$

The left side of (2.25) is a diffusion process, and the right side is also a diffusion process since $(\mathbf{X}_t^*, \mathbf{p}_t^*, \mathbf{k}_t)$ is a diffusion process and $\partial \widetilde{V}/\partial X_i$ is continuously differentiable. Matching the drift coefficients of the diffusion processes on both sides, we have

$$\mu_i^*(\mathbf{p}_t^*, \mathbf{k}_t) = L^{\Pi^*} \frac{\partial \widetilde{V}(\mathbf{X}_t^*, \mathbf{p}_t^*, \mathbf{k}_t)}{\partial X_i}, \qquad i = 1, 2, \ \forall t \geq 0. \tag{2.26}$$

On the other hand, differentiating the HJB equation (2.20) with respect to X_1 and X_2 gives

$$\rho(\mathbf{k}_t) \frac{\partial \widetilde{V}(\mathbf{X}_t^*, \mathbf{p}_t^*, \mathbf{k}_t)}{\partial X_i} = -\frac{\partial G(Q_t^*, \mathbf{X}_t^*, \mathbf{k}_t)}{\partial X_i} - \frac{\partial H(\mathbf{X}_t^*, \mathbf{k}_t)}{\partial X_i} + L^{\Pi^*} \frac{\partial \widetilde{V}(\mathbf{X}_t^*, \mathbf{p}_t^*, \mathbf{k}_t)}{\partial X_i}, \quad i = 1, 2. \tag{2.27}$$

where we have used the fact that $\frac{\partial}{\partial X_i} L^{\Pi} = L^{\Pi} \frac{\partial}{\partial X_i}$ for $i = 1, 2$.

Combining equations (2.25)-(2.27), we can eliminate the unknown value function $\widetilde{V}$ and obtain the following equilibrium price and inventory relations:

$$\mu_i^*(\mathbf{p}_t^*, \mathbf{k}_t) - \frac{\partial G(Q_t^*, \mathbf{X}_t^*, \mathbf{k}_t)}{\partial X_i} = \frac{\partial H(\mathbf{X}_t^*, \mathbf{k}_t)}{\partial X_i} + \rho(\mathbf{k}_t) p_{it}^*, \qquad i = 1, 2, \ \forall t \geq 0. \tag{2.28}$$

From (2.15) and (2.25), if the equilibrium production rate Q_t^* is in the interior of $[\underline{Q}, \overline{Q}]$, then it is uniquely determined by

$$p_{2t}^* - p_{1t}^* = \frac{\partial G(Q_t^*, \mathbf{X}_t^*, \mathbf{k}_t)}{\partial Q}. \tag{2.29}$$

If Q_t^* solved from the above equation is outside the control space $[\underline{Q}, \overline{Q}]$, then the optimal production is the nearest boundary point. In any case, we can write

$$Q_t^* = Q_t^*(\mathbf{p}_t^*, \mathbf{X}_t^*). \tag{2.30}$$

Similar conditions to those in (2.28)-(2.30) also holds at the individual firm level, that is,

$$\mu_i^*(\mathbf{p}_t^*, \mathbf{k}_t) - \frac{\partial g(q_t^*, \mathbf{x}_t^*, \mathbf{k}_t)}{\partial x_i} = \frac{\partial h(\mathbf{x}_t^*, \mathbf{k}_t)}{\partial x_i} + \rho(\mathbf{k}_t) p_{it}^*, \qquad i = 1, 2, \tag{2.31}$$

$$p_{2t}^* - p_{1t}^* = \frac{\partial g(q_t^*, \mathbf{x}_t^*, \mathbf{k}_t)}{\partial q}, \quad \text{if } q \in (\underline{q}, \overline{q}). \tag{2.32}$$

Equation (2.31) describes the trade-offs in making inventory decisions. Suppose the firm holds $\mathbf{x}_t^*$ units of inventory, and considers purchasing an extra unit of inventory at time t, holding it from t to $t + dt$, and then selling it at $t + dt$. This extra unit may either appreciate or depreciate in its market value; its expected price change is $\mu_i^*(\mathbf{p}_t^*, \mathbf{k}_t)dt$. This extra unit of inventory helps to reduce operating cost by $-\partial g/\partial x_i dt$, but incurs extra holding cost $-\partial h/\partial x_i dt$. Purchasing this extra unit and holding it for dt incurs opportunity cost $\rho(\mathbf{k}_t)p_{it}^* dt$. Thus, every firm trades off the costs and benefits of holding a marginal unit of inventory; in equilibrium, the equality in (2.31) is maintained all the time.

Equation (2.31) actually characterizes the thresholds for procuring raw materials and selling finished goods. For any $\mathbf{k}_t$, $\mathbf{p}_t^*$ and q_t^*, equation (2.31) describes two curves on the inventory plane, which are just the thresholds. In the firm's problem in Section 2.2, the characterization of the thresholds involve the unknown value function. Here, in the competitive equilibrium, the value function is related to the market price, and the thresholds become more explicit and provide more insights about firm-level trade-offs. If the benefit of holding an extra unit of inventory is higher (lower) than the cost, the firm is below (above) the threshold, so the firm would increase (decrease) inventory until it reaches the threshold.

Equation (2.32) says that the firm trades off the costs and benefits of producing a marginal unit of product, and maintains the marginal cost of production to be equal to the price spread (i.e., gross margin).

If we subtract equation (2.31) with $i = 1$ from the same equation with $i = 2$, and combine the resulting equation with that in (2.32), we have

$$L^{\pi^*}\frac{\partial g}{\partial q} = \left(\frac{\partial g}{\partial x_2} + \frac{\partial h}{\partial x_2} - \frac{\partial g}{\partial x_1} - \frac{\partial h}{\partial x_1}\right) + \rho(\mathbf{k}_t)\frac{\partial g}{\partial q}, \tag{2.33}$$

where L^{π^*} is defined as in (2.7), and all of the derivatives are evaluated at the equilibrium (optimal) point. This equation describes the temporal trade-off of production. Consider producing an extra unit at time t, holding it from t to $t + dt$, and then producing one unit less at $t + dt$. The term in the parentheses is the change in the holding cost and other operating cost if producing one more unit. The last term on the right side is the opportunity cost of producing one more unit. They must equal to the expected change of marginal production cost on the left side of the equation.

We now show how the equilibrium processes can be endogenously determined. First note that the market clearing conditions (2.23)-(2.24) together with the balance equation (2.18) and the optimal production (2.30) lead to

$$A(p_{1t}^*, \mathbf{k}_t) = Q_t^*(\mathbf{p}_t^*, \mathbf{X}_t^*) + dX_{1t}^*/dt, \qquad D(p_{2t}^*, \mathbf{k}_t) = Q_t^*(\mathbf{p}_t^*, \mathbf{X}_t^*) - dX_{2t}^*/dt. \tag{2.34}$$

Using the four equations in (2.28) and (2.34), we can apply the following procedure to determine the equilibrium price and inventory processes. We first choose a functional specification of the price drift $\boldsymbol{\mu}^*(\cdot,\cdot)$, and then write inventory $\mathbf{X}^*$ as a function of $\mathbf{p}^*$ using (2.28). Differentiating this function we can write inventory change $d\mathbf{X}^*$ in terms of $\mathbf{p}^*$ and $d\mathbf{p}^*$. Replacing $\mathbf{X}^*$ and $d\mathbf{X}^*$ in (2.34) by the functions of $\mathbf{p}^*$ and $d\mathbf{p}^*$ obtained previously, we have a differential equation for $\mathbf{p}^*$. To find a rational expectations equilibrium, we equate the drift of $\mathbf{p}^*$ with the initially chosen $\boldsymbol{\mu}^*(\cdot,\cdot)$ and solve for $\boldsymbol{\mu}^*(\cdot,\cdot)$. The diffusion part of the price belief can be assigned to be equal to that of $\mathbf{p}^*$. Thus, an equilibrium price process is determined. To find the equilibrium inventory process, we can write $\mathbf{p}^*$ as a function of $\mathbf{X}^*$ using (2.28). Substituting $\mathbf{p}^*$ in (2.34) by that function gives a differential equation that governs the equilibrium evolution of $\mathbf{X}^*$.

If the above procedure yields a solution, then it must be a rational expectations equilibrium, because (2.34) ensures markets clearing and the belief about price matches the equilibrium price. The existence and the uniqueness of the rational expectations equilibrium can be established for some special economy described in the following section. It is an open question whether the equilibrium exists and is unique in general.

2.4 A Special Economy

In this section, we consider a special case of the economy studied in the previous section. We derive the rational expectations equilibrium explicitly, and investigate the equilibrium inventory and price dynamics. The special economy is detailed below.

1. Two economic factors $\mathbf{k} = [k_1, k_2]^\top$ are used to model the random fluctuation in the supply and demand. The factors follow an exogenously given two-dimensional mean-reverting process:

$$d\mathbf{k}_t = \mathbf{K}\mathbf{k}_t dt + \boldsymbol{\sigma}_0(\mathbf{k}_t) d\mathbf{w}_t, \tag{2.35}$$

where $\mathbf{K} = \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix}$ and $\boldsymbol{\sigma}_0(\mathbf{k}_t)$ is a $2 \times m$ matrix function. The supply and demand rates are specified as follows:

$$A(p_{1t}, \mathbf{k}_t) = A_{01} + A_1 p_{1t} + k_{1t}, \qquad D(p_{2t}, \mathbf{k}_t) = A_{02} - A_2 p_{2t} + k_{2t}, \tag{2.36}$$

where A_{01}, A_{02}, A_1 and A_2 are given positive constants. A necessary condition for Assumption 2.4 to be satisfied is that $\mathbf{k}_t$ must be uniformly bounded for all $t \geq 0$, which requires that the eigenvalues of $\mathbf{K}$ have negative real parts, and that $\boldsymbol{\sigma}_0(\mathbf{k}) \to 0$ as $\mathbf{k}$ approaches to the boundary.

2. Quadratic convex operating cost function (independent of $\mathbf{k}$):

$$G(Q, \mathbf{X}) + H(\mathbf{X}) \;=\; a_0 + a_1 Q + \frac{1}{2} a_2 Q^2 - \mathbf{c}^\mathsf{T} \mathbf{X} + \frac{1}{2} \mathbf{X}^\mathsf{T} \mathbf{C} \mathbf{X}, \tag{2.37}$$

where $a_1 > 0$, $a_2 > 0$, $\mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} > 0$, and $\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix}$ is symmetric and positive definite ($c_{11} > 0, c_{22} > 0$, $c_{12} = c_{21}$ and $c_{11}c_{22} > c_{12}^2$). The value of a_0 does not affect the optimal decision. We also assume that the range $[\underline{Q}, \overline{Q}]$ is large enough so that the optimal production rate is always an interior solution. We will prove that the equilibrium production rate is uniformly bounded, so this condition can be met indeed.

3. All firms have constant discount rate: $\rho(\mathbf{k}) \equiv \rho$, where $\rho > 0$ is a given constant.

4. All firms *believe* that the price follows a stochastic process of the following form:

$$d\mathbf{p}_t = \big(\mathbf{B}(\mathbf{p}_t - \mathbf{m}) + \mathbf{D}\mathbf{k}_t\big)dt + \boldsymbol{\sigma}(\mathbf{p}_t, \mathbf{k}_t) d\mathbf{w}_t, \tag{2.38}$$

where $\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix}$, $\mathbf{m} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix}$, and $\mathbf{D} = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix}$ are parameters, and $\boldsymbol{\sigma}$ is a $2 \times m$ matrix function. These are the parameters that will be determined *endogenously* from the equilibrium. We are only interested in those equilibrium price processes that have stationary distributions. This requires all the eigenvalues of $\mathbf{B}$ to have negative real parts, or equivalently,

$$\mathsf{tr}[\mathbf{B}] < 0, \qquad \mathsf{det}[\mathbf{B}] > 0. \tag{2.39}$$

In Theorem 2.2, we prove that in equilibrium, the eigenvalues of $\mathbf{B}$ are negative real numbers.

Given the economy specified as above, the system (2.28)-(2.29) and (2.34) becomes a linear system of equations (for notational convenience, we omit superscript '*' in equilibrium quantities hereafter):

$$\begin{cases} \mathbf{B}(\mathbf{p}_t - \mathbf{m}) + \mathbf{D}\mathbf{k}_t - \rho\mathbf{p}_t &= -\mathbf{c} + \mathbf{C}\mathbf{X}_t, \\ p_{2t} - p_{1t} &= a_1 + a_2 Q_t, \\ A_{01} + A_1 p_{1t} + k_{1t} &= Q_t + dX_{1t}/dt, \\ A_{02} - A_2 p_{2t} + k_{2t} &= Q_t - dX_{2t}/dt. \end{cases} \tag{2.40}$$

From the last three equations of (2.40), we have

$$\frac{d\mathbf{X}_t}{dt} = \mathbf{A}\mathbf{p}_t + \mathbf{a} + \mathbf{I}_1\mathbf{k}_t, \tag{2.41}$$

where

$$\mathbf{A} = \begin{bmatrix} A_1 + \frac{1}{a_2} & -\frac{1}{a_2} \\ -\frac{1}{a_2} & A_2 + \frac{1}{a_2} \end{bmatrix}, \qquad \mathbf{a} = \begin{bmatrix} A_{01} + \frac{a_1}{a_2} \\ -A_{02} - \frac{a_1}{a_2} \end{bmatrix}, \qquad \mathbf{I}_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Differentiating the first equation of (2.40), we have

$$d\mathbf{p}_t = \mathbf{C}_1\, d\mathbf{X}_t - \mathbf{D}_1\, d\mathbf{k}_t,$$

where $\mathbf{C}_1 \equiv \mathbf{B}_1\mathbf{C}$, $\mathbf{D}_1 \equiv \mathbf{B}_1\mathbf{D}$, and $\mathbf{B}_1 = (\mathbf{B} - \rho\mathbf{I})^{-1}$, where $\mathbf{I}$ is an identity matrix. Substituting (2.41) and (2.35) into the above equation yields

$$d\mathbf{p}_t = \left[\mathbf{C}_1\mathbf{A}(\mathbf{p}_t + \mathbf{A}^{-1}\mathbf{a}) + (\mathbf{C}_1\mathbf{I}_1 - \mathbf{D}_1\mathbf{K})\mathbf{k}_t\right]dt - \mathbf{D}_1\boldsymbol{\sigma}_0(\mathbf{k}_t)d\mathbf{w}_t. \tag{2.42}$$

To find the rational expectations equilibrium, we match the coefficients in (2.42) with those in (2.38), which leads to the following conditions:

$$\mathbf{B}^2 - \rho\mathbf{B} = \mathbf{C}\mathbf{A}, \qquad \mathbf{m} = -\mathbf{A}^{-1}\mathbf{a}, \qquad \mathbf{D} = \mathbf{C}_1\mathbf{I}_1 - \mathbf{B}_1\mathbf{D}\mathbf{K}, \qquad \boldsymbol{\sigma}(\mathbf{p}, \mathbf{k}) = -\mathbf{D}_1\boldsymbol{\sigma}_0(\mathbf{k}). \tag{2.43}$$

We solve the first equation above for $\mathbf{B}$, and then solve the third equation for $\mathbf{D}$. The second and the fourth equations in (2.43) directly specify the equilibrium value of $\mathbf{m}$ and $\boldsymbol{\sigma}(\mathbf{p}, \mathbf{k})$. The following theorem asserts the existence and uniqueness of such solution.

Theorem 2.2 *Suppose the firms' belief of price is in the form of (2.38). Then, there exists a unique rational expectations equilibrium with stationary distributions. In particular,*

(i) There exists a unique solution to $\mathbf{B}^2 - \rho\mathbf{B} = \mathbf{C}\mathbf{A}$ that satisfies (2.39), which is:

$$\mathbf{B} = \frac{\rho}{2}\mathbf{I} - \sqrt{\mathbf{F}},$$

where $\mathbf{F} = \mathbf{C}\mathbf{A} + \frac{\rho^2}{4}\mathbf{I}$ and $\sqrt{\mathbf{F}} := \mathbf{V}\mathsf{diag}[\sqrt{\xi_1}, \sqrt{\xi_2}]\mathbf{V}^{-1}$, where $\xi_1 > 0$ and $\xi_2 > 0$ are eigen-

values of $\mathbf{F}$*, and* $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2]$ *contains linearly independent eigenvectors corresponding to* ξ_1 *and* ξ_2*, respectively. Furthermore, the eigenvalues of* $\mathbf{B}$ *are negative real numbers.*

(ii) There exists a unique solution to $\mathbf{D} = \mathbf{C}_1\mathbf{I}_1 - \mathbf{B}_1\mathbf{D}\mathbf{K}$*, which is:*

$$\mathsf{vec}[\mathbf{D}] = \left(\mathbf{K}^\top \otimes \mathbf{B}_1 + \mathbf{I}\right)^{-1}\mathsf{vec}[\mathbf{C}_1\mathbf{I}_1],$$

where $\mathsf{vec}[\cdot]$ *denotes the vector formed by collecting the columns of a matrix in one long vector:* $\mathsf{vec}[\mathbf{D}] = [d_{11}, d_{21}, d_{12}, d_{22}]^\top$*, and* $\otimes$ *denotes the Kronecker product:* $\mathbf{K}^\top \otimes \mathbf{B}_1 = \begin{bmatrix} k_{11}\mathbf{B}_1 & k_{21}\mathbf{B}_1 \\ k_{12}\mathbf{B}_1 & k_{22}\mathbf{B}_1 \end{bmatrix}$.

(iii) With $\mathbf{B}$ *and* $\mathbf{D}$ *solved above, the unique equilibrium price and inventory are given by:*

$$\begin{aligned} d\mathbf{p}_t &= \left(\mathbf{B}(\mathbf{p}_t - \mathbf{m}) + \mathbf{D}\mathbf{k}_t\right)dt \;-\; \mathbf{D}_1\sigma_0(\mathbf{k}_t)d\mathbf{w}_t, \\ d\mathbf{X}_t &= \left(\mathbf{B}^\top(\mathbf{X}_t - \mathbf{m}_X) + (\mathbf{I}_1 - \mathbf{A}\mathbf{D}_1)\mathbf{k}_t\right)dt, \end{aligned}$$

where $\mathbf{m} = -\mathbf{A}^{-1}\mathbf{a}$ *and* $\mathbf{m}_X = \mathbf{C}^{-1}(\mathbf{c} - \rho\mathbf{m})$*. Furthermore, the equilibrium processes have the following relations:*

$$\begin{aligned} \mathbf{p}_t &= \mathbf{m} + \mathbf{C}_1(\mathbf{X}_t - \mathbf{m}_X) - \mathbf{D}_1\mathbf{k}_t, \\ Q_t &= \left([\,-1,\ 1\,]\ \mathbf{p}_t - a_1\right)/a_2, \\ \begin{bmatrix} A_t \\ S_t \end{bmatrix} &= \begin{bmatrix} A_{01} \\ A_{02} \end{bmatrix} + \begin{bmatrix} A_1 & 0 \\ 0 & -A_2 \end{bmatrix}\mathbf{p}_t + \mathbf{k}_t, \end{aligned}$$

and $(\mathbf{p}, \mathbf{X}, \mathbf{\Pi})$ *is uniformly bounded for all* $t \geq 0$.

It is worth noting that the "mean" price $\mathbf{m}$ is such that the long run average rate of raw material flow $A_{01} + A_1 m_1$ equals the average production rate $\frac{m_2 - m_1 - a_1}{a_2}$, and equals the average rate of finished goods flow $A_{02} - A_2 m_2$. The "mean" inventory level $\mathbf{m}_X$ is the minimizer of $\frac{1}{2}\mathbf{X}^\top\mathbf{C}\mathbf{X} - \mathbf{c}^\top\mathbf{X} + \rho\mathbf{m}^\top\mathbf{X}$, which is the operating cost of inventory plus the opportunity cost of inventory investment. Lemma A.1.3 in the Appendix shows that $\mathbf{C}_1 < 0$, which implies that at the same level of $\mathbf{k}$, higher inventory is associated with lower price. The parameters of the economy can be chosen such that the equilibrium process is uniformly bounded in a positive compact set. First, we can choose parameters such that $\mathbf{m}$ and $\mathbf{m}_X$ are positive. Notice that if $\mathbf{k}_t = 0$ for all $t \geq 0$, then the equilibrium price and inventory processes monotonically approach to $\mathbf{m}$ and $\mathbf{m}_X$, respectively. We can choose $\mathcal{K}$, the range of $\mathbf{k}$, sufficiently small such that the equilibrium price and inventory processes are positive all the time. To save space, we do not discuss these conditions in detail.

The above theorem asserts the uniqueness of the competitive rational expectations equilibrium under the class of beliefs in (2.38) and the stability condition (2.39). In fact, other rational expectation equilibria are possible, e.g., $\mathbf{B} = \frac{\rho}{2}\mathbf{I} + \sqrt{\mathbf{F}}$, which is also a solution to the first equation in (2.43), though this solution results in unstable equilibrium processes.

We next examine how demand and supply fluctuations affect the equilibrium processes. We will only consider the case where the demand and supply fluctuations are independent. That is, $\mathbf{K} = \mathsf{diag}[-\kappa_1, -\kappa_2]$ with $\kappa_1 > 0, \kappa_2 > 0$, and $\boldsymbol{\sigma}_0(\mathbf{k}_t) = \mathsf{diag}[\sigma_{01}(k_{1t}), \sigma_{02}(k_{2t})]$, and thus,

$$dk_{it} = -\kappa_i k_{it} dt + \sigma_{0i}(k_{it}) dw_{it}, \quad i = 1, 2. \tag{2.44}$$

We refer to the diffusion terms in (2.44) as *supply and demand shocks.* These shocks have instantaneous effects on prices, which is $-\mathbf{D}_1 \boldsymbol{\sigma}_0(\mathbf{k}_t) d\mathbf{w}_t = -\mathbf{D}_1 \begin{bmatrix} \sigma_{01}(k_{1t}) dw_1 \\ \sigma_{02}(k_{2t}) dw_2 \end{bmatrix}$ seen from Theorem 2.2, but no instantaneous effects on inventory levels (inventory is a smooth process). The supply and demand levels affect price drift through the term $\mathbf{D}\mathbf{k}_t$, and affect the drift of inventory levels via the term $(\mathbf{I}_1 - \mathbf{A}\mathbf{D}_1)\mathbf{k}_t$. These effects are described in the following theorem with a proof included in the appendix.

Theorem 2.3 *Suppose the factor process follows (2.44). Then, under a rational expectations equilibrium,*

(i) The elements of $\mathbf{D}_1$ *have signs* $\begin{bmatrix} + & - \\ + & - \end{bmatrix}$. *That is, a positive supply shock has a negative instantaneous impact on* p_1 *and* p_2*, and a positive demand shock has a positive instantaneous impact on* p_1 *and* p_2*;*

(ii) The composition effect of a positive supply shock and a positive demand shock can be negative on p_1*, but positive on* p_2*;*

(iii) The elements of $\mathbf{D}$ *have signs* $\begin{bmatrix} - & + \\ - & + \end{bmatrix}$. *That is, the price drift is decreasing in the supply level and increasing in the demand level;*

(iv) If $b_{11} < 0$ *and* $b_{22} < 0$*, then the elements of* $\mathbf{I}_1 - \mathbf{A}\mathbf{D}_1$ *have signs* $\begin{bmatrix} + & \\ & - \end{bmatrix}$. *That is, the drift of raw material inventory is increasing in the supply level, and the drift of finished goods inventory is decreasing in the demand level.*

Most results in the above theorem are rather intuitive. Part (ii) implies that it is possible to see negative correlation between the input price and the output price. Part (iv) is conditioned on the signs of the diagonal elements of $\mathbf{B}$. We find that this condition may not hold when κ_1 is much smaller than κ_2, and c_{11} is much larger than c_{22}. Intuitively, if there is a raw material supply disruption that would take long to recover (κ_1 small), and lower raw material inventory incurs more operational costs than lower finished goods, then increasing raw material inventory may become preferable (and thus the up-left element in $\mathbf{I}_1 - \mathbf{AD}_1$ becomes negative).

2.5 Simulation

An important step in assessing the empirical usefulness of the theory is to gain a better understanding of its implications for the behavior of prices, inventories, and production. In this section, we simulate and study the evolution of the equilibrium processes for the special economy in Section 2.4.

Different scenarios can be generated by varying the parameters of the market demand and supply functions in (2.36) and the operating cost function in (2.37). For each scenario, we numerically determine the equilibrium processes based on Theorem 2.2. We choose parameter values such that the inventory levels, price, and production rate are positive throughout the simulation.

Two types of simulations are conducted. First, we study the equilibrium responses to a single supply or demand shock (i.e., impulse response functions). In other words, we simulate the equilibrium driven by a particular sample path of $\mathbf{w}_t$ that increases only in a very short period and is constant everywhere else. In the second simulation, we examine how the equilibrium evolves under "continuously" fluctuating demand and supply.

Example 2.1 Suppose the demand and supply functions in (2.36) are

$$A(p_{1t}, \mathbf{k}_t) = 16 + 0.5p_{1t} + k_{1t}, \qquad D(p_{2t}, \mathbf{k}_t) = 30 - 0.5p_{2t} + k_{2t},$$

where $\mathbf{k}$ satisfies (2.44): $dk_{it} = -k_{it}dt + 25\big(1 - (\frac{k_{it}}{20})^4\big)dw_{it}$, for $i = 1, 2$. That is, $\mathbf{K} = \mathsf{diag}[-1, -1]$ and $\boldsymbol{\sigma}_0(\mathbf{k}_t) = 25\,\mathsf{diag}\big[1 - (\frac{k_{1t}}{20})^4, 1 - (\frac{k_{2t}}{20})^4\big]$, k_1 and k_2 are independent factors, and bounded within $[-20, 20]$. Suppose the operating cost function in (2.37) is

$$G(Q, \mathbf{X}) + H(\mathbf{X}) = a_0 + 0.2Q + 0.1Q^2 - 10X_1 - 9X_2 + 0.05(X_1^2 + X_2^2 + X_1X_2),$$

where a_0 is such that the above operating cost is always positive, but the value of a_0 does not affect the equilibrium. Suppose the firm's discount rate is 0.05 per unit of time. In this case, the equilibrium is given by Theorem 2.2 with parameters:

$$\mathbf{B} = \begin{bmatrix} -0.475 & 0.225 \\ 0.225 & -0.475 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} -.0720 & 0.043 \\ -0.043 & 0.072 \end{bmatrix}, \quad \mathbf{m} = \begin{bmatrix} 11.71 \\ 16.29 \end{bmatrix}, \quad \mathbf{m}_X = \begin{bmatrix} 70.95 \\ 46.38 \end{bmatrix}.$$

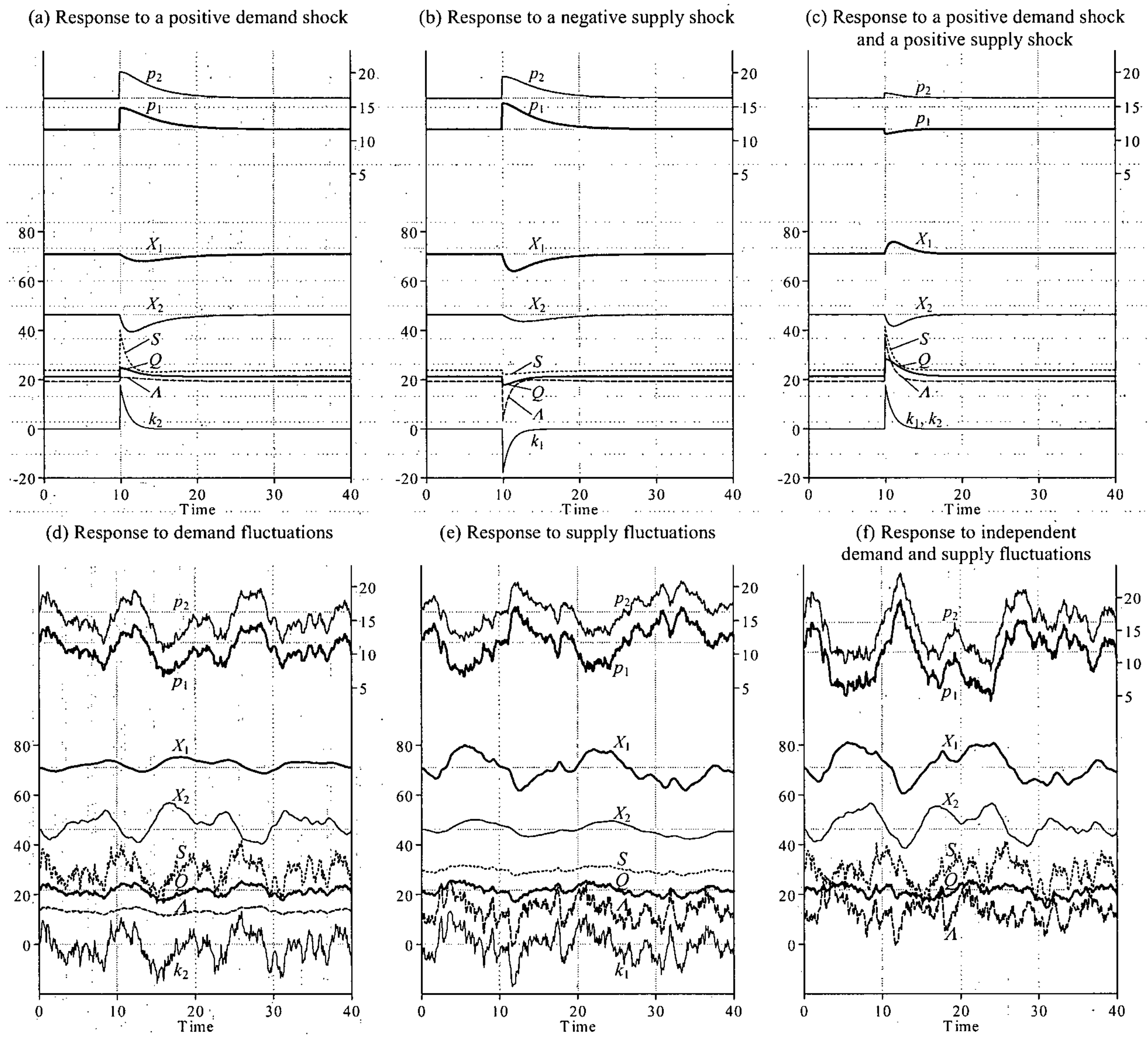
The impulse responses are shown in Figure 2.3(a)-(c), and the responses to "continuously" fluctuating factors are shown in Figure 2.3(d)-(f). Each simulation lasts for 40 units of time with time step size of 0.05 (i.e., 800 steps in total). Prices (p_1, p_2) are measured by the right scale; inventory levels (X_1, X_2), factors (k_1, k_2), inflow rate (Λ), production rate (Q), and outflow rate (S) are measured by the left scale. The long-run average rates of the three flows should be the same, but for illustrative purposes, we shift them slightly apart from each other.

A positive demand shock in Figure 2.3(a) translates to instantaneous increases in both finished goods price and raw material price, as asserted in Theorem 2.3(i). The price spread also increases, resulting in an instantaneous increase in the production rate. The increase in sales is greater than in production, which, in turn, is greater than the increase in procurement. Thus, the effect of the demand shock is dampened by the inventories. The inventories drop to the lowest levels some time after the demand shock occurs, and slowly revert back to the original levels. Thus, at time 15 or so, the prices stay above the original level even though the demand shock has already died out. As a result, the sales rate drops below the original level before reverting back to its original level. Figure 2.3(b) shows that a negative supply shock affects inventories and prices in the same directions as those in (a), but the rates of flows decrease. Similar to Figure 2.3(a), the effect of the supply shock is partly absorbed by the inventories.

Figure 2.3(c) combines the effects in (a) and (b). As the effects of k_1 and k_2 are additive (because k_1 and k_2 are independent, and the drifts of the equilibrium processes are linear in $\mathbf{p}$ and $\mathbf{k}$), Figure 2.3(c) actually shows the responses in (a) minus that in (b). When a supply shock and a demand shock of the same size occur simultaneously, the raw material price and the finished goods price are negatively correlated, as are the inventory levels.

In Figure 2.3(d), we simulate the equilibrium driven by a particular sample path of $\mathbf{k}_t$ where k_{2t} is a typical mean-reverting path while k_{1t} is constantly zero. In Figure 2.3(e), k_{1t} is chosen to be another independent mean-reverting path while k_{2t} is set to zero. Figure 2.3(f) shows a more realistic scenario where both demand and supply fluctuations exist. It actually shows the responses

Figure 2.3: Simulation example 1



in (d) plus those in (e), as the effects of k_1 and k_2 are additive. The factors are not shown in Figure 2.3(f) to allow for a clear illustration.

Since inventory levels and prices move in opposite directions under a supply shock and/or a demand shock, we expect to see them in negative correlation under fluctuating supply and demand. Figure 2.3(d)-(f) generally confirm this for the whole simulation period. Nevertheless, inventory fluctuations are lagged behind, and consequently, the shorter-term correlation between inventory levels and prices is weaker. These facts also are exhibited for the petroleum industry data in the next section.

In Figure 2.3(f), production is seen to be less variable than procurement and sales, because of the production smoothing role of the inventories. This fact can also be observed in the actual data in the next section. ∎

Example 2.2 Consider an alternative operating cost function,

$$G(Q,\mathbf{X}) + H(\mathbf{X}) = a_0 + 0.2Q + 0.1Q^2 - 10X_1 - 13.2X_2 + 0.04X_1^2 + 0.07X_2^2 + 0.1X_1X_2.$$

Suppose the demand and supply functions and the discount rate are the same as Example 1. Then, the equilibrium is given by Theorem 2.2 with parameters:

$$\mathbf{B} = \begin{bmatrix} -0.009 & -0.246 \\ 0.246 & -0.550 \end{bmatrix}, \quad \mathbf{D} = \begin{bmatrix} -0.059 & 0.072 \\ -0.072 & 0.098 \end{bmatrix}, \quad \mathbf{m} = \begin{bmatrix} 11.71 \\ 16.29 \end{bmatrix}, \quad \mathbf{m}_X = \begin{bmatrix} 66.19 \\ 41.19 \end{bmatrix}.$$

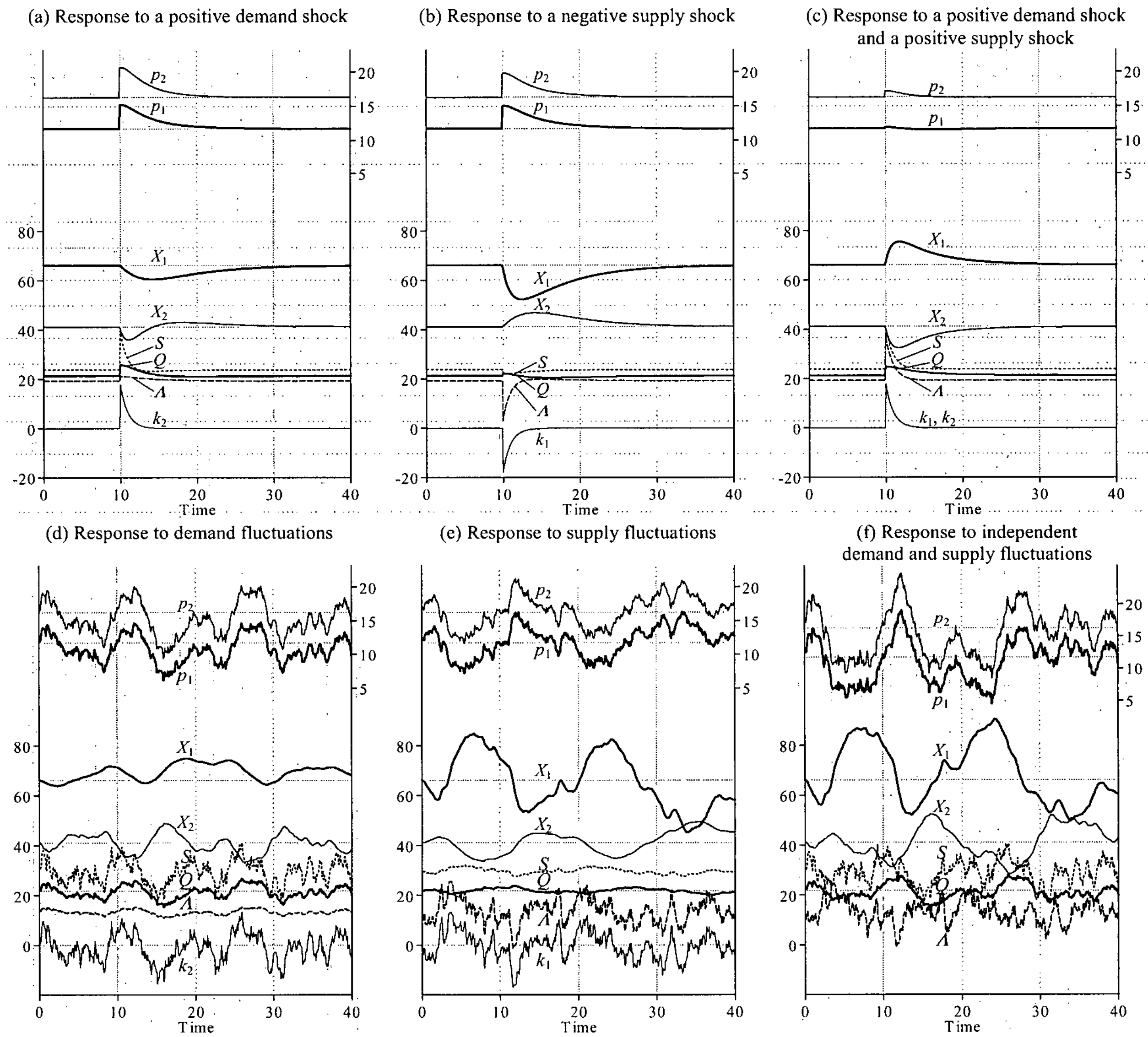
Figure 2.4(a) and (b) show that the impulse responses of the prices and the raw material inventory are similar to Example 1, but the finished goods inventory and production behave rather differently. In Figure 2.4(a), the finished goods inventory first drops, then picks up even above the original level, before going back to the original level. The production increases first, but then drops slightly below its original level (by only 0.9) before going back to the original level (21.86).

In Figure 2.4(b), differing from Example 1, the production rate increases slightly even with a negative supply shock, and then drops below the original level before going back to the original level. These changes are relatively small compared to Example 1. Thus, the raw material inventory drops, while the finished goods inventory builds up, and then both revert back to the original levels.

These patterns of responses are due to the fact that the finished goods inventory and its interaction with the raw material inventory contribute to a larger degree to the nonlinear part of the operating cost ($0.07X_2^2 + 0.1X_1X_2$), in contrast to the contribution by the raw material inventory ($0.04X_1^2$). When the demand surges, the optimal response is to produce faster to compensate for the drop in the finished goods, and even "over" produce to raise the finished goods inventory slightly above the original level, in contrast to Example 1. When the supply drops, lowering the production rate is undesirable since maintaining both types of inventories at low levels is more expensive.

These differences from Example 1 are also shown in Figure 2.4(d) and (e). Over the whole period, the raw material inventory is generally negatively correlated with prices, while the finished

Figure 2.4: Simulation example 2



goods inventory is negatively correlated with prices in (d) but positively correlated with prices in (e). Inventory fluctuations are lagged behind to various degrees. Figure 2.4(f) shows that the composite effects under both demand and supply fluctuations are more mixed and the relationship between inventory and price (especially for the finished goods) cannot be described as a more negative or a more positive correlation.

In any case, production is less variable than procurement and sales, a fact that is observed in the actual data in the next section. ∎

We have also produced other examples with different inventory and price behaviors. For instance, demand shocks can be translated to positive correlations between raw material price and lagged raw material inventory, while supply shocks can be translated to positive correlations between finished goods price and lagged finished goods inventory. The magnitudes of the responses and the lag effects vary under different settings.

Since rich inventory and price behavior patterns can be generated from our simple model, the equilibrium framework in Section 2.3 may be able to explain the real inventory and price fluctuations.

The results of the simulations have two important implications for the study of commodity inventory and price behavior:

- The study of raw material markets should be integrated with the study of the finished goods markets. Looking at only one commodity, as is usually the approach in the literature, is not sufficient to identify the inventory and price dynamics.
- The inventory level and the price are not always negatively correlated. The lag effects further complicate the relationship between the inventory level and the price. To analyze the mixed effect, a dynamic model should be used.

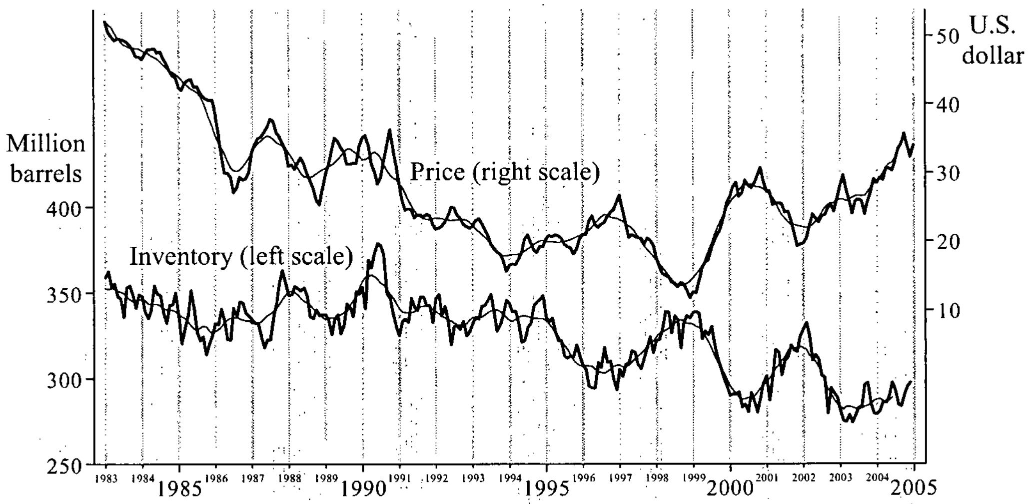
2.6 Empirical Evidence from the Petroleum Refining Industry

In this section, we use the actual data of U.S. petroleum inventories and prices to fit the equilibrium model and study its implications. We test the special economy model in Section 2.4. Statistical estimation suggests that the special economy model explains some of the variations in the actual data. All the results are meaningful, but should be interpreted with caution, as our model in Section 2.4 is a simple approximation of the real price and inventory behavior. The complete econometric estimation of the general model in Section 2.3 will be explored in future research.

Let p_1 be the petroleum refineries' acquisition cost for crude oil. The refineries maintain crude oil inventory and petroleum products inventories. We do not break down inventories into many product categories, since our model considers only one finished product. We define p_2 as a composite price of refineries' wholesale prices of petroleum products. Specifically, p_2 is the weighted average of the wholesale prices of gasoline, distillates, jet fuel, propane, and residual oil. The weights are determined by the amount of supply. The data for inventories, acquisition cost, petroleum products wholesale prices and supplies are all from the Monthly Energy Review of the Energy Information Administration. The refineries' wholesale price data is available since 1982, while the other data is available since 1973. As a result, our period of study is from January 1982 to January 2005 (277 months).

Figure 2.5: Petroleum price and inventory

(a) Adjusted crude oil price (refinery acquisition cost) and inventory (without SPR)



(b) Adjusted composite petroleum product wholesale price and inventory

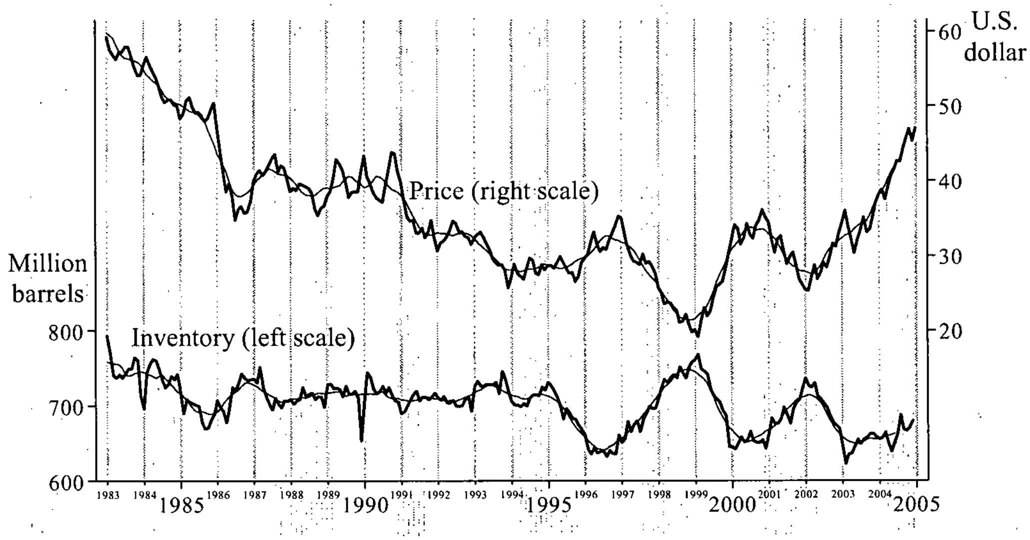
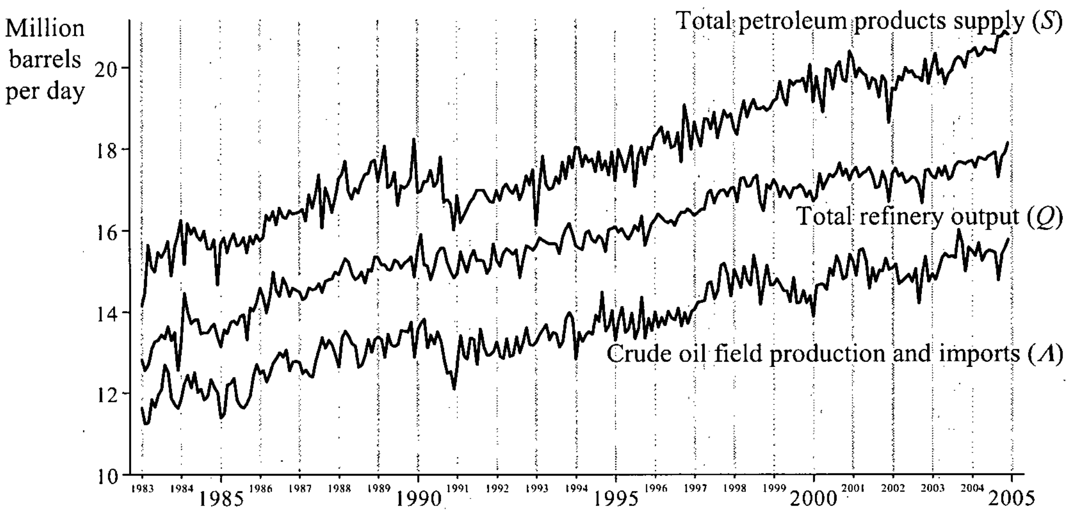


Figure 2.5 shows the seasonally adjusted petroleum price and inventory. (The seasonal adjust-

Figure 2.6: Crude oil procurement, refinery production, and petroleum products supply (seasonally adjusted)



ment is discussed below.) In terms of the long-term trends, inventories and prices seem negatively correlated from about 1994 to 2003. The short term variations are more complicated: inventory levels and prices are often seen to be in negative correlation, but lag effects and positive correlation also appear occasionally.

Figure 2.6 shows the seasonally adjusted rates of inflow, production, and outflow. The refinery output rate is higher than the crude oil procurement rate because some other liquids are needed in the refining process and there also exist processing gains in the refining process. The petroleum products supply includes imports, hence is higher than the refinery output. It can be seen that refinery production is less variable than crude oil procurement and products supply. This is what has been observed in the simulations as well.

We assume that firms discount their profit using the risk-free rate r_t, which is the 3-month treasury bill rate. We estimate the special economy as specified in (2.40). The market-clearing conditions are identities for the data, and are thus dropped from estimation. The system that we estimate is:

$$\begin{cases} \mu_{1t} - r_t p_{1t} &= -c_1 + c_{11} X_{1t} + c_{12} X_{2t} + \epsilon_{1t}, \\ \mu_{2t} - r_t p_{2t} &= -c_2 + c_{12} X_{1t} + c_{22} X_{2t} + \epsilon_{2t}, \\ p_{2t} - p_{1t} &= a_1 + a_2 Q_t + \epsilon_{3t}, \end{cases} \tag{2.45}$$

where $\boldsymbol{\mu}_t = \mathbf{B}(\mathbf{p}_t - \mathbf{m}) + \mathbf{D}\mathbf{k}_t$.

The special economy model assumes that the supply and demand levels revert to constant means, that the operating cost function is stationary over time, and that the discount rate is constant. To fit such a model with stationary parameters, we need to identify periods in which all processes appear to be stationary. But then, we do not have enough data for each period.

Our approach is to make the data series stationary by removing the trends, and then fit the model with the detrended data. In theory, we will need a non-stationary model to support this approach, but this approach is intuitively appealing, and is similar to that of Ye et al. (2002), who used the deviation of actual inventory from its normal level to forecast the oil price.

Two components of the trend are present: a seasonal trend and a long-term trend. We first remove the seasonal component using the X-12-ARIMA program available from the U.S. Census Bureau (http://www.census.gov/srd/www/x12a). The seasonally adjusted prices are shown in Figure 2.5. Then, for every month t, we compute the average price from month $t-6$ to $t+5$, and adjust it for sharp price changes. The series of these average prices forms the long-term trend, shown as thin curves in Figure 2.5. The difference between the seasonally adjusted price and the long-term trend is the detrended price process. The same detrending procedure is applied to the inventory, production, crude oil supply, and total consumption data.

The discrete version of the price diffusion process in (2.38) is a vector autoregressive of order one (VAR(1)) process:

$$\mathbf{p}_{t+1} = (\mathbf{B}+\mathbf{I})\mathbf{p}_t + \mathbf{D}\mathbf{k}_t - \mathbf{B}\mathbf{m} + \boldsymbol{\epsilon}_t, \tag{2.46}$$

where $\boldsymbol{\epsilon}_t$ has certain correlation structure. Many statistical software packages provide VAR estimation. Our estimation for the system of equations in (2.46) is provided in Table 2.1.

Table 2.1: Estimation of equations (2.46) using petroleum data

$$\mathbf{B} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} -0.214\ (-2.39)^{***} & -0.059\ (-0.67) \\ 0.332\ (3.44)^{***} & -0.616\ (-6.44)^{***} \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{bmatrix} = \begin{bmatrix} -0.315\times 10^{-3}\ (-1.64) & 0.353\times 10^{-3}\ (1.93)^{*} \\ -0.278\times 10^{-3}\ (-1.34) & 0.411\times 10^{-3}\ (2.08)^{**} \end{bmatrix}$$

R^2 is 0.54 and 0.47 for the two equations in (2.46), respectively.

*, ** and *** denote statistical significance at 10%, 5% and 1% level, respectively.

t statistics are reported in brackets.

Note that the estimated **B** satisfies condition (2.39), and the estimated **D** have the right signs as specified in Theorem 2.3. Using these estimates, we compute the price drift in (2.45), and then estimate the whole system (2.45). We apply the three-stage least squares (3SLS) regression with the constraint that the coefficient of X_{2t} in the first equation of (2.45) must equal the coefficient of X_{1t} in the second equation, as the operating cost coefficient matrix **C** is symmetric. The results are shown in Table 2.2. Although R^2's are small, since the model is simple, the data supports the simple model in the sense that all coefficients have the right signs and most of them are statistically significantly different from zero, and the estimators satisfy the convex operating cost assumption.

Table 2.2: Estimation of equations (2.45) using petroleum data

Equation (2.45-1)	
Operating cost coefficient c_1	8.61(6.00)***
Operating cost coefficient c_{11}	10.64(5.38)***
Operating cost coefficient c_{12}	7.13(4.63)***
R^2	0.134
Equation (2.45-2)	
Operating cost coefficient c_2	8.77(4.85)***
Operating cost coefficient c_{22}	8.94(4.41)***
Operating cost coefficient c_{12}	7.13(4.63)***
R^2	0.102
Equation (2.45-3)	
Production cost coefficient a_1	2.19(0.97)
Production cost coefficient a_2	356.05(2.17)**
R^2	0.031

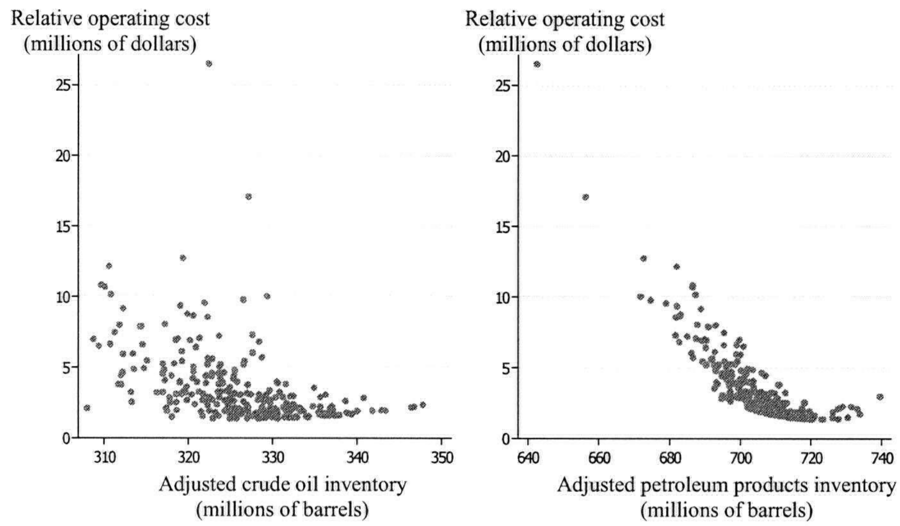
*, ** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.
t-statistics are reported in brackets.

An interesting question is whether or not a significant convenience yield is present for holding inventories, as advocated in the commodity markets literature. To study the convenience yield implied by our estimation, we plot the estimated industry operating cost in Figure 2.7.

Since the coefficient a_0 in the operating cost is not estimated, only the relative level of operating cost is shown. Figure 2.7 suggests that operating cost increases significantly when the inventory levels drop. This result is consistence with the significant convenience yield identified by Pindyck (1994) and Considine (1997) among others.

We also tried other specifications of the operating cost, such as linear holding cost and other

Figure 2.7: Estimated relative operating cost



operating costs modeled as an inverse function of inventory levels. The results generally confirm our findings.

2.7 Conclusion and Extensions

This chapter provides a rational expectations equilibrium model for storable goods. Compared with other equilibrium models in the literature, our model provides more firm level operational insights and equilibrium properties that facilitate the analysis of the price and the inventory dynamics in the economy. We demonstrated the equilibrium inventory and price dynamics through a special economy and its simulation. The price and inventory behaviors suggest that our model has a potential to explain the real price and inventory behavior. The data from the petroleum refining industry provides empirical support for our model.

We have seen that an individual firm's optimal control policy is of a bang-bang type. This is a consequence of the firm's risk-neutrality and zero adjustment cost. An extension to the risk-averse firm is straightforward. Assuming that the firm's utility function is an increasing concave function of its profit rate, then the utility-maximizing firm will control inventories such that the marginal utility rate equals the marginal long-run expected utility. This is also an important condition under the rational expectations equilibrium of the economy with risk-averse firms. We are able to extend the conditions (2.28) to include a new term that adjusts for risk-aversion. However, finding a rational expectations equilibrium is likely to be more difficult and the subject for future research.

One could also extend the model to include the operations adjustment cost. In this approach, both the state space and the control space must be redefined. The control becomes the speed of adjustment, and the state space needs to be expanded to include the rates of inflow, production and outflow. Another extension is to consider a multi-product model, as studied by Considine (1997). Firms may alter the timing of production relative to sales to take the advantage of cost complementarities. The appealing empirical question would be to test all of the above extended models and perhaps their combination, to see which one best fits the data.

References

Akella, R., P. R. Kumar. 1986. Optimal control of production rate in a failure prone manufacturing system. *IEEE Transactions on Automatic Control*, **31**(2) 116-126.

Bielecki, T., P. R. Kumar. 1988. Optimality of zero-inventory policies for unreliable manufacturing systems. *Operations Research*, **36**(4) 532-541.

Brennan, M. J. 1958. The supply of storage. *American Economic Review*, **47**(1) 50-72.

Cafiero, C., B. D. Wright. 2003. Is the storage model a 'closed' empirical issue? The empirical ability of the storage model to explain price dynamics. *Symposium on the State of Agricultural Commodity Market Research*, FAO Commodity and Trade Division, Rome, Italy.

Considine, T. J. 1997. Inventories under joint production: an empirical analysis of petroleum refining. *Review of Economics and Statistics*, **79**(3) 493-502.

Considine, T. J., D. F. Larson. 2001. Uncertainty and the convenience yield in crude oil price backwardations. *Energy Economics*, **23** 533-548.

Deaton, A., G. Laroque. 1992. On the behavior of commodity prices. *Review of Economic Studies*, **59**(1) 1-23.

Deaton, A., G. Laroque. 1996. Competitive storage and commodity price dynamics. *Journal of Political Economy*, **104**(5) 896-923.

Fleming, W. H., H. M. Soner. 1993. *Controlled Markov Processes and Viscosity Solutions.* Springer-Verlag, New York.

Fleming, W. H., S.P. Sethi, H. M. Soner. 1987. An optimal stochastic production planning problem with randomly fluctuating demand. *SIAM Journal on Control and Optimization*, **25**(6) 1494-1502.

Gavirneni, S. 2004. Periodic review inventory control with fluctuating purchasing costs. *Operations Research Letters*, **32** 374-379.

Harrison J. M., T. L. Sellke and A. J. Taylor. 1983. Impulse control of Brownian motion. *Mathematics of Operations Research.* **8**(3) 454-466.

Harrison J. M., M. I. Taksar. 1983. Instantaneous control of Brownian motion. *Mathematics of Operations Research*, **8**(3) 439-453.

Horn, R. A., C. R. Johnson. 1991. *Topics in Matrix Analysis.* Cambridge University Press, Cambridge, New York.

Karatzas, I. 1980. Optimal discounted linear control of the Wiener process. *Journal of Optimization Theory and Applications*, **31**(3) 431-440.

Muth, J. F. 1961. Rational expectations and the theory of price movements. *Econometrica*, **29**(3) 315-335.

Pindyck, R. S. 1994. Inventories and the short-run dynamics of commodity prices. *RAND Journal of Economics*, **25**(1) 141-159.

Routledge, B. R., D. J. Seppi, C. S. Spatt. 2000. Equilibrium forward curves for commodities. *Journal of Finance*, **55**(3) 1297-1338.

Sethi, S.P., H.M. Soner, Q. Zhang, J. Jiang. 1992. Turnpike sets and their analysis in stochastic production planning problems. *Mathematics of Operations Research*, **17**(4) 932-950.

Sapra, A., P. L. Jackson. 2005. *The martingale evolution of price forecasts in a supply chain market for capacity.* Working paper, Cornell University.

Scheller-Wolf A., S. Tayur. 1998. Reducing international risk through quantity contracts. GSIA working paper, Carnegie Mellon University.

Sethi, S. P., H. Yan, H. Zhang, Q. Zhang. 2002. Optimal and hierarchical controls in dynamic stochastic manufacturing systems: a survey. *Manufacturing and Service Operations Management*, **4**(2) 133-170.

Tayur, S., W. Yang. 2002. Equilibrium analysis of a natural gas supply chain. Working paper, Carnegie Mellon University.

Williams, J. C. and B. D. Wright. 1991. *Storage and commodity markets.* Cambridge University Press, Cambridge, New York.

Ye, M., J. Zyren, J. Shore. 2002. Forecasting crude oil spot price using OECD petroleum inventory levels. *International Advances in Economic Research*, **8**(4) 324-334.

Zyren, J. 1995. What drives motor gasoline prices? *Petroleum Marketing Monthly*, June 1995, Energy Information Administration.

Chapter 3

Optimal Pricing and Replenishment in a Single-Product Inventory System with Brownian Demand

3.1 Introduction

In recent years, substantially more research has been done on joint inventory and pricing strategies. The focus has been on establishing the optimality of structural inventory and pricing policies. One of the key determinants of the complexity and generality of these models is the assumption about the demand processes. Various demand models have been proposed for periodic-review settings (e.g., additive demand, multiplicative demand, and other general demand models), and for continuous-review settings (e.g., Poisson and renewal demand models).

In this chapter, the basic setting is a single-product continuous-review inventory system with a *Brownian demand model.* Brownian motion, with its continuous path, is particularly appropriate for modeling the demands of fast-moving products, such as sugar and coffee. It is a natural model when demand forecast involves Gaussian noises. The drift of the Brownian demand (i.e., the instantaneous average demand rate) depends on the price set by the firm. The diffusion term represents the demand variability.

We assume that inventory replenishment is instantaneous, which is appropriate when the lead time is insignificant relative to the length of the replenishment cycle. Thus, inventory is depleted by the demands in a continuous-time fashion and then replenished immediately when it drops to the reorder point. We assume, in our base model, that the demands must be satisfied immediately upon arrival. Consequently, the reorder point is zero, and the replenishment follows a simple order-up-to policy, with the order-up-to level denoted by S. We allow the pricing decisions to be dynamically adjusted within the replenishment cycle, but we do not adjust prices continuously over time, as most realistic pricing strategies do not involve too many different prices. Instead, we explicitly consider the number of prices to use. We divide S into N equal segments, N being a given integer. We charge one price when the inventory level falls into each segment. All N prices are optimally determined, jointly with the replenishment level S, so as to maximize the expected long-run average or the discounted profit.

The rest of the introductory section of this chapter contains a literature review and a summary of the features of the Brownian demand model.

3.1.1 Literature Review

There has been substantial literature on joint pricing and inventory control. We refer the reader to three excellent survey papers: Yano and Gilbert (2003), Elmaghraby and Keskinocak (2003), and Chan et al. (2004). Our review below focuses on those papers which are closely related to our study.

Whitin (1955), Porteus (1985a), Rajan, Rakesh and Steinberg (1992), among others, study demands that are deterministic functions of price. Whitin (1955) connects pricing and inventory control in the EOQ framework, and Porteus (1985a) provides an explicit solution for the linear demand instance. Rajan, Rakesh and Steinberg (1992) investigate continuous pricing for perishable products for which demands may diminish as products age.

In periodic-review stochastic inventory and pricing models, several types of demand assumptions are made. The demand function in each period typically consists of a deterministic demand function and a random component. These two components can be additive (e.g., Chen and Simchi-Levi 2004a, Chen, Ray and Song 2006), or multiplicative (e.g., Song, Ray and Boyaci 2006), or both (e.g., Chen and Simchi-Levi 2004a,b). Some other models allow the random component to affect the demand in more general forms (e.g., Federgruen and Heching 1999, Polatoglu and Sahin 2000, Feng and Chen 2004, Huh and Janakiraman 2005).

Federgruen and Heching (1999) assume that the demand in each period depends on the price charged in the period and a random term. The dependence can be quite general, but every realization of demand function is assumed to be concave in price. The replenishment cost is linear, without a fixed setup cost. The authors show that a base-stock list-price policy is optimal for both average and discounted objectives. Earlier related works include those by Zabel (1972) and Thowsen (1975).

The above periodic-review models have been extended to include a replenishment setup cost. In the backordering setting, it was first conjectured by Thomas (1974), and then proved by Chen and Simchi-Levi (2004a), that an (s, S, p) policy is optimal for additive demand in a finite horizon. Chen and Simchi-Levi (2004b) further prove the optimality of the stationary (s, S, p) policy for both additive and multiplicative demand in an infinite horizon. Feng and Chen (2004) prove the optimality of (s, S, p) policy under more general demand functions, but restricting the prices to a finite set. Assuming lost sales, Polatoglu and Sahin (2000) obtain rather involved optimal policies under a general demand model and provide restrictive conditions under which the (s, S, p) policy is optimal. Chen, Ray and Song (2006) prove that the (s, S, p) policy is optimal under additive demand and lost sales. Huh and Janakiraman (2005) provide a novel approach for proving and generalizing many of the early results for both backorder and lost sales settings.

Continuous-review models are studied by Feng and Chen (2003), and Chen and Simchi-Levi (2006). Feng and Chen (2003) model the demand as a Poisson process with price-sensitive intensity. Pricing and replenishment decisions are made upon finishing serving each demand, but the prices are restricted to a given finite set. An (s, S, p) policy is proven to be optimal. Chen and Simchi-Levi (2006) generalize this model by considering a compound renewal demand process with both the inter-arrival times and the size of the demand depending on the price. It is shown that the (s, S, p) policy is still optimal.

There are also works that involve production, for instance, Li (1988) considers a make-to-order production system with price-sensitive demand. Both production and demand are modeled by Poisson processes with controllable intensities, and the control of demand intensity is through pricing. A barrier policy is shown to be optimal: when the inventory level reaches an upper barrier, production stops; when the inventory level drops to zero, the demand stops (or the demand is lost). The optimal price is shown to be a non-increasing function of the inventory level.

3.1.2 Features of the Brownian Demand Model

First, the Brownian demand model allows us to explicitly and naturally bring out the impact of demand variability on the optimal pricing and replenishment decisions, whereas results along this line were previously limited to only a few numerical studies. Federgruen and Heching (1999) experimented with a multiplicative demand case, and found that the optimal base stock and the optimal prices are increasing in demand variability. Polatoglu and Sahin (2000) numerically examined an additive demand model with uniform random noise, and found that the order-up-to and reorder levels both tend to be higher when the demand variability increases, while the prices do not exhibit a clear monotone property. To the best of our knowledge, our study is the first analytical examination of the impact of demand variability.

Second, with the Brownian demand model and the zero lead time assumption, the order-up-to replenishment policy is immediately seen as being optimal. This allows us to focus on how, given the order-up-to policy, the optimal order quantity and the optimal prices vary and interact with other model parameters, particularly the demand variability. This is very different from most of the literature, where the focus is on establishing the optimality of the order-up-to policy.

Third, using the Brownian demand model, we can derive an upper bound for the profit improvement generated from the use of the dynamic pricing strategy in comparison to a static strategy. We find that dynamic pricing results in only a limited profit improvement over the single price strategy (when both are optimally determined). The relative profit improvement, however, becomes more significant when the profit margin is low. This result is consistent with the numerical results in Feng and Chen (2004), and Chen, Ray and Song (2006).

For most part of this chapter, we focus on the Brownian demand model in (3.1) below, which treats the demand variability as a constant. This is a continuous-time analogy to the additive demand model in periodic-review setting. The key results in this chapter extend readily to the case with price-dependent demand variability, which includes a special case that is analogous to the multiplicative demand (see Section 3.4.3).

The constant demand variability model serves as an important base case for two reasons. First, the noise part in the demand model is usually due to factors other than price, such as forecast error and inventory accounting error; the former mainly depends on the information available, whereas the latter relies heavily on the prevailing technology. (For instance, the radio frequency identification (RFID) technology has demonstrated great advantages in improving inventory accounting.) Second, in some empirical studies of demand price relations, the residual of the regression is usually found to be a sequence of normally distributed random variables independent of the price. Refer to, for example, Reiss and Wolak (2006), and Genesove and Mullin (1998). The latter studies the demand functions for the U.S. sugar industry.

The rest of the chapter is organized as follows. In Section 3.2, we present a formal description of our model. In Section 3.3, we study the optimal pricing and replenishment decisions under the long-run average objective. We start by making these decisions separately, so as to highlight the comparisons with prior studies and known results, and then present the joint optimization model and demonstrate the profit improvement. We conclude this chapter by pointing out several possible extensions in Section 3.4.

3.2 Model Description and Preliminary Results

We consider a continuous-review inventory model with a price-sensitive demand. The objective is to determine the inventory replenishment and pricing decisions that strike a balance between the sales revenue and the cost for holding and replenishing inventory over time, so as to maximize the expected long-run average or discounted profit.

3.2.1 The Demand Model

The cumulative demand up to time t is denoted as $D(t)$, and modeled by a diffusion process:

$$D(t) \quad = \quad \int_0^t \lambda(p_u)du + \sigma B(t), \qquad t \geq 0, \tag{3.1}$$

where p_t is the price charged at time t, $\lambda(p_t)$ is the demand rate at time t, $B(t)$ denotes the standard Brownian motion, and σ is a positive constant measuring the variability of the demand (or the error of demand forecast).

Let $\mathcal{P}$ and $\mathcal{L}$ denote, respectively, the domain and the range of the demand rate function $\lambda(\cdot)$. Both are assumed to be intervals of $\Re_+$ (the set of nonnegative real numbers).

Assumption 3.1 (on demand rate) *The demand rate $\lambda(p)$ and its inverse $p(\lambda)$ are both strictly decreasing and twice continuously differentiable in the interior of $\mathcal{P}$ and $\mathcal{L}$, respectively. The revenue* *rate* $r(\lambda) = p(\lambda)\lambda$ *is strictly concave in* λ.

Many commonly-used demand functions satisfy the above assumption, including the following examples, where the parameters α, β and δ are all positive:

- The linear demand function $\lambda = \alpha - \beta p$, $p \in [0, \alpha/\beta]$: $r(\lambda) = \frac{\alpha}{\beta}\lambda - \frac{1}{\beta}\lambda^2$ is strictly concave;
- The exponential demand function $\lambda = \alpha e^{-\beta p}$, $p \geq 0$: $r(\lambda) = -\frac{1}{\beta}\lambda \log(\lambda/\alpha)$ is strictly concave;
- The power demand function $\lambda = \beta p^{-\delta}$, $p \geq 0$: $r(\lambda) = \lambda^{-\frac{1}{\delta}+1}\beta^{\frac{1}{\delta}}$ is strictly concave if $\delta \geq 1$.

3.2.2 Cost Parameters

We assume that the holding cost is linear in the quantity held. Let h be the cost for holding one unit of inventory per unit of time. The replenishment cost is a function of the replenishment quantity, denoted as $c(S)$, which satisfies the following assumption.

Assumption 3.2 (on replenishment cost) *The replenishment cost function* $c(S)$ *is twice continuously differentiable and increasing in* S *for* $S \in (0, \infty)$. *The average cost* $a(S) = c(S)/S$ *is strictly convex in* S, *and* $a(S) \to \infty$, *as* $S \to 0$.

Consider a special case: $c(S) = K + cS^{\delta}$, $S > 0$, where $K, c, \delta > 0$. When $\delta = 1$, this is the most commonly used linear function with a setup cost K. The average cost function, $a(S) = \frac{K}{S} + cS^{\delta-1}$ is convex if $\delta \in (0, 1] \cup [2, \infty]$. If $\delta \in (1, 2)$, $a(S)$ is convex in the region where $a'(S) \leq 0$. (This region is of particular interest; refer to Section 3.3.1.) To see this, note that

$$a'(S) = \frac{1}{S^2}(-K + c(\delta - 1)S^{\delta}),$$

and when $\delta \in (1, 2)$ and $a'(S) \leq 0$, we have

$$a''(S) = -\frac{2-\delta}{S^3}\Big(- \frac{2}{2-\delta}K + c(\delta - 1)S^{\delta}\Big) \geq -\frac{a'(S)(2-\delta)}{S} \geq 0.$$

3.2.3 Pricing and Replenishment Policies

Assume replenishment is instantaneous, i.e., zero lead time. We further assume that all demands will be supplied immediately upon arrival; i.e., no backorder is allowed, or there is an infinite backorder cost penalty. (Our results extend readily to the backorder case; refer to Section 3.4.2.)

The replenishment follows a continuous-review order-up-to policy. With the Brownian demand model and the zero lead time assumption, the order-up-to replenishment policy is immediately seen as being optimal. Specifically, whenever the inventory level drops to zero, it is brought up to S instantaneously via a replenishment, where S is a decision variable. We shall refer to the time between two consecutive replenishment epochs as a *cycle*.

We adopt the following dynamic pricing strategy. Let $N \geq 1$ be a given integer, and let $S = S_0 > S_1 > \cdots > S_{N-1} > S_N = 0$. Immediately after a replenishment at the beginning of a cycle, price p_1 is charged until the inventory drops to S_1; price p_2 is then charged until the inventory drops to S_2; ...; and finally when the inventory level drops to S_{N-1}, price p_N is charged until the inventory drops to $S_N = 0$, when another cycle begins. The same pricing strategy applies to all cycles. For simplicity, we set $S_n = S(N-n)/N$. That is, we divide the full inventory of S units into N equal segments, and price each segment with a different price as the inventory is depleted by demand.

In general, the segments are not necessarily equal, but our experiments show that the optimal segments (optimized jointly with prices and replenishment level) are quite even and the corresponding profit is almost the same as using equal segments. Furthermore, in practical situations when the prices are within a discrete set, our algorithm based on equal segments can effectively tell how many prices to use, when to change prices and what prices to change to. (See Example 3.12 for elaboration.)

In summary, the decision variables are: $(S, \mathbf{p})$, where $S \in \Re_+$, and $\mathbf{p} = (p_1, \ldots, p_N) \in \mathcal{P}^N$. Within a cycle, we shall refer to the time when the price p_n is applied as period n.

3.2.4 The Inventory Process

Without loss of generality, suppose at time zero the inventory is filled up to S. We focus on the first cycle which ends at the time when inventory reaches zero. Let $T_0 = 0$, and let T_n be the first time when inventory drops to S_n:

$$T_n \quad := \quad \inf\left\{\, t \geq 0 \;:\; D(t) = nS/N \,\right\}, \qquad n = 1, 2, \ldots, N.$$

The length of period n is therefore $\tau_n := T_n - T_{n-1}$. Since T_n's are stopping times, by the strong Markov property of Brownian motion, τ_n is just the time during which S/N units of demand has occurred under the price p_n. That is,

$$\tau_n \quad \stackrel{dist.}{=} \quad \inf\left\{\, t \geq 0 \;:\; \lambda(p_n)t + \sigma B(t) = S/N \,\right\}. \tag{3.2}$$

Let $X(t)$ denote the inventory-level at t. We have,

$$X(t) \;=\; S - D(t), \qquad t \in [0, T_N).$$

Since our replenishment-pricing policy is the same for all cycles, $X(t)$ is a regenerative process with the replenishment epochs being its regenerative points.

We conclude this section with a lemma, which gives the first two moments and the generating function of the stopping time τ_n. (The proof is in the appendix.)

Lemma 3.1 *For the stopping time τ_n in (3.2), we have*

$$\begin{aligned}
\mathsf{E}[\tau_n] &= \frac{S}{N\lambda_n}, \\
\mathsf{E}[\tau_n^2] &= \frac{\sigma^2}{\lambda_n^2}\mathsf{E}[\tau_n] + \mathsf{E}^2[\tau_n] = \frac{\sigma^2 S}{N\lambda_n^3} + \frac{S^2}{N^2\lambda_n^2}, \\
\mathsf{E}[e^{-\gamma\tau_n}] &= \exp\Big[-\frac{\sqrt{\lambda_n^2 + 2\sigma^2\gamma} - \lambda_n}{\sigma^2}\frac{S}{N}\Big],
\end{aligned}$$

where $\lambda_n = \lambda(p_n) > 0$, and $\gamma > 0$ is a parameter.

3.3 Long-Run Average Objective

To optimize the long-run average profit, thanks to the regenerative structure of the inventory process, it suffices to focus on the first cycle. Recall that period n refers to the period in which the price p_n applies, and the inventory drops from $S_{n-1} = \frac{(N-n+1)S}{N}$ to $S_n = \frac{(N-n)S}{N}$. Applying integration by parts, and recognizing

$$dX(t) = -dD(t), \qquad X(T_{n-1}) = S_{n-1}, \quad X(T_n) = S_n,$$

we have

$$\begin{aligned}
\int_{T_{n-1}}^{T_n} X(t)dt &= T_n S_n - T_{n-1}S_{n-1} - \int_{T_{n-1}}^{T_n} t dX(t) \\
&= T_n S_n - T_{n-1}S_{n-1} - \int_{T_{n-1}}^{T_n} T_{n-1} dX(t) + \int_{T_{n-1}}^{T_n} (t - T_{n-1}) dD(t) \\
&= \tau_n S_n + \int_{T_{n-1}}^{T_n} (t - T_{n-1})\big[\lambda(p_n)dt + \sigma dB(t)\big].
\end{aligned}$$

A simple change of variable yields

$$\int_{T_{n-1}}^{T_n} (t - T_{n-1})\lambda(p_n)dt \;=\; \int_0^{\tau_n} u\lambda(p_n)du \;=\; \frac{1}{2}\lambda(p_n)\tau_n^2;$$

whereas

$$\mathsf{E}\Big[\int_{T_{n-1}}^{T_n} (t-T_{n-1})dB(t)\Big] = \mathsf{E}\Big[\int_{T_{n-1}}^{T_n} tdB(t)\Big] - \mathsf{E}\big[T_{n-1}\big]\mathsf{E}\big[B(T_n)-B(T_{n-1})\big] = 0,$$

follows from the martingale property of $B(t)$ and the optional stopping theorem.

Let $v_n(S,p_n)$ denote the expected profit (sales revenue minus inventory holding cost) during period n. Then, making use of the above derivation, along with Lemma 3.1, we have

$$\begin{aligned} v_n(S,p_n) &= \frac{p_n S}{N} - \mathsf{E}\Big[\int_{T_{n-1}}^{T_n} hX(t)dt\Big] \\ &= \frac{p_n S}{N} - h\mathsf{E}[\tau_n]S_n - \frac{1}{2}h\lambda(p_n)\mathsf{E}[\tau_n^2] \\ &= \frac{p_n S}{N} - \frac{hS^2(N-n+\frac{1}{2})}{N^2\lambda(p_n)} - \frac{h\sigma^2 S}{2N\lambda(p_n)^2}, \end{aligned} \tag{3.3}$$

Note in the above expression, the first term is the sales revenue from period n, the second term is the inventory holding cost attributed to the deterministic part of the demand (i.e., the drift part of the Brownian motion), and the last term is the additional holding cost incurred by demand uncertainty.

For ease of analysis, below we shall often use $\{\mu_n = \lambda(p_n)^{-1}, n=1,\ldots,N\}$ as decision variables and denote $\boldsymbol{\mu} = (\mu_1,\ldots,\mu_N) \in \mathcal{M}^N$ and $\mathcal{M} = \{\lambda^{-1} : \lambda \in \mathcal{L}\}$. Then, the long-run average objective can be written as follows:

$$V(S,\boldsymbol{\mu}) = \frac{\sum_{n=1}^N v_n(S,p_n) - c(S)}{\frac{S}{N}\sum_{n=1}^N \mu_n} = \frac{\sum_{n=1}^N \Big[p(\frac{1}{\mu_n}) - \frac{hS}{N}(N-n+\frac{1}{2})\mu_n - \frac{h\sigma^2}{2}\mu_n^2 - a(S)\Big]}{\sum_{n=1}^N \mu_n}. \tag{3.4}$$

The additional holding cost rate due to demand uncertainty is represented by $\frac{h\sigma^2\sum_{n=1}^N \mu_n^2}{2\sum_{n=1}^N \mu_n}$.

For the special case when $N=1$, the price and the demand rate are both constants in a cycle (and hence constant throughout the horizon). The objective function (3.4) takes a simpler form. For comparison with some classical work, we use λ as the decision variable and denote the long-run average profit under single-price policy as $V(S,\lambda)$. Then,

$$V(S,\lambda) = \frac{v_1(S,p(\lambda)) - c(S)}{S/\lambda} = r(\lambda) - \frac{hS}{2} - \lambda a(S) - \frac{h\sigma^2}{2\lambda}. \tag{3.5}$$

The classical EOQ model only consists of the second and third terms in (3.5), which are the average inventory cost if the demand is deterministic with a constant rate. Whitin (1955) and Porteus (1985a) considered the price-sensitive EOQ model, which involves the first three terms in (3.5). Our model gives rise to an additional holding cost due to demand variability.

3.3.1 Optimal Replenishment with Fixed Prices

In this case, the set of prices $\mathbf{p}$, or equivalently, $\boldsymbol{\mu}$ is given, and the firm's problem is

$$\max_{S>0} \quad V(S, \boldsymbol{\mu}).$$

Under Assumption 3.1, $V(S, \boldsymbol{\mu})$ is strictly concave in S. Note that $V(S, \boldsymbol{\mu}) \to -\infty$ as $S \to \infty$ or $S \to 0$ (the latter is due to Assumption 3.2 that $a(S) \to \infty$ as $S \to 0$). Thus, the unique optimal replenishment level is determined by the first-order condition:

$$S^* \;=\; a'^{-1}\Big(-\frac{h}{N^2}\sum_{n=1}^{N}(N-n+\frac{1}{2})\mu_n\Big), \tag{3.6}$$

where a'^{-1} is well-defined since $a(\cdot)$ is strictly convex and $a'(\cdot)$ is strictly increasing under Assumption 3.2. In practice, the average cost as a function of quantity is usually first decreasing (due to economy of scale) and then increasing (due to capacity or other technological restrictions). However, at the optimal replenishment level, we have $a'(S^*) < 0$ for any fixed prices. This observation helps to reduce the search space when the replenishment level is optimized jointly with pricing decisions (see Section 3.3.3).

The way that the demand variability and the holding cost impact the optimal replenishment level can be readily derived from (3.6).

Proposition 3.1 *With prices fixed,*

(i) the optimal replenishment level S^* is independent of σ, and decreasing in h and in p_n (for any n);

(ii) the optimal profit is decreasing in σ and h.

Proof. Part (i) is obvious from (3.6). Part (ii) is also obvious since the objective in (3.4) is decreasing in σ and h in any case. ∎

The interpretation of this proposition is quite intuitive, except that the optimal replenishment level is independent of σ. The latter results from the fact that the demand variability impacts on the average profit function through the additional holding cost term, $h\sigma^2/(2\lambda)$, which does not affect the replenishment decision. This will not be the case under a discounted objective, as will be shown in Section 3.4.1.

Example 3.1 Consider $c(S) = K + cS$, where $K, c > 0$. The first order condition in (3.6) leads to the familiar EOQ formula:

$$S^* = \sqrt{\frac{2K\lambda_a}{h}}, \qquad \text{where} \qquad \lambda_a = \Big(\sum_{n=1}^{N} \frac{2N-2n+1}{N^2}\mu_n\Big)^{-1}. \tag{3.7}$$

In the standard EOQ model, the demand rate is taken as the average demand per time unit, which in our setting becomes

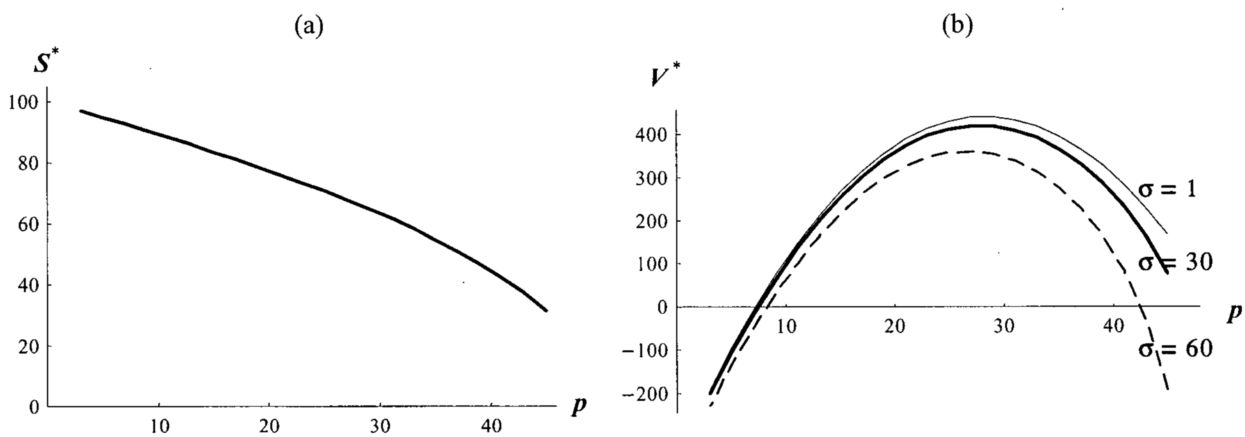
$$\bar{\lambda} \;=\; \Big(\frac{1}{N}\sum_{n=1}^{N}\mu_n\Big)^{-1}. \tag{3.8}$$

Suppose that the fixed prices satisfy $p_1 \le p_2 \le \cdots \le p_N$ (or $\mu_1 \le \mu_2 \le \cdots \le \mu_N$), i.e., a lower price is charged when the inventory level is higher (we will see this price pattern in Theorem 3.1). Higher weights are given to smaller μ_n's in (3.7), while μ_n's are equally weighted in (3.8). Hence, $\bar{\lambda} \le \lambda_a$. This implies that the standard EOQ with a time-average demand rate may lead to a replenishment level lower than the optimally desirable. ∎

Example 3.2 Let $c(S) = 100 + 5S$, $\lambda(p) = 50 - p$, $h = 1$, and $N = 1$. In Figure 3.1(a), the EOQ quantity is shown as a function of the single fixed price, $S^* = 10\sqrt{100-2p}$. Figure 3.1(b) illustrates the optimal profit as a function of the price under various levels of demand variability. The peak of each V^* curve is reached when the price and the replenishment level are jointly optimized. When the price deviates away from the optimum, the profit falls dramatically. We can also see that the optimal price decreases slightly as σ increases. This will be formally investigated in Section 3.3.3.

∎

Figure 3.1: Optimal replenishment level and profit with a fixed price



3.3.2 Optimal Pricing with a Fixed Replenishment Level

3.3.2.1 The Single-Price Problem

The firm chooses a single price, or equivalently, the demand rate, to maximize the average profit:

$$\max_{\lambda \in \mathcal{L}} \quad V(S,\lambda) = r(\lambda) - \frac{hS}{2} - \lambda a(S) - \frac{h\sigma^2}{2\lambda}.$$

Under Assumptions 3.1 and 3.2, $V(S,\lambda)$ is strictly concave in λ. Assuming an interior solution (which is rather innocuous, since the last two terms in the objective prevent λ from being extreme), the unique optimal λ follows from the first-order condition:

$$r'(\lambda) + \frac{h\sigma^2}{2\lambda^2} = a(S). \tag{3.9}$$

We have the following proposition.

Proposition 3.2 *Suppose a single price is optimized while the replenishment level is held fixed.*
(i) The optimal price p^* is decreasing in σ and h.
(ii) The optimal price p^* is increasing in $a(\cdot)$; if $a(S)$ is decreasing in S, then p^* is decreasing in S.
(iii) The optimal profit is decreasing in σ and h.

Proof. For (i) and (ii), note that $\frac{\partial^2 V}{\partial\lambda\partial\sigma} = \frac{h\sigma}{\lambda^2} > 0$, $\frac{\partial^2 V}{\partial\lambda\partial h} = \frac{\sigma^2}{2\lambda^2} > 0$, and $\frac{\partial^2 V}{\partial\lambda\partial(a(S))} = -1 < 0$. That is, the objective function V is supermodular in (λ,σ), supermodular in (λ,h) and submodular in $(\lambda, a(S))$. Hence, (i) and (ii) follow from standard results for supermodular/submodular functions (Topkis 1978). Part (iii) is obvious. ∎

Example 3.3 Let $\lambda(p) = \beta p^{-1}$, $c(S) = K + cS$, where $\beta, K, c > 0$. Then $r(\lambda) = \beta$, and the first-order condition (3.9) becomes

$$\frac{h\sigma^2}{2\lambda^2} = \frac{K}{S} + c,$$

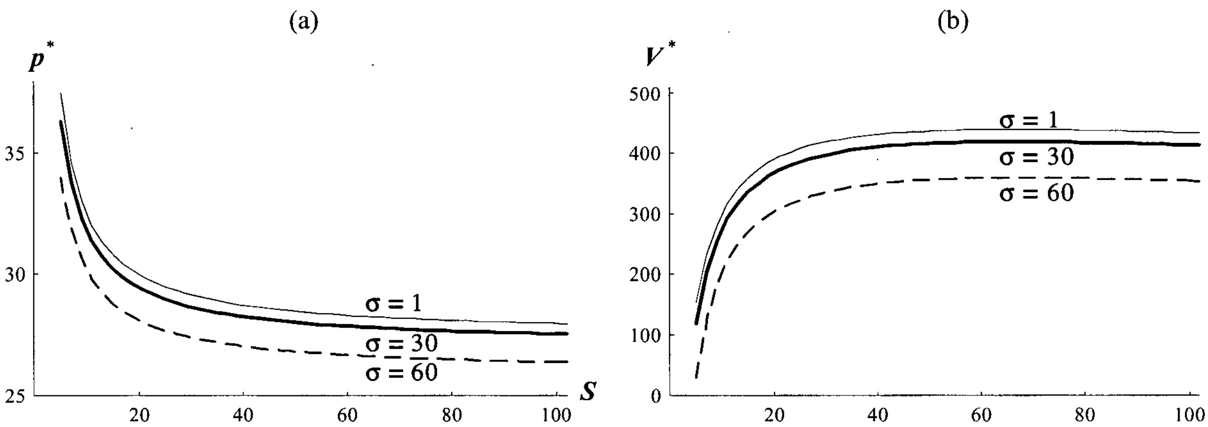
leading to the optimal price

$$p^* = \frac{\beta}{\lambda^*} = \frac{\beta}{\sigma}\sqrt{\frac{2(\frac{K}{S}+c)}{h}},$$

which is clearly decreasing in σ, h and S. ∎

Example 3.4 Consider the same setting as Example 3.2. Figure 3.2 plots the optimal price and the corresponding profit against various levels of replenishment and demand variability. The results depict the qualitative trends in Proposition 3.3.2. The peak of each V^* curve is reached when the price and inventory are jointly optimized. In contrast to Figure 3.1(b), the profit here appears less sensitive to the replenishment level, as long as the pricing decision is optimized. ∎

Figure 3.2: Optimal price and profit with a fixed replenishment level



3.3.2.2 The N-Price Problem

In this case, the decision variables are the set of prices $\mathbf{p}$, or equivalently, $\boldsymbol{\mu}$. Specifically, the problem is

$$\max_{\boldsymbol{\mu}\in\mathcal{M}^N} \qquad V(S,\boldsymbol{\mu}) = \frac{\sum_{n=1}^{N}\left[p(\frac{1}{\mu_n}) - \frac{hS}{N}(N-n+\frac{1}{2})\mu_n - \frac{h\sigma^2}{2}\mu_n^2 - a(S)\right]}{\sum_{n=1}^{N}\mu_n}. \tag{3.10}$$

The strict concavity of the revenue function $r(\lambda)$ (Assumption 3.1) implies that the function $p(\frac{1}{\mu})$ is strictly concave in μ. This is because $r''(\lambda) = 2p'(\lambda)+\lambda p''(\lambda)$ and $d^2p(\frac{1}{\mu})/d\mu^2 = \frac{1}{\mu^3}(2p'(\frac{1}{\mu})+\frac{1}{\mu}p''(\frac{1}{\mu}))$ have the same sign. Thus, the numerator of the objective is strictly concave in $\boldsymbol{\mu}$. The ratio of a concave function over a positive linear function is known to be pseudo-concave (Mangasarian 1970). Hence, $V(S,\boldsymbol{\mu})$ is pseudo-concave in $\boldsymbol{\mu}$. Furthermore, it is strictly pseudo-concave in the sense that

$$\nabla_\mu V(S,\boldsymbol{\mu}_1)(\boldsymbol{\mu}_2-\boldsymbol{\mu}_1)\le 0 \qquad \Rightarrow \qquad V(S,\boldsymbol{\mu}_2) < V(S,\boldsymbol{\mu}_1).$$

Hence, the optimal $\boldsymbol{\mu}$ must be unique. (This follows from a straightforward extension of Mangasarian (1970) from pseudo-concavity to strict pseudo-concavity.) Suppose the optimality is achieved at an interior point. Then the optimal $\boldsymbol{\mu}$ is uniquely determined by the first-order condition:

$$\frac{\partial \widetilde{v}_n(S,\mu_n)}{\partial \mu_n} = \frac{\sum_{k=1}^N \widetilde{v}_k(S,\mu_k) - c(S)}{\sum_{k=1}^N \mu_k}, \qquad n = 1,\ldots,N, \tag{3.11}$$

where $\widetilde{v}_n(S,\mu_n) = v_n(S, p(\frac{1}{\mu_n}))$. Note that the right side of the above does not depend on n. This implies that the optimal pricing must be such that the marginal profit is the same across all periods (assuming an interior optimum).

The following theorem describes the basic optimal pricing pattern.

Theorem 3.1 *For any replenishment quantity S, the optimal prices are increasing over the periods, i.e., $p_1^* \leq p_2^* \leq \cdots \leq p_N^*$.*

Proof. Since $p_n = p(\frac{1}{\mu_n})$ is increasing in μ_n, it suffices to show that the optimal μ_n^* is increasing in n. But this follows immediately from (3.10), since given any $(\mu_1,\ldots,\mu_N)$, rearranging these variables (but not changing their values) in increasing order will maximize the term $\sum_{n=1}^N n\mu_n$ in the numerator, while all other terms remain unchanged. ∎

In other words, within each cycle, the optimal prices are increasing over the periods. Intuitively, when the inventory is higher at the beginning of a cycle, we charge a lower price to induce higher demand so as to reduce the inventory holding cost.

The next proposition is about the impact of the parameters, σ, h and S on the optimal prices. (The proof of this proposition is in the Appendix.)

Proposition 3.3 *Assuming interior optimum,*

(i) The highest price p_N^* is decreasing in σ.

(ii) The lowest price p_1^* is decreasing in h.

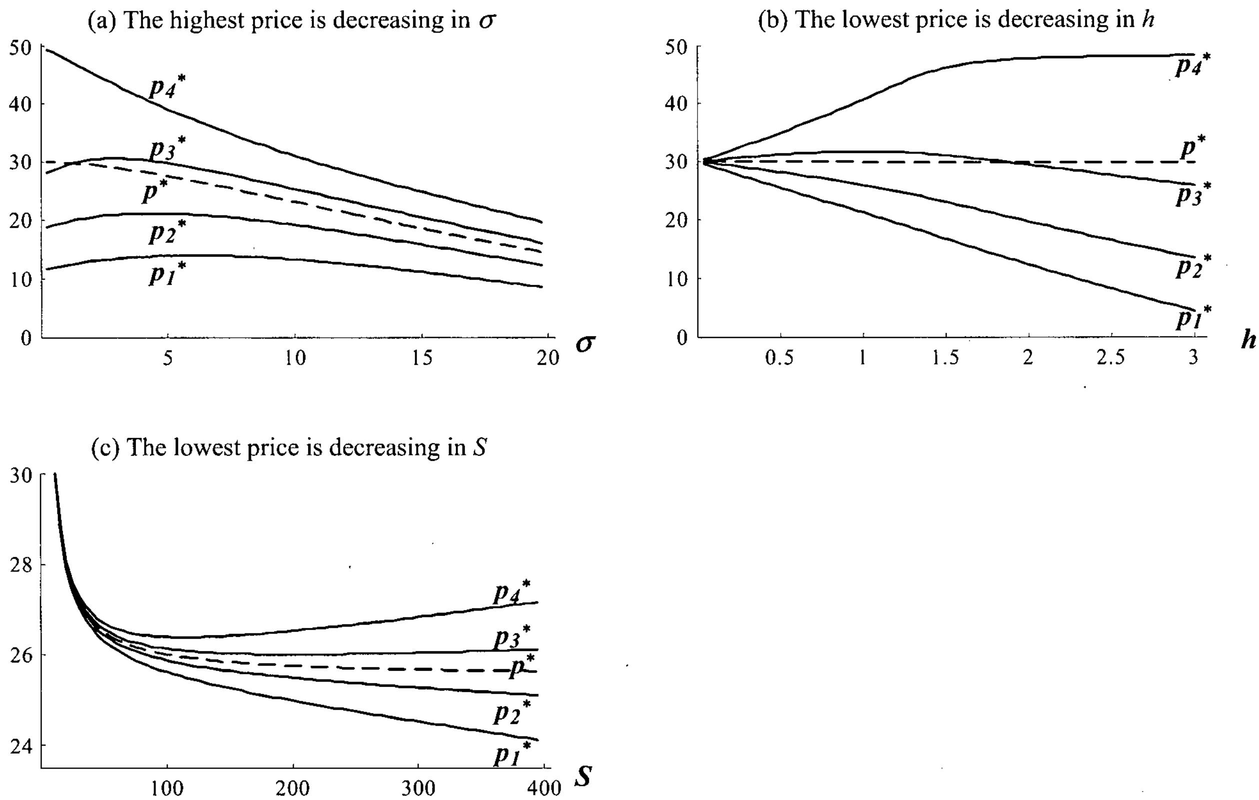
(iii) If $a'(S) < 0$, then p_1^* is decreasing in S.

(iv) The optimal profit is decreasing in σ and h.

Part (i)-(iii) of the above proposition rely on the interior optimum assumption, but we find through numerical tests that these monotonicity results hold even without the interior optimum assumption. This proposition shows that the monotonicity properties in Proposition 3.2 only partially hold here. The next example provides comparison plots for these two propositions and also provides an illustration for Theorem 3.1.

Example 3.5 (a) Let $p(\lambda) = 50 - 10\lambda$, $c(S) = 500 + 5S$, $S = 100$, $h = 2$, and let σ vary from 0 to 20. For $N = 4$, the optimal prices are plotted in Figure 3.3(a). The highest price is decreasing in σ while the others are not. There also exist other instances in which all of the prices are decreasing in σ. The demand variability incurs additional holding cost $\frac{h\sigma^2 \sum \mu_n^2}{2 \sum \mu_n}$. When σ increases, in order to balance out the increase of this cost, intuitively we need to decrease $\boldsymbol{\mu}$ and at the same time decrease the spread of $\boldsymbol{\mu}$ (i.e., $\mu_N - \mu_1$). The composite effect is mixed except for μ_N^*.
(b) Let $p(\lambda) = 50 - 10\lambda$, $c(S) = 500 + 5S$, $S = 100$, $\sigma = 1$, and let h vary from 0 to 3. The optimal prices are shown in Figure 3.3 (b). The lowest price is decreasing in h while the others are not. The spread of the prices increases significantly with the increase in h.
(c) Let $p(\lambda) = 50 - \lambda$, $c(S) = 100 + S$, $h = 0.5$, $\sigma = 3$, and let S vary from 0 to 400. The optimal prices are shown in Figure 3.3 (c). Note that $a'(S) < 0$ for all S. The lowest price is decreasing in S while the others are not. The spread of the prices increases in S.

Figure 3.3: Optimal N prices and the optimal single price with a fixed replenishment level



All of the above examples show $p_1^* < \cdots < p_N^*$. For comparison, the optimal price for the single-price problem is also shown as p^* in Figure 3.3. ∎

Finally we derive an upper bound to quantify the potential benefit of multiple price changes over a single price. To provide intuition of what enhances and what limits the profit improvement, we note that the holding cost attributed to the deterministic part of the demand is the *only* motive for varying prices over the periods. In the numerator of (3.10), the additional holding cost term $-\sum \frac{1}{2}h\sigma^2\mu_n^2$ and the revenue term $\sum p(\frac{1}{\mu_n})$ are, on the contrary, suggesting not varying prices, since they are both concave in $\boldsymbol{\mu}$. These two terms limit the extend to which the optimal prices vary, and thus limit the potential profit improvement. The following lemma and the theorem formalize this intuition.

Lemma 3.2 *For fixed S, let $\boldsymbol{\mu}^*$ be the optimal N prices. Then for $1 \le m < n \le N$,*

$$\mu_n^* - \mu_m^* \quad \le \quad \frac{S(n-m)}{N\left(\sigma^2 + \frac{B}{h}\right)},$$

where $B = \inf\left\{ -\frac{d^2p(\frac{1}{\mu})}{d\mu^2} \; : \; \mu \in [\mu_1^*, \mu_N^*] \right\}$.

This lemma indicates that the optimal dynamic pricing may not involve marked price changes under some situations. Note that σ^2 and B measure the concavity of $-\sum \frac{1}{2}h\sigma^2\mu_n^2$ and $\sum p(\frac{1}{\mu_n})$, respectively. Consistent with our previous intuition, increasing either σ^2 or B results in a smaller bound.

Theorem 3.2 *For fixed S, let $\boldsymbol{\mu}^*$ be the optimal N prices, and let V_N^* be the optimal average profit defined in (3.10). Then,*

$$V_N^* - V_1^* \quad \le \quad \frac{hS^2\left(1 - N^{-2}\right)}{12\bar{\mu}\left(\sigma^2 + \frac{B}{h}\right)},$$

where $\bar{\mu} = \frac{1}{N}\sum_{n=1}^{N}\mu_n^*$ and $B = \inf\left\{ -\frac{d^2p(\frac{1}{\mu})}{d\mu^2} \; : \; \mu \in [\mu_1^*, \mu_N^*] \right\}$.

Proof. We consider a feasible single-price policy that charges a price $p(\bar{\mu}^{-1})$, and compare this with the optimal N-price policy. We show the latter has a lower average revenue, a higher additional holding cost due to demand variability, and the same average ordering cost. Hence, for the optimal N-price policy to yield a higher profit, it must have a lower average holding cost attributed to the deterministic part of the demand. The difference between these two holding cost terms corresponding to the two policies constitute an upper bound on the profit improvement from the single-price policy to the N-price policy.

Specifically, the average profit of charging a single price $p(\bar{\mu}^{-1})$ is

$$V_1 = \frac{p(\frac{1}{\bar{\mu}}) - \frac{hS}{2}\bar{\mu} - \frac{h\sigma^2}{2}\bar{\mu}^2 - a(S)}{\bar{\mu}}.$$

We have

$$\begin{aligned} & V_N^* - V_1^* \\ \leq\ & V_N^* - V_1 \\ =\ & \frac{\sum_{n=1}^N \left[p(\frac{1}{\mu_n^*}) - p(\frac{1}{\bar{\mu}})\right] - \frac{hS}{N}\sum_{n=1}^N \left[(N-n+\frac{1}{2})\mu_n^* - \frac{N}{2}\bar{\mu})\right] - \frac{h\sigma^2 N}{2}\left[\frac{1}{N}\sum_{n=1}^N \mu_n^{*2} - \bar{\mu}^2\right]}{\sum_{n=1}^N \mu_n^*} \end{aligned}$$

The first term in the numerator, $\sum_{n=1}^N \left[p(\frac{1}{\mu_n^*}) - p(\frac{1}{\bar{\mu}})\right] \leq 0$, since $p(\frac{1}{\mu})$ is concave in μ; and the last term in the numerator $\frac{1}{N}\sum_{n=1}^N \mu_n^{*2} - \bar{\mu}^2 \geq 0$, which is essentially the Cauchy-Schwartz inequality. Thus,

$$\begin{aligned} V_N^* - V_1^* &\leq \frac{-\frac{hS}{N}\sum_{n=1}^N \left[(N-n+\frac{1}{2})\mu_n^* - \frac{N}{2}\bar{\mu})\right]}{\sum_{n=1}^N \mu_n^*} \\ &= \frac{hS}{N}\left(\frac{\sum_{n=1}^N n\mu_n^*}{\sum_{n=1}^N \mu_n^*} - \frac{N+1}{2}\right) \\ &= \frac{hS}{N}\left(\frac{2\mu_1^* + 4\mu_2^* + \cdots + 2N\mu_N^* - (N+1)(\mu_1^* + \cdots + \mu_N^*)}{2\sum_{n=1}^N \mu_n^*}\right) \\ &= \frac{hS}{N}\left(\frac{(N-1)(\mu_N^* - \mu_1^*) + (N-3)(\mu_{N-1}^* - \mu_2^*) + \ldots}{2N\bar{\mu}}\right) \end{aligned}$$

where in the last line, the series ends with $\mu_{N/2+1}^* - \mu_{N/2}^*$ if N is even, and ends with $2\Big(\mu_{(N+3)/2}^* - \mu_{(N-1)/2}^*\Big)$ if N is odd.

Applying Lemma 3.2 and the identity:

$$(N-1)^2 + (N-3)^2 + \cdots + (N+1-2\Big\lfloor\frac{N}{2}\Big\rfloor)^2 = \frac{(N-1)N(N+1)}{6},$$

we have

$$\begin{aligned} V_N^* - V_1^* &\leq \frac{hS^2}{\bar{\mu}\left(\sigma^2 + \frac{B}{h}\right)}\left(\frac{(N-1)^2 + (N-3)^2 + \cdots + (N+1-2\lfloor\frac{N}{2}\rfloor)^2}{2N^3}\right) \\ &= \frac{hS^2}{\bar{\mu}\left(\sigma^2 + \frac{B}{h}\right)}\frac{(N-1)N(N+1)}{12N^3} \\ &= \frac{hS^2\left(1 - N^{-2}\right)}{12\bar{\mu}\left(\sigma^2 + \frac{B}{h}\right)}. \end{aligned}$$

∎

The bound in the above proposition indicates that the N-price policy cannot improve much over the single-price policy when the replenishment quantity S and the holding cost rate h are low, and the demand variability σ^2 and the concavity of the revenue function B are high. These parameters affect the above bound in the same way as they affect the price increment bound in Lemma 3.2.

Also note that this bound does not rely on the interior optimum assumption.

3.3.3 Joint Pricing-Replenishment Optimization

3.3.3.1 Single Price

In this case, the firm's problem is

$$\max_{S>0,\, \lambda\in\mathcal{L}} \quad V(S,\lambda) = r(\lambda) - \frac{hS}{2} - \lambda a(S) - \frac{h\sigma^2}{2\lambda}. \tag{3.12}$$

The first-order conditions are given by (3.6) and (3.9). The difficulty is, $V(S,\lambda)$ may not be concave in (S,λ), and it may even have multiple local maxima (see Example 3.6 below). However, for some special case, a solution satisfying the first order conditions and yielding positive profit must be the global optimum, as demonstrated in the following proposition and example. See appendix for the proof.

Proposition 3.4 *Suppose $\lambda(p) = \alpha - \beta p$, $c(S) = K + cS$, where $\alpha, \beta, K, c > 0$. If $S^* > 0$ and $\lambda^* \in (0, \alpha]$ satisfy the first-order conditions in (3.6) and (3.9), and $V(S^*, \lambda^*) > 0$, then (S^*, λ^*) solves problem (3.12).*

Example 3.6 Consider $\lambda(p) = 50 - p$, $c(S) = 500 + 2S$, $\sigma = 0.2$, and $h = 1$. The optimal S for any fixed λ is given by $S(\lambda) = 10\sqrt{10\lambda}$. Substituting this EOQ into the first-order condition (3.9) gives

$$50 - 2\lambda + \frac{1}{50\lambda^2} = \frac{5\sqrt{10}}{\sqrt{\lambda}} + 2.$$

However, simply solving the above equation yields three stationary points: $\lambda_1 = 0.01619$, $\lambda_2 = 0.1010$, and $\lambda_3 = 22.327$. Using the second-order condition, it can be verified that λ_1 and λ_3 are local maxima, while λ_2 is a local minimum. Furthermore, $V(S(\lambda_1), \lambda_1) = -4.482$ and $V(S(\lambda_3), \lambda_3) = 423.8$. Hence $\lambda^* = \lambda_3 = 22.327$ and $S^* = 149.42$. ∎

The monotonicity of the joint optimum is explored in the following proposition (compared with Proposition 3.1 and 3.2 where a single decision is optimized).

Proposition 3.5 *If there is a unique optimal price and replenishment level, then*
(i) the optimal price is decreasing in σ;
(ii) the optimal replenishment level is increasing in σ and decreasing in h;
(iii) the optimal profit is decreasing in σ and h.

Remark: When the optimal solutions are not unique, the results in the above proposition continue to hold. In lieu of increasing and decreasing, the relevant properties are *ascending* and *descending*, respectively. Refer to Topkis (1978).

In Figure 3.1(b), we have seen that the optimal price decreases in σ, and the optimal profit also decreases in σ, which are consistent with the results in Proposition 3.5.

Intuitively, the higher the unit holding cost h, the less inventory should be held; and the less order quantity means the higher average cost and thus the higher price. But this intuition is correct only when the demand is deterministic ($\sigma = 0$). When $\sigma > 0$, a lower price may offset the additional holding cost due to demand variability. The composite effect is mixed, as shown in the following example.

Example 3.7 Let $\lambda(p) = \alpha e^{-p}$ with $\alpha > 0$, and $c(S) = S(K - \log S)$ for $0 \leq S \leq e^{K-1}$, where $K > 0$ is given. Note that $r(\lambda) = \lambda \log(\alpha/\lambda)$ is strictly concave for $\lambda \in (0, \alpha]$, $c(S)$ is strictly increasing for $S \in [0, e^{K-1}]$ and $a(S) = K - \log S$ is strictly convex and approaches to infinity as $S \to 0$, so Assumptions 3.1 and 3.2 are satisfied.

In this case, the first-order conditions (3.6) and (3.9) become

$$\log \alpha - \log \lambda - 1 = K - \log S - \frac{h\sigma^2}{2\lambda^2} \qquad \text{and} \qquad -\frac{1}{S} = -\frac{h}{2\lambda},$$

which determine the optimal solution:

$$\lambda^* = \frac{\sigma}{2}\sqrt{\frac{2h}{K + 1 + \log(h/2\alpha)}} \qquad \text{and} \qquad S^* = \frac{2\lambda^*}{h}.$$

The above is indeed a global optimal solution when $\lambda^* < \alpha$, and $S^* \leq e^{K-1}$; the latter is satisfied if we choose K large enough. Now, the partial derivative,

$$\frac{\partial(\lambda^{*2})}{\partial h} = \frac{\sigma^2(K + \log(h/2\alpha))}{2(K + 1 + \log(h/2\alpha))^2}$$

varies from negative to positive, as h increases from zero. Thus, λ^* first decreases and then increases in h, or equivalently, p^* first increases and then decreases in h. ∎

Finally, we compare the joint optimization here with the sequential optimization scheme that is usually followed in practice: the marketing/sales department first makes the pricing decision, and then the purchasing department decides the replenishment quantity based on demand projection as a consequence of the pricing decision. For instance, the marketing/sales department solves the problem:

$$\lambda^\dagger = \arg\max_{\lambda}\{r(\lambda)\},$$

and sets $p^\dagger = p(\lambda^\dagger)$. Then, the purchasing department takes $\lambda^\dagger$ as given and solves the problem:

$$S^\dagger = \arg\min_{S}\left\{\lambda^\dagger a(S) + hS/2\right\}.$$

Clearly, the sequential decision procedure does not take demand variability into account. The optimal price and inventory level found by the sequential scheme certainly satisfy the first-order condition in (3.6). But the first-order condition in (3.9) holds only when the demand variability happens to be $\widehat{\sigma} = \lambda^\dagger\sqrt{2a(S^\dagger)/h}$.

Proposition 3.6 *Comparing to the joint optimization,*

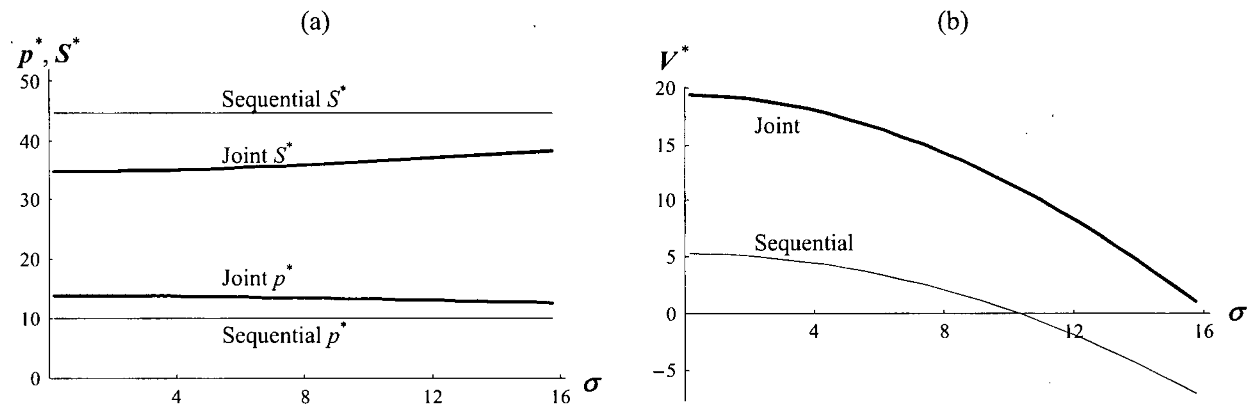
(i) if $\sigma < \widehat{\sigma}$, then the sequential decision underprices and overstocks;

(ii) if $\sigma > \widehat{\sigma}$, then the sequential decision overprices and understocks.

The proof is a straightforward application of the monotonicity result in Proposition 3.5, and will be omitted. In general, the sequential optimization is sub-optimal. If the demand variability is far away from $\widehat{\sigma}$, the sequential decision can lead to significant profit loss, as illustrated in the following examples.

Example 3.8 Let $\lambda(p) = 20 - p$, $c(S) = 100 + 5S$ and $h = 1$. Suppose the marketing/sales department sets price at $p^\dagger = 10$ in order to maximize the revenue rate. Then, the purchasing department minimizes the operating cost using the EOQ model: $S^\dagger = 20\sqrt{5} \approx 44.7$. The threshold $\widehat{\sigma} = \lambda^\dagger\sqrt{2a(S^\dagger)/h} \approx 38$. In Figure 3.4 we compare the sequential decision with the joint decision.

Figure 3.4: Sequential vs. joint pricing and replenishment decisions



The parameter range is chosen such that the best decision will be able to achieve a positive profit. Substantial profit loss is observed: when $\sigma = 0$, the sequential decision both underprices and overstocks by 28%, resulting in 73% profit loss compared to the joint decision; when $\sigma = 10$, the sequential decision underprices by 25% and overstocks by 22%, making almost no profit. ∎

3.3.3.2 N Prices

The problem is

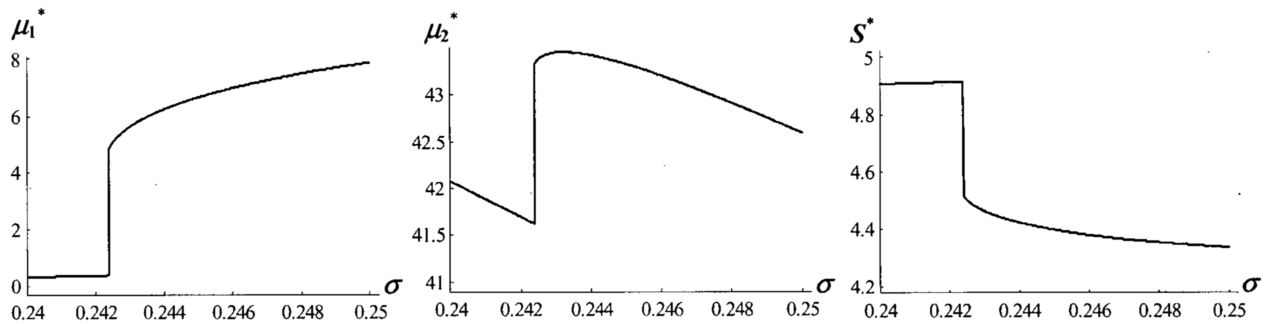
$$\max_{S>0,\, \boldsymbol{\mu}\in\mathcal{M}^N} \quad V(S,\boldsymbol{\mu}) = \frac{\sum_{n=1}^{N}\left[p(\frac{1}{\mu_n}) - \frac{hS}{N}(N-n+\frac{1}{2})\mu_n - \frac{h\sigma^2}{2}\mu_n^2 - a(S)\right]}{\sum_{n=1}^{N}\mu_n}. \tag{3.13}$$

As in the single price case, the objective function $V(S,\boldsymbol{\mu})$ may not be jointly concave. Under the interior optimum assumption, the optimal solution must satisfy the first-order conditions in (3.6) and (3.11).

The result in Theorem 3.1 (which holds for any replenishment level) continues to hold here, i.e., $p_1^* \leq p_2^* \leq \cdots \leq p_N^*$. However, the monotonicity in Proposition 3.3 and Proposition 3.5 need not hold here, as evident from the following example.

Example 3.9 Let $p(\lambda) = 10 - 10^{-3}\lambda + \lambda^{-1}$, $c(S) = 50 + S^2$ and $h = 0.2$. (Note the term λ^{-1} in $p(\lambda)$ only adds a constant to the objective function to ensure that the average profit is positive.) Consider $N = 2$. The optimal solutions are plotted in Figure 3.5. Two observations emerge from the figure. First, there exist jumps in the optimal policy, due to multiple local maxima. For example, for $\sigma = 0.243$, there are at least two local maximizers: $(S = 4.917, \mu_1 = 0.401, \mu_2 = 41.51)$ and

Figure 3.5: Impact of demand variability on joint optimal solutions



($S = 4.464, \mu_1 = 5.637, \mu_2 = 43.44$). The first yields an objective value $V = 0.2639$, which is slightly higher than the second one ($V = 0.2638$). For $\sigma = 0.244$, the two local maximizers are slightly different: ($S = 4.918, \mu_1 = 0.448, \mu_2 = 41.33$) and ($S = 4.425, \mu_1 = 6.248, \mu_2 = 43.41$). However, the objective value corresponds to the second one ($V = 0.261908$) is slightly better than the first ($V = 0.261903$). Hence, the optimal policy exhibits discontinuity when σ varies from 0.243 to 0.244. (These numerical results are accurate to the decimal places used. Furthermore, these phenomena can also be verified analytically.) Our second observation is that when the optimal solutions are continuous in σ, there exists a range in which both μ_1 and μ_2 are increasing in σ, while S is decreasing in σ. This is in sharp contrast with the results in the single-price case. ∎

Next, we develop a bound on the profit improvement as the number of prices (N) increases.

Theorem 3.3 *Let $(S_N^*, \boldsymbol{\mu}^*)$ be the optimal joint pricing-replenishment decision, and V_N^* be the corresponding optimal profit in (3.13). Then,*

$$V_N^* - V_1^* \quad \leq \quad \frac{h {S_N^*}^2 \left(1 - N^{-2}\right)}{12\bar{\mu}\left(\sigma^2 + \frac{B_N}{h}\right)},$$

where $\bar{\mu} = \frac{1}{N}\sum_{n=1}^{N} \mu_n^*$ and $B_N = \inf\left\{ -\frac{d^2 p(\frac{1}{\mu})}{d\mu^2} \;:\; \mu \in [\mu_1^*, \mu_N^*] \right\}$.

Proof. Let $V_N^*(S)$ denote the optimal profit when S is given (i.e., pricing decision only). If S happens to be fixed at S_N^*, then the profit is V_N^*, i.e., $V_N^* = V_N^*(S_N^*)$. Applying Theorem 3.2, we have the desired bound immediately:

$$V_N^* - V_1^* \;=\; V_N^*(S_N^*) - V_1^*(S_1^*) \;\leq\; V_N^*(S_N^*) - V_1^*(S_N^*) \;\leq\; \frac{h {S_N^*}^2 \left(1 - N^{-2}\right)}{12\bar{\mu}\left(\sigma^2 + \frac{B_N}{h}\right)}.$$

∎

The shortfall of the above bound is that it involves the solution to the N-price problem. Heuristically, we can use the solution to the joint single-price and replenishment problem, denoted by (S_1^*, μ^*), in the upper bound. Specifically, replace S_N^* by S_1^*, replace $\bar{\mu}$ by μ^*, replace B_N by $B_1 = -\frac{d^2p}{d\mu^2}(\frac{1}{\mu^*})$, and replace $1 - N^{-2}$ by $3/4$ to tighten the bound. That is, the bound in Theorem 3.3 can be evaluated heuristically as follows:

$$\frac{hS_1^{*2}}{16\mu^*\left(\sigma^2 + \frac{B_1}{h}\right)}. \tag{3.14}$$

Example 3.10 Let $c(S) = K + cS$ and $p(\lambda) = a - b\lambda$, where a, b, c, K are all positive parameters. The optimization problem in (3.13) becomes

$$\max_{S>0,\, \boldsymbol{\mu}\in[b/a,\infty)^N} \quad V(S, \boldsymbol{\mu}) = \frac{\sum_{n=1}^N \left[a - \frac{b}{\mu_n} - \frac{hS}{N}(N - n + \frac{1}{2})\mu_n - \frac{h\sigma^2}{2}\mu_n^2 - \frac{K}{S} - c\right]}{\sum_{n=1}^N \mu_n}.$$

Applying a change of variables, $\boldsymbol{\mu} = b\widetilde{\boldsymbol{\mu}}$ and $S = K\widetilde{S}$, we can rewrite the above problem as follows:

$$\max_{\widetilde{S}>0,\, \widetilde{\boldsymbol{\mu}}\in[1/a,\infty)^N} \quad V(\widetilde{S}, \widetilde{\boldsymbol{\mu}}) = \frac{\sum_{n=1}^N \left[a - c - \frac{1}{\widetilde{\mu}_n} - \frac{Khb\widetilde{S}}{N}(N - n + \frac{1}{2})\widetilde{\mu}_n - \frac{hb^2\sigma^2}{2}\widetilde{\mu}_n^2 - \frac{1}{\widetilde{S}}\right]}{b\sum_{n=1}^N \widetilde{\mu}_n}. \tag{3.15}$$

Clearly, the above expression indicates that there are four degrees of freedom in terms of independent parameters: $(N,\ a - c,\ Khb,\ hb^2\sigma^2)$. Specifically, the four degrees of freedom are determined by N, either a or c, and two from (K, h, b, σ). In the numerical studies reported here and below, we choose to vary (N, c, h, σ) while fixing (K, a, b).

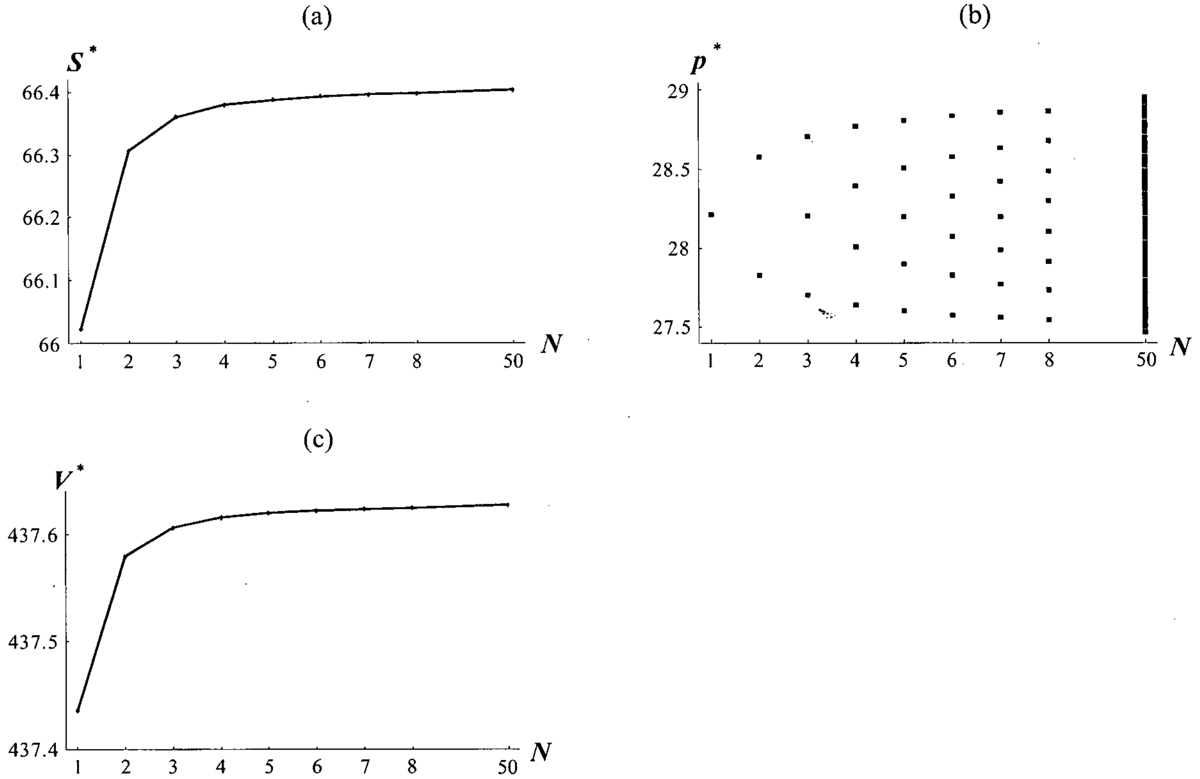
Figure 3.6 shows the optimal replenishment levels, prices, and profit values corresponding to different values of N, the number of price changes. We see that as N increases the optimal prices are inter-leaved (e.g., the optimal single price is sandwiched between the two-price solutions, which, in turn, are each sandwiched between a neighboring pair of the three-price solutions). This interleaving property seems to persist in all examples we have studied. Figure 3.6(c) shows that N has a decreasing marginal effect on profit. Using two prices already achieves most of the potential profit improvement, and beyond $N = 8$, the marginal improvement is essentially nil. ∎

Example 3.11 We continue with the last example, but focus on comparing the optimal profits under $N = 1$ and $N = 8$. We consider the parameter values $c \in (0, 30], h \in (0, 50], \sigma \in (0, 50]$, while fixing the others at $K = 100,\ a = 50,\ b = 1$.

The results reported in Figure 3.7 are for $c = 1$. Similar results are observed for $c = 5, 10, 20, 30$.

Figure 3.7(a) shows the optimal profit corresponding to a single price. The profit is clearly

Figure 3.6: Effect of the number of price changes.

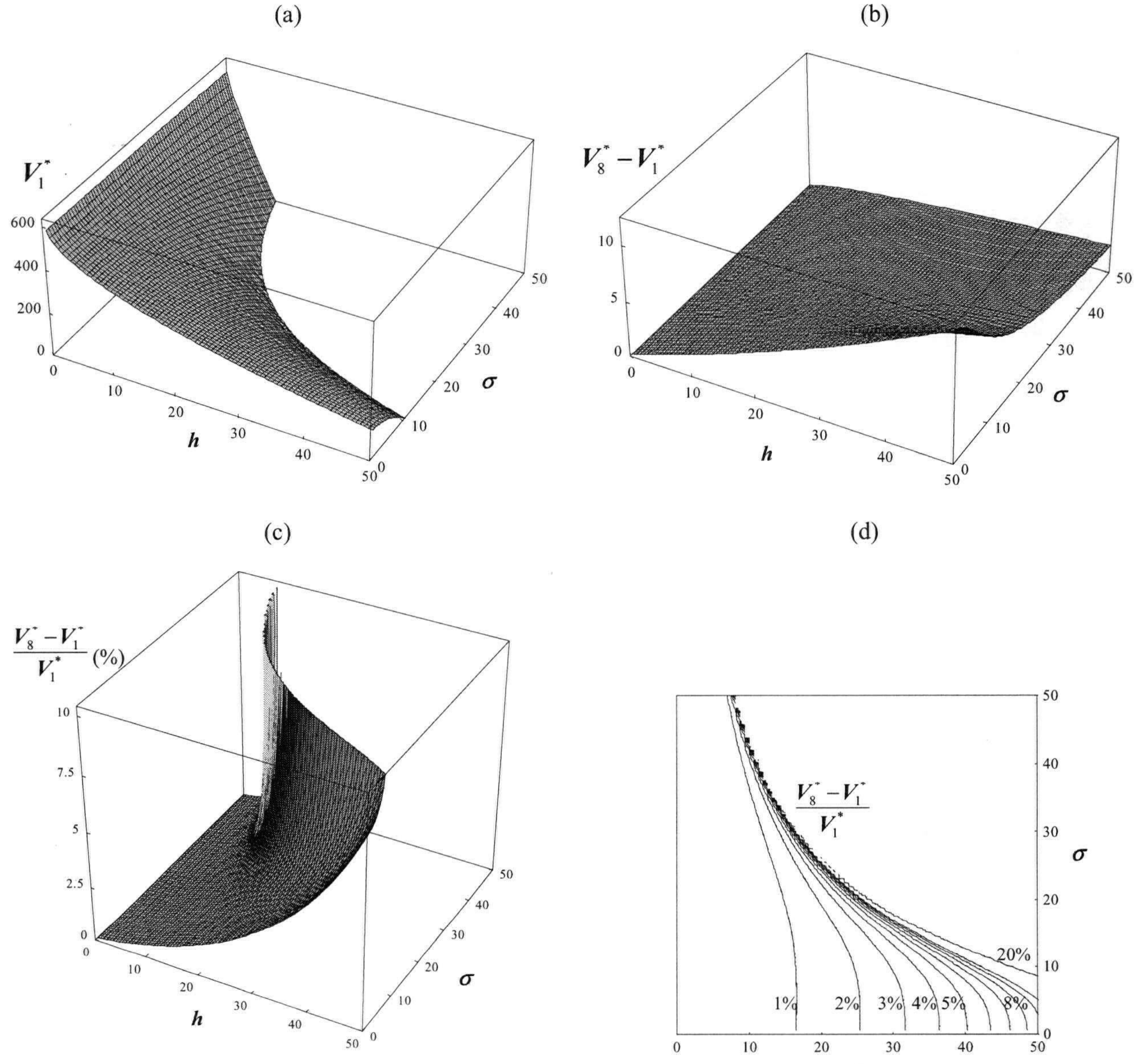


decreasing in h and σ (Proposition 3.5). Furthermore, the negative effect of σ is larger when h is larger and vice versa, suggesting that the profit function is submodular in (h, σ). Intuitively, since the effect of σ shows up in the profit function through the additional holding cost, the higher the h (resp. σ) value, the more sensitive is the profit to σ (resp. h).

Figure 3.7(b) shows the absolute improvement in profit when using 8 prices ($V_8^* - V_1^*$). The improvement is increasing in h and decreasing in σ. Intuitively, as the inventory holding cost increases, the right trade-off among revenue, holding cost and replenishment cost becomes more important, thus more pricing options over time is more beneficial. However, pricing becomes less effective as demand variability increases.

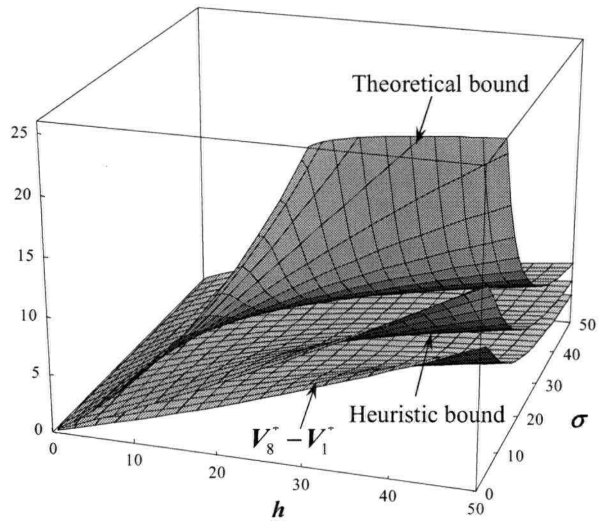
Note that when the profit under a single price V_1^* approaches zero, the absolute improvement ($V_8^* - V_1^*$) does not diminish. In particular, while V_1^* decreases in h, the improvement increases in h. Indeed, the *relative* improvement ($= \frac{V_8^* - V_1^*}{V_1^*}$) is increasing in h, and approaching infinity when V_1^* is close to zero, as demonstrated in Figure 3.7(c) and (d).

Figure 3.7: Profit improvement of multiple price changes over a single price



To conclude this example, we show in Figure 3.8 the profit improvement in comparison with the theoretical bound in Theorem 3.3 and heuristic bound in (3.14). The heuristic bound can be seen in this example as giving a good estimate of the maximum potential improvement. ∎

Figure 3.8: Bounds on the profit improvement



3.3.3.3 The Algorithm: Fractional Programming

To solve the fractional optimization problem in (3.13), we can instead solve the following:

$$\max_{S>0,\,\boldsymbol{\mu}\in\mathcal{M}^N} V_\eta := \sum_{n=1}^{N}\left[p\Big(\frac{1}{\mu_n}\Big) - \frac{hS(N-n+\frac{1}{2})}{N}\mu_n - \frac{h\sigma^2}{2}\mu_n^2 - a(S)\right] - \eta\sum_{n=1}^{N}\mu_n, \qquad (3.16)$$

where η is a parameter. When the optimal objective value of the problem in (3.16) is zero, the corresponding solution is the optimal solution to the original problem in (3.13), and the corresponding η is the optimal value of the original problem. To see the equivalence, write the optimal value in (3.13) as $V^* = A^*/B^*$, where A and B denote the numerator and denominator in (3.13) respectively. When $\eta = V^*$, the optimal value of (3.16) is zero. On the other hand, suppose there exists an η which yields a zero objective value in (3.16), specifically, $A^* - \eta B^* = 0$, then for any feasible A, B, we have $A - \eta B \leq 0$, which implies $A/B \leq \eta = A^*/B^*$.

The algorithm described below takes advantage of the separability of the objective function with respect to $\boldsymbol{\mu}$ when S and η are given. Specifically, the sub-problems are:

$$\max_{\mu_n\in\mathcal{M}} g_n(\mu_n) := p\Big(\frac{1}{\mu_n}\Big) - \frac{hS(N-n+\frac{1}{2})}{N}\mu_n - \frac{h\sigma^2}{2}\mu_n^2 - a(S) - \eta\mu_n, \quad n = 1,\ldots,N. \qquad (3.17)$$

Under Assumption 3.1, the objective in (3.17) is concave in μ_n.

`Algorithm for solving (3.13)`

1. `Initialize` $\eta=\eta_0$, $S=S_0$, `and` $\mu_n=\mu$ `for` $n=1,\ldots,N$.

(For instance, $\eta_0 = 0$; $S_0 = S^*$ and $\mu_n = \mu^*$ for all n,
with (S^*, μ^*) being the optimal solution to the single-price problem.)
Set ε at small positive values (according to the required precision).

2. Solve the N single-dimensional concave function maximization problems in (3.17).

3. Update S following a specified stepsize or use equation (3.6).
If the difference between the new and the old S values is smaller than ε,
go to step 4; otherwise, return to step 2.

4. $V_\eta =$ sum of the objective values of the N sub-problems.
If $|V_\eta| \leq \varepsilon$, stop. Otherwise,
if $V_\eta > \varepsilon$, increase η;
if $V_\eta < -\varepsilon$, decrease η.
Go to step 2.

In step 2, there are ways to accelerate the search procedure. First, we can take the advantage of the monotonicity of μ_n^* in n following Theorem 3.1. Second, we have $g_n(\mu_n^*) \leq g_{n+1}(\mu_n^*) \leq g_{n+1}(\mu_{n+1}^*)$, for $n = 1, \ldots, N-1$. Hence, if we find μ_1^* such that $g_1(\mu_1^*) > \varepsilon/N$, then we can be sure that $V_\eta > \varepsilon$, and to bypass the rest of the algorithm to increase η directly and proceed to the next loop. Similarly, if we first solve for μ_N^* and find that $g_N(\mu_N^*) < -\varepsilon/N$, we can decrease η directly. Third, we can make use of the monotonicity of μ_n^* in S according to Proposition 3.3 (iii). This is particularly useful when the stepsize for updating S in step 3 is small.

In step 3, if we are in the early stage of the algorithm, i.e., when V_η is still substantially away from zero, then the updating of S needs to cover a wide range so as not to miss the true optimum. This can be done by using large stepsize first and then reduce them gradually. When we are at the late stage of fine-tuning V_η, we can update S to nearby values.

In step 3, we can use (3.6) as an updating scheme. This is in fact the coordinate ascent method: we alternate between optimizing $\boldsymbol{\mu}$ for fixed S and optimizing S for fixed $\boldsymbol{\mu}$. This procedure is guaranteed to converge to a local maximum, but not necessary the global maximum (see Luenberger (1984)). So it is useful once the algorithm enters the region containing the true optimum without other local maxima.

It can be verified that the optimal objective value in (3.16) is strictly decreasing in η, so there is a unique zero-crossing point at which V_η is zero. Thus, in step 4, we can update η following a standard line search algorithm, such as the bi-section or the golden ratio.

The above algorithm with a slight modification appears more suitable for applications. In practice, it is often relevant to ask how many prices to use, when to change prices and what prices to change to. In addition, the prices and the replenishment level may be restricted to some given (discrete) sets of values. It is readily verified that the equivalence between (3.13) and (3.16) still holds under discrete sets. To solve the discrete version of (3.16), we simply modify step 2 of the algorithm such that N single-dimensional concave maximization problems are solved, and modify step 3 such that the updating scheme for S is restricted to the discrete set. Using a large N, the solution can effectively answer the questions posed above, as evident from the following example.

Example 3.12 Let $\lambda(p) = 50 - p$, $c(S) = 100 + S$, $h = 1$, $\sigma = 10$. Suppose the prices have to be integer valued, and S has to be a multiple of 5. Using the algorithm (adapted as above), and choosing $N = 140$, we find the optimal policy is to order 70 units for each cycle, charge a price of \$25 until the inventory level drops to 67 units, charge \$26 until inventory drops to 19, and charge \$27 until the inventory runs out. The policy uses only three prices (despite the choice of a large N value) and yields a profit of 528.745.

In other words, the equal partitioning of $[0, S]$ (into 140 segments) does not prevent the algorithm to find the optimal partitioning (of three uneven segments). ∎

3.4 Conclusion and Extensions

We have demonstrated the effectiveness of the Brownian motion model for price-sensitive demand, in particular in making optimal pricing and replenishment decisions, quantifying the profit improvement, and bringing out the impact of demand variability. The Brownian demand model also facilitates making connections to and comparisons with, wherever applicable, previously known results in both deterministic and stochastic settings.

In the rest of this chapter, we elaborate on three possible extensions of our model and results developed above. More details are provided in Appendix B.

3.4.1 Discounted Objective: An Example

Many results in Section 3.3 continue to hold if the long-run average objective is replaced by a discounted objective. While we relegate the technical details to the appendix, we present below a simple example to highlight the contrasts between the two cases.

Let $\gamma > 0$ be the discount rate. Under a single price policy, the discounted profit starting from zero inventory takes the form (equality (A.38) in the appendix),

$$V_\gamma(S,\lambda) \;=\; \frac{r(\lambda)}{\gamma} - \frac{S(h/\gamma + a(S))}{1 - e^{-b(\lambda)S}} + \frac{h\lambda}{\gamma^2},$$

where $b(\lambda) = 2\gamma/(\sqrt{\lambda^2 + 2\sigma^2\gamma} + \lambda)$. Unlike the average objective case, the additional cost due to demand variability, defined as the difference between $V_\gamma(S,\lambda)$ with a positive σ and $V_\gamma(S,\lambda)$ with σ set to zero, now depends on S.

We first fix the price and only optimize the replenishment level, that is

$$\max_{S>0} \; V_\gamma(S,\lambda).$$

The first-order condition is

$$[h/\gamma + a(S) + Sa'(S)](e^{bS} - 1) \;=\; bS[h/\gamma + a(S)]. \tag{3.18}$$

Recall, under the average objective and fixed price, the demand variability has no effect on the optimal replenishment level. In contrast, the demand variability will raise the replenishment level under the discounted objective. (See the proof in the appendix.) The following example illustrates this effect.

Suppose $c(S) = K + cS$, where $K, c > 0$. When γ is small, b is also small and the first order condition (3.18) can be written as

$$(h/\gamma + c)(bS + \tfrac{1}{2}b^2S^2 + o(b^2)) \;=\; bS(h/\gamma + K/S + c).$$

Then,

$$S^*_\gamma \;\approx\; \sqrt{\frac{2K\gamma}{b(h + c\gamma)}}.$$

Approximating b by Taylor series: $b \approx \frac{\gamma}{\lambda} - \frac{\sigma^2\gamma^2}{2\lambda^3}$, we have

$$S^*_\gamma \;\approx\; \sqrt{\frac{2K(\lambda + \frac{\sigma^2\gamma}{2\lambda})}{h + c\gamma}}.$$

Porteus (1985b) pointed out that the EOQ solution under a discounted objective can be approximated by $\sqrt{\frac{2K\lambda}{h+c\gamma}}$. Our result generalizes this result by incorporating the effect of demand variability.

3.4.2 Full Backordering

Our model can be extended to allow backordering, in which case the replenishment takes the form of a $(-s, S)$ policy: a replenishment order is issued whenever the number of backorders has reached s, to bring the inventory level back to S. Hence, the replenishment quantity is $S + s$. (As before, zero lead time is assumed.) The pricing policy is modified accordingly: equally divide S and s into N and M (positive integers) segments, respectively, such that

$$S = S_0 > S_1 > \cdots > S_N = 0 > S_{N+1} > \cdots > S_{N+M} = -s,$$

and apply price p_n until the net inventory (inventory on hand net the backorders) falls to S_n, $n = 1, \ldots, N + M$.

Assuming linear backordering cost, most of the results we have derived for the no-backorder case will continue to hold, with suitable modifications. For instance, the optimal pricing will satisfy the monotonicity:

$$p_1^* \leq p_2^* \leq \cdots \leq p_N^*, \qquad p_{N+1}^* \geq p_{N+2}^* \geq \cdots \geq p_{N+M}^*.$$

That is, the optimal prices increase as the on-hand inventory is depleted, and decrease as the backorders accumulate. The bound on the profit improvement is (assuming $M = N$):

$$V_N^* - V_1^* \leq \frac{(1 - N^{-2})}{12} \left[\frac{hS^2}{\bar{\mu}_h \left(\sigma^2 + \frac{B}{h}\right)} + \frac{bs^2}{\bar{\mu}_b \left(\sigma^2 + \frac{B}{b}\right)} \right],$$

where b is the cost for holding one backorder per unit of time, $\bar{\mu}_h = \frac{1}{N} \sum_{n=1}^{N} \mu_n^*$, $\bar{\mu}_b = \frac{1}{N} \sum_{n=N+1}^{2N} \mu_n^*$, and B denotes, as before, the lower bound for $-\frac{d^2 p(\frac{1}{\mu})}{d\mu^2}$ over the spread of $\boldsymbol{\mu}^*$.

3.4.3 Price-Sensitive Demand Variability

Here we extend the Brownian demand model in (3.1) by allowing the diffusion coefficient to depend on the demand rate, which varies with the price. Specifically,

$$D(t) \quad = \quad \int_0^t \lambda(p_u) du + \int_0^t \sigma(\lambda(p_u)) dB(u), \qquad t \geq 0, \tag{3.19}$$

where the diffusion coefficient $\sigma(\lambda)$ is a function of the demand rate. A special case is $\sigma(\lambda) = \lambda\sigma$, i.e., the demand variability is linear in the demand rate. Under $\sigma(\lambda) = \lambda\sigma$, (3.19) can be written as

$$dD(t) \quad = \quad \lambda(p_t)\big(dt + \sigma dB(t)\big), \tag{3.20}$$

which is an analogy to the multiplicative demand case in the periodic-review setting.

We summarize the results for the model with price-sensitive demand variability as follows:

1. The properties of the optimal replenishment level in Proposition 3.1 still holds. The conclusion in Theorem 3.1 that the optimal prices are increasing over the periods within each cycle still holds.

2. The upper bounds in Theorem 3.2 and 3.3 can be extended to the general case with price-sensitive demand variability.

3. The monotonicity properties of the optimal decisions with respect to demand variability depend on the functional form of $\sigma(\lambda)$. For the special case $\sigma(\lambda) = \lambda\sigma$, some of the monotonicity results in Proposition 3.2, 3.3, 3.5, and 3.6 will be in the opposite direction.

The above findings are elaborated below. The proofs are included in the appendix.

With the general model in (3.19), under the same pricing and replenishment policies, the long-run average profit functions (3.4) and (3.5) become

$$V(S,\boldsymbol{\mu}) = \frac{\sum_{n=1}^{N}\left[p(\frac{1}{\mu_n}) - \frac{hS}{N}(N-n+\frac{1}{2})\mu_n - \frac{1}{2}h\sigma(\mu_n^{-1})^2\mu_n^2 - a(S)\right]}{\sum_{n=1}^{N}\mu_n}, \tag{3.21}$$

$$V(S,\lambda) = r(\lambda) - \frac{hS}{2} - \lambda a(S) - \frac{h\sigma(\lambda)^2}{2\lambda}. \tag{3.22}$$

We assume that $C_v(\lambda) \equiv \sigma(\lambda)^2/\lambda$ is convex in λ. (This assumption holds for both $\sigma(\lambda) = \sigma$ and $\sigma(\lambda) = \lambda\sigma$). The convexity of $C_v(\lambda)$ implies that $\sigma(\frac{1}{\mu})^2\mu^2$ is convex in μ. Then, $V(S,\lambda)$ is concave in λ, and $V(S,\boldsymbol{\mu})$ is pseudo-concave in $\boldsymbol{\mu}$. The first-order conditions are thus sufficient to determine the optimal pricing decisions.

It can be easily verified that Proposition 3.1 and Theorem 3.1 still hold (with the same line of the original proofs).

The results on the upper bounds in Theorem 3.2 and 3.3 require some modification, with σ^2 in the original upper bound replaced by $G = \inf\left\{ \frac{d^2\left(\frac{1}{2}\sigma(\frac{1}{\mu})^2\mu^2\right)}{d\mu^2} \ : \ \mu \in [\mu_1^*, \mu_N^*] \right\}$. Hence, the bound in Theorem 3.3 becomes

$$V_N^* - V_1^* \le \frac{hS_N^{*\,2}\left(1-N^{-2}\right)}{12\bar{\mu}\left(G + \frac{B_N}{h}\right)}.$$

Note that if $\sigma(\lambda)$ is a constant, then $G = \sigma^2$ and we recover the original bound. The derivation of this bound relies on the convexity of $\sigma(\lambda)^2/\lambda$ and a modification of Lemma 3.2: for $1 \leq m < n \leq N$, $\mu_n^* - \mu_m^* \leq \frac{S(n-m)}{N(G+\frac{B}{h})}$. All these results are proven the appendix.

The demand variability affects the profit function through the additional holding cost term: $\frac{h\sigma(\lambda)^2}{2\lambda}$ in (3.22) and $\frac{\sum_{n=1}^N h\sigma(\mu_n^{-1})^2\mu_n^2}{2\sum_{n=1}^N \mu_n}$ in (3.21). Therefore, the monotonicity properties of the optimal decisions with respect to demand variability depend on the specification of $\sigma(\lambda)$.

For the special case $\sigma(\lambda) = \lambda\sigma$, some monotonicity results with respect to the demand variability need to be modified. For fixed S, the optimal prices $p_1^*, \ldots, p_N^*$ are all increasing in σ, which is opposite to Proposition 3.2(i) and Proposition 3.3(i). If a single price is jointly optimized with the replenishment level, the optimal price is increasing in σ, and the optimal replenishment level is decreasing in σ, which is opposite to Proposition 3.5(i) and (ii). Furthermore, from the first-order condition, the optimal price is such that $r'(\lambda) = a(S) + \frac{1}{2}h\sigma^2 > 0$. The sequential decision chooses a price satisfying $r'(\lambda) = 0$, which is always lower than the optimal price from the joint optimization. This is different from Proposition 3.6. The literature has also seen the differences in the monotonicity properties between the additive demand model and multiplicative demand model (refer to Federgruen and Heching 1999, and Polatoglu and Sahin 2000 reviewed in Section 3.1).

References

Chan, L.M.A., Z.J. Shen, D. Simchi-Levi, J.L. Swann. 2004 Coordination of pricing and inventory decisions: a survey and classification. In *Supply Chain Analysis in the eBusiness Era*, D. Simchi-Levi, D. Wu and Z.J. Shen (eds), Kluwer Academic Publishers.

Chen, F.Y., S. Ray, Y. Song. 2006. Optimal pricing and inventory control policy in periodic-review systems with fixed ordering cost and lost sales. *Naval Research Logistics*, **53**(2) 117-136.

Chen, X., D. Simchi-Levi. 2006. Coordinating inventory control and pricing strategies: the continuous review model. *Operations Research Letters*, **34**(3) 323-332.

Chen, X., D. Simchi-Levi. 2004a. Coordinating inventory control and pricing strategies with random demand and fixed ordering cost: the finite horizon case. *Operations Research*, **52**(6) 887-896.

Chen, X., D. Simchi-Levi. 2004b. Coordinating inventory control and pricing strategies with random demand and fixed ordering cost: the infinite horizon case. *Mathematics of Operations Research*, **29**(3) 698-723.

Elmaghraby, W., P. Keskinocak. 2003. Dynamic pricing in the presence of inventory considerations: research overview, current practices and future directions. *Management Science*, **49**(10) 1287-

1309.

Federgruen, A., A. Heching. 1999. Combined pricing and inventory control under uncertainty. *Operations Research*, **47**(3) 454-475.

Feng, Y., F.Y. Chen. 2003. Joint pricing and inventory control with setup costs and demand uncertainty. Working paper, Chinese University of Hong Kong.

Feng, Y., F.Y. Chen. 2004. Optimality and optimization of a joint pricing and inventory-control policy for a periodic-review system. Working paper, Chinese University of Hong Kong.

Genesove, D., W.P. Mullin. 1998. Testing static oligopoly models: conduct and cost in the sugar industry, 1890-1914. *RAND Journal of Economics*, **29**(2) 355-377.

Huh, W.T., G. Janakiraman. 2005. Optimality results in inventory-pricing control: an alternate approach. Working paper, Columbia University, New York University.

Karlin, S., H. Taylor 1975. *A First Course in Stochastic Processes* (2nd ed.), Academic Press, New York.

Li, L. 1988. A stochastic theory of the firm. *Mathematics of Operations Research*, **13**(3) 447-466.

Luenberger, D.G. 1984. *Linear and Nonlinear Programming* (2nd ed.), Addison-Wesley.

Mangasarian, O.L. 1970. Convexity, pseudo-convexity and quasi-convexity of composite functions. *Cahiers du Centre d'Études de Recherche Opérationnelle*, **12** 114-122.

Polatoglu, H., I. Sahin. 2000. Optimal procurement policies under price-dependent demand. *International Journal of Production Economics*, **65**(2) 141-171.

Porteus, E.L. 1985a. Investing in reduced setups in the EOQ model. *Management Science*, **31**(8) 998-1010.

Porteus, E.L. 1985b. Undiscounted approximations of discounted regenerative models. *Operations Research Letters*, **3** 293-300.

Rajan, A., Rakesh, R. Steinberg. 1992. Dynamic pricing and ordering decisions by a monopolist. *Management Science*, **38**(2) 240-262.

Reiss, P.C., F.A. Wolak. 2006. Structural econometric modeling: rationales and examples from industrial organization. In *Handbook of Econometrics*, Volume 6, J.J. Heckman, E.E. Leamer (eds), North-Holland.

Ross, S.M. 1996. *Stochastic Processes* (2nd ed.) Wiley, New York.

Song, Y., S. Ray, T. Boyaci. 2006. Optimal dynamic joint pricing and inventory control for multiplicative demand with fixed order costs. Working paper, McGill University.

Thomas, L.J. 1974. Price and production decisions with random demand. *Operations Research*, **22**(3) 513-518.

Thowsen, G.T. 1975. A dynamic, nonstationary inventory problem for a price/quantity setting firm. *Navel Research Logistics Quarterly*, **22** 461-476.

Topkis, D.M. 1978. Minimizing a submodular function on a lattice. *Operations Research*, **26**(2) 305-321.

Whitin, T.M. 1955. Inventory control and price theory. *Management Science*, **2**(1) 61-68.

Yano, C.A., S.M. Gilbert. 2003. Coordinated pricing and production/procurement decisions: a review. In: *Managing Business Interfaces: Marketing, Engineering and Manufacturing Perspectives*, A. Chakravarty and J. Eliashberg (eds.), Kluwer Academic Publishers.

Zabel, E. 1972. Multiperiod monopoly under uncertainty. *Journal of Economic Theory*, **5**(3) 524-536.

Chapter 4

Inventory Performance of U.S. Manufacturing Companies from 1981 to 2000*

4.1 Introduction

In the 1970s and 1980s Japanese manufacturing companies made substantial market share gains in the U.S. markets in a range of industries, including most notably the car industry. This stimulated a significant search for the reasons for their success. The "Just-In-Time" (JIT) inventory system was often identified as a key element. There were many calls for a revolution in inventory policies of American firms. It was said that American firms needed to reduce their inventories. It was said that the financial markets would reward firms that cut inventories and punish those that did not do so.

Twenty years later much less is heard about the need for revolutionary changes to inventory policies. Is this due to the fact that the revolution took place and inventories were dramatically reduced? Or did inventory policy remain largely unchanged while other issues became more topical? The only way to tell is to look at the actual inventory holdings of a large number of firms.

In this chapter we study the changes in inventories on the part of American manufacturing firms over the 1981-2000 period. We examine whether these firms actually reduced their inventories as recommended by the gurus in the late 1970s and early 1980s. Of course there is variation in inventory policy across firms. According to the gurus, firms with lean inventories are more valuable

*A version of this chapter has been published:

Hong Chen, Murray Z. Frank, Owen Q. Wu. 2005. What actually happened to the inventories of American companies between 1981 and 2000? *Management Science* **51**(7) 1015–1031.

than firms with bloated inventories. We examine whether the financial markets value firms in this manner. For both of these questions we examine the data both unconditionally and conditioning on conventional factors. In each case we find consistent results conditionally and unconditionally. Thus the trends that we document are not driven by factors that are already known in the literature.

We find that inventories were significantly reduced over the 1981-2000 period. Inventory days declined on average by about 2% per year. Inventory days declined most rapidly in computer equipment, electronic equipment, and printing and publishing industries.

When we look at the components of the overall inventory interesting differences emerge. The largest decline is found for work-in-process inventory days which declined about 6% per year. Raw materials declined about 3% per year. Finished goods inventory did not decline. In some industries, finished goods inventory days actually increased – notably in tobacco, leather goods, and medical instruments industries.

A firm that deals efficiently with its suppliers will have low raw materials inventories. A firm that has efficient internal operations will have low work-in-process inventory. From this perspective it appears that firms' inventory holdings in raw materials and work-in-process seem to have generally improved significantly over time. However, there is no strong empirical evidence regarding the finished goods inventory. Intuitively, a firm that produces based on forecasting may have higher finished goods inventory in order to have a higher service level (based on the goods availability). But the firm may have its finished goods inventory reduced with better forecasting through better supply chain coordination such as vendor managed inventory. Neither of these contradictory predictions can be shown to have a dominant effect.

When considering stock market valuation, it is important to distinguish valuation differences at a moment in time (cross-section) from valuation differences that only emerge as time progresses (time series). In the cross-section there is no evidence that the market places a higher value on firms with lower inventories. Over time, however, interesting differences do in fact emerge.

A firm with a high Tobin's q (or a high market-to-book ratio) is a firm that the market is valuing particularly highly relative to the accounting measure of value. If lean inventories are highly valued, then firms with lean inventories will have particularly high Tobin's q. We find no evidence of any such relationship in the data.

Suppose that valuation differences only emerge gradually. An investor who holds a portfolio of lean inventory firms will accumulate more wealth. An investor who ignores inventories or who holds the shares of firms with bloated inventories will accumulate less wealth.

To test this idea, we form portfolios based on each firm's abnormal inventory holdings relative to their industry peers. The portfolios are rebalanced annually to reflect changes in corporate inventory positions. The long term value of these portfolios is then compared to the values found in a large number of randomly formed portfolios. If inventory is irrelevant, no statistically significant differences should be observed.

We do find that significant valuation differences emerge over time. Firms with abnormally high inventories have abnormally poor stock returns. Firms with abnormally low inventories have ordinary stock returns. Firms with slightly lower than average inventories perform best over time. They outperform average firms by about 4.5% per year on average.[1]

Of course, many things other than inventory affect stock returns. Accordingly it is important to study whether the portfolio effects that we identify are simply proxying for some factor that is already known in the empirical literature on asset pricing. To address this concern we use the Fama and French (1993) three-factor model. This is by far the most popular empirical model of stock returns in recent years.[2] We find that the abnormal inventory effect is not accounted for by the standard model from the empirical finance literature.

There are a few previous studies that are related to our work. Balakrishnan, Linsmeier and Venkatachalam (1996) studied the accounting performance of 46 firms that adopted JIT over the 1985-1989 period. Compared to a matched sample of non-adopters, on average there is no effect on the reported return on assets – if anything, it declines slightly. The authors are particularly cautious about their results due to the small sample size.

Huson and Nanda (1995) reach a different conclusion. They studied a sample of 55 firms that adopted JIT. They report that the JIT firms do increase the inventory turnover subsequent to adoption, and that they have an increase in earnings per share. Oddly enough, the JIT firms also report a direct increase in unit costs while nonadopters were cutting their unit costs.

Sakakibara, Flynn, Schroeder and Morris (1997) carried out a questionnaire study of 41 plants

[1] Zipkin (1991) distinguished pragmatic JIT from romantic JIT. Proponents of romantic JIT really support the idea of firms having zero inventory. Proponents of pragmatic JIT support reduced inventory, but do not take the idea to an extreme. The valuation evidence is quite suggestive of pragmatic JIT.

[2] We have also tried adding the effect of stock market momentum to the model. It makes no important difference to our conclusions, so we do not report those results separately.

in the transportation components, electronics and machinery industries. In their surveys they found mixed evidence about how JIT practices and manufacturing performance are related. It is suggested that the effect of JIT comes from its effect on manufacturing strategy and quality management. They do not consider how these practices relate to financial performance.

Surveys of managers necessarily produce a relatively limited sample of firms. Caution is needed because it may be unclear how well the results from fewer than a hundred firms generalize to the thousands of publicly traded firms. However, surveys also have benefits. Those conducting surveys are not limited in the kinds of questions that can be asked.

It is also possible to study the problem at the industry level rather than the firm level. Rajagopalan and Malhotra (2001) look at inventory holdings for a number of 2-digit SIC code industries in the manufacturing sector of the U.S. economy. Overall inventories declined. They suggest a need for subsequent research of a different character: "A firm-level analysis may yield insights into the true causes of changes in inventory ratios." "Further, it would be valuable to explore the linkages between inventory performance and financial performance using firm level data." This is very much in the spirit of this study.

Hendricks and Singhal (2003) investigated the stock market reaction to the public announcement by a firm that they are experiencing supply chain glitches that are causing production or shipping delays. These commonly result from inventory problems. Based on a sample of 861 announcements, they found that the supply chain glitches significantly decrease the shareholder value. This shows that when problems in normal inventory control are large enough to be "material" and so require announcement, the market cares. The paper does not show whether "normal", but inefficient inventory control practices are deemed important. Hendricks and Singhal (2003) method also does not provide information about the trends in inventory holdings.

The rest of this chapter is organized as follows. Section 4.2 sets out the key questions to be examined. Section 4.3 describes inventory measures and the data. Section 4.4 examines whether inventory actually declined. Section 4.5 studies the financial impact of inventories. Section 4.6 provides our conclusions.

4.2 Key Questions

There are two basic questions that we set out to answer. First, did inventories actually decline? Second, are the abnormal inventories of some firms related to abnormal financial returns after controlling for the established factors that are usually used to account for stock returns? In each case there are a range of related issues that arise as we study alternative control factors, groupings of the data, and functional form specifications.

Both of our questions were stimulated by the early 1980s literature on Just-In-Time practice. A particularly striking overview was provided by Zipkin (1991). He suggests that Just-In-Time can be approached from two perspectives: pragmatic JIT and romantic JIT. Pragmatic JIT promotes inventory reduction, but not zero inventory; it focuses instead on the concrete details of the production process. Romantic JIT calls for a dramatic action and believes in zero inventory. According to Zipkin (1991), "Schonberger and others repeatedly describe inventory as wasteful, excessive, indeed 'inherently evil.' The aim should be not just to reduce it but to stamp it out." For advocacy examples, see Schonberger (1982) and Hall (1983). Zipkin (1991) also argues that stock market valuations might be affected by inventory reductions.[3]

A subsidiary question is suggested by Schonberger (1982). He argues that much of the interest in inventory reductions by American managers focused on their interactions with suppliers. If his claim provides a good characterization of what was really taking place, then we should observe the largest declines on raw materials and finished goods inventories.

The second basic issue is whether low inventories are actually desirable. If inventory reductions are a good thing, then investors should pay more for the firms that reduce inventory. Is this what we see?

To answer this question we need a measure of stock market valuation. Standard measures of valuation include the market-to-book ratio and Tobin's q. They both record the ratio of the amount that the market is willing to pay to own the firm, relative to its book value. They are measures of valuation at a point in time. If investors are willing to pay more for low inventory firms, then

[3] "The turmoil in the financial markets over the past decade has certainly contributed to the appeal of the more radical versions of JIT. Obviously, any company concerned about the price of its shares would have a strong incentive to reduce inventories, and even more to project inventory reductions in the future. Someone planning a takeover or an LBO would also find such a concept attractive. With working capital freed up, or rather with the promise of lower working-capital requirements, more debt securities can be issued to finance the transaction. And such reductions appear even more tempting when they come easily and without new capital investment." Zipkin (1991).

inventory will be negatively associated with the market-to-book ratio or Tobin's q.

It is also possible to consider longer term valuation effects. To do this we create portfolios as a function of the inventory levels, and then study how these portfolios perform over time relative to other portfolios. If abnormally low inventories are good, then we would expect to see such firms having high stock returns. It is, of course, necessary to control for the standard asset pricing factors (see Fama and French 1993).

Inventories are likely to be influenced by macroeconomic conditions. For this reason we consider the effect of using a number of standard macroeconomic factors. Specifically, we consider the effects of: the interest rate (R_f), growth in gross domestic product (GGDP), inflation rate (Infl), and the optimism expressed by purchasing managers (PMI). For our purposes the macro factors are intended as controls. Nonetheless it is worth asking what kinds of effects we expect them to have.

When the interest rate rises, inventories are more expensive relative to holding bonds. Inventory levels should drop.

The effect of a booming economy depends on whether it was anticipated or not. When the economy is expanding more rapidly than anticipated (high GDP growth), firms may have trouble keeping up with demand. Lower inventories should be seen. This should particularly affect finished goods. Conversely, when the economy is growing less rapidly than anticipated, inventories might tend to build up. These predictions depend on what the firm had been expecting. A booming economy might be booming less than had been anticipated. In that case the prediction is reversed. Accordingly the business cycle predictions are theoretically ambiguous. It is an empirical question whether there is a systematic relationship.

The expected effect of inflation has both a direct cost effect and an indirect effect that operates through the effect of inflation on interest rates.[4] High inflation makes it desirable to buy inputs early – before their prices inflate still further. Thus raw materials inventories should rise. High inflation is also associated with high interest rates. This tends to make it expensive to hold inventories. The effect on overall inventories is ambiguous for this reason.

The purchasing managers index measures optimism about the state of the economy. When managers are feeling optimistic, presumably they will prepare for extra sales and inventories should

[4] Since we have included an interest rate measure already, one might have guessed that the indirect effect would not occur. This is correct if our measure is sufficient to fully control for the full effect of the term structure of interest rates. We expect our measure to capture part of that effect, but not all of it.

increase. The relative impact on raw materials, work-in-process, and finished goods is an empirical question.

It is worth noting that there are many interesting business level factors that have changed over the period. For instance, due to trucking deregulation and the rise of FedEx, rapid delivery of goods became easier and cheaper. Due to computerization, record keeping functions likely improved. These and a host of other changes are all real. We have been unable to find good empirical measures for these changes that we could match with our panel for firms over the period. While our data permits us to ask whether inventories declined, we do not have the data needed to sort out alternative business level factors that have contributed to the decline we are measuring. Accordingly we make no claim about the relative importance of these factors.

4.3 Measuring Inventory and Data

There are a number of different inventory ratios that are frequently considered. The appropriate measure depends on the purpose. White, Sondhi and Fried (1994) provide a helpful treatment of the standard accounting ratios that we use.

From an operations management point of view we are most interested in how long inventory is held. It is important to have productive inputs available when needed. But, as stressed by the advocates of Just-In-Time, holding inventory takes up space and can permit slack attitudes to become pervasive with damaging effects overall.

Inventory days (ID) measures how many days on average it takes for the inventory to turn over. In year t, let firm i's inventory be I_{it} and let COGS_{it} denote the cost of goods sold. Then inventory days of firm i in year t is

$$\mathrm{ID}_{it} = \frac{\mathrm{I}_{it} \times 365 \text{ days}}{\mathrm{COGS}_{it}}.$$

A second popular measure is the inventory-to-sales ratio. If we use sales to replace COGS in the above ratio, then we have the inventory-to-sales ratio, which we denote as IS. The third measure that we studied is the inventory-to-assets ratio, which is denoted as IA. Let TA_{it} denote total assets of firm i in year t. The inventory-to-assets ratio is

$$\mathrm{IA}_{it} = \frac{\mathrm{I}_{it}}{\mathrm{TA}_{it}}.$$

This measurement examines the fraction of a firm's assets that are tied up in inventory. Asset ratios are particularly useful for making comparisons across years. This ratio automatically normalizes for firm size. We systematically studied IS, IA, and ID. Generally they provide confirmation of the same results. To save space we focus primarily on ID in the tables.

For each of the above inventory measures we can replace overall inventory with raw materials, work-in-process, or finished goods. The interpretations change in the obvious manner.

Different industries have different inventory needs. There are, of course, many possible ways to control for industry effects. We have tried a range of alternative methods but found little differences in the conclusions to be drawn. Accordingly, we settled on a particularly simple method. We take the normalized deviation from the industry norm as a measure of whether a particular firm has lean or bloated inventory. To be specific, let AbI_{it} denote abnormal inventory of firm i in year t,

$$\text{AbI}_{it} = \frac{\text{ID}_{it} - \text{mean inventory days of firm } i\text{'s 3-digit SIC industry in year } t}{\text{standard deviation of inventory days of firm } i\text{'s 3-digit SIC industry in year } t}.$$

An attractive feature of AbI is that it is unit free. The interpretation of AbI is quite simple. If $\text{AbI}_{it} > 0$ then in year t firm i is holding inventory longer than do other firms in the same industry. Firms with $\text{AbI}_{it} < 0$ hold their inventory for a shorter period of time than do their industry peers.[5]

This study is based on balance sheet data from 7433 U.S. manufacturing firms over a 20-year period. The data includes all publicly traded manufacturing firms. These are almost all of the manufacturers in the U.S. economy.

The firm specific data comes from the COMPUSTAT database available through WRDS (Wharton Research Data Services, University of Pennsylvania). The stock returns data is from the December 2002 edition of the CRSP (Center for Research in Security Prices, University of Chicago) database, and it is merged with the COMPUSTAT data using the CCM (CRSP/COMPUSTAT Merged) database also available from WRDS. Data on financial risk factors is from Ken French's web page[6].

We use data of U.S. firms whose COMPUSTAT incorporation codes are zero. We only include firms with SIC codes from 2000 to 3999 inclusive, i.e., manufacturing firms. Inventories either do

[5] Many firms actually span multiple industry segments, but only report firm-level data. We use the firms' primary SIC codes to identify the industries they belong to. So AbI is an approximation of a firm's deviation from the industry mean. Based on the COMPUSTAT Business Segment data, we repeated the same portfolio analysis with all multi-segment firms removed (about half of the data is dropped). The results are consistent with those reported in Section 4.5.

[6] See http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/.

not exist, or have rather different interpretations in other sectors of the economy. When available, we replace SIC codes with historical SIC codes. Both SIC and historical SIC codes are from COMPUSTAT. When historical data is unavailable we approximate it. For example, if a firm exists over the 20-year period and its historical SIC is reported only for the period 1987-1995, then we replace SIC with historical SIC for 1987-1995, make no change for 1996-2000, and replace SIC with historical SIC in 1987 for the period 1981-1986. For the analysis of financial impact of inventories, we use CRSP share codes 10 and 11 (i.e., ordinary common shares), provided that they are listed as exchanges codes 1, 2 and 3 (i.e., NYSE, AMEX and NASDAQ, respectively).

We take the COMPUSTAT identifier GVKEY as our empirical definition of a firm. A handful of stocks are matched to more than one GVKEY. Removing or keeping such duplications has almost no effect on the result.[7]

COMPUSTAT has several missing value codes. We replace the code "insignificant figure" with zero, while all other codes are replaced by missing values. Since companies may restate data for accounting changes, we use the restated data when available. Data items that may be restated are: assets, sales and cost of goods sold.

When applying a log transformation we replace zero values with 0.001 to deal with the fact that the log of zero is not defined. If we replace zero with too small a number, then the log becomes a big negative number – an outlier will be created. We found that 0.001 is the smallest positive data that COMPUSTAT reports for inventory, so we use it to replace the zero values. We experimented using slightly smaller values, and found that the conclusions are not affected.

Outlier observations can cause problems, particularly when taking ratios. Accordingly, we follow the common procedure of winsorizing the data. For inventory, costs of goods sold, sales and assets, we replace the top 1% of the data by the highest value that is not removed. It is important to down-weight the extreme tails, but exactly where that cut-off is defined does not make much difference. We experimented with slightly lower and slightly higher cut-offs, and compared the results. The exact location of the winsorization does not affect the conclusions.

[7]To be specific there are 37 PERMNO (the CRSP identifier) matched to more than one GVKEY. In the reported results we resolved the duplications based on the following rules. We remove 26 GVKEY companies who report pre-FASB data. (FASB is the Financial Accounting Standards Board. For its history see http://www.fasb.org/facts/.) We remove 9 GVKEY companies whose existence periods overlap with the existence periods of their duplicates. We remove two GVKEY companies whose data do not agree with their duplicates, and whose existence periods are shorter than their duplicates.

Winsorization is a standard method to deal with outliers. But what about observations that are measured with error but are not necessarily outliers? In a regression, what matters are additive errors in the explanatory variables. Time itself is not error prone, so it might appear that all is well. However, if there are systematic patterns in the variable being explained that are not additive, there can still be a problem.

The potential concern is that accounting rules give the accountants some discretion. What is worse, sometimes accountants are not honest altogether. Some accountants might be trying to help their clients 'look good.' In the early 1980s, looking good generally meant having a low inventory due to the popularity of JIT. In the late 1990s it was quite different. Consider the famous accounting scandals associated with Arthur Andersen and its clients. Some firms may have been concerned about appearing over-levered. Inventory typically serves as collateral for debt. From this perspective having a higher inventory might help the firm look good to investors who were worried about being repaid. To the extent that this might have affected the accounting in the late 1990s, it could cause firms to exaggerate the size of their inventory.

Our best guess is that the bias in accounting is not all that large. But there is no easy way to measure its impact. To the extent that the accounting bias matters, it will cause us to underestimate the rate at which inventory declines.

For the analysis of inventory decline, over the full period, there are more than 61,000 firm-years for which we have information on total inventories. More than 3/4 of our sample firms provide a breakdown into raw materials, work-in-process, and finished goods inventory. For the analysis of financial impact of inventories, over 41,000 firm-years of information are available over the 20-year period.

The macroeconomic control factors that we include are conventional.[8]

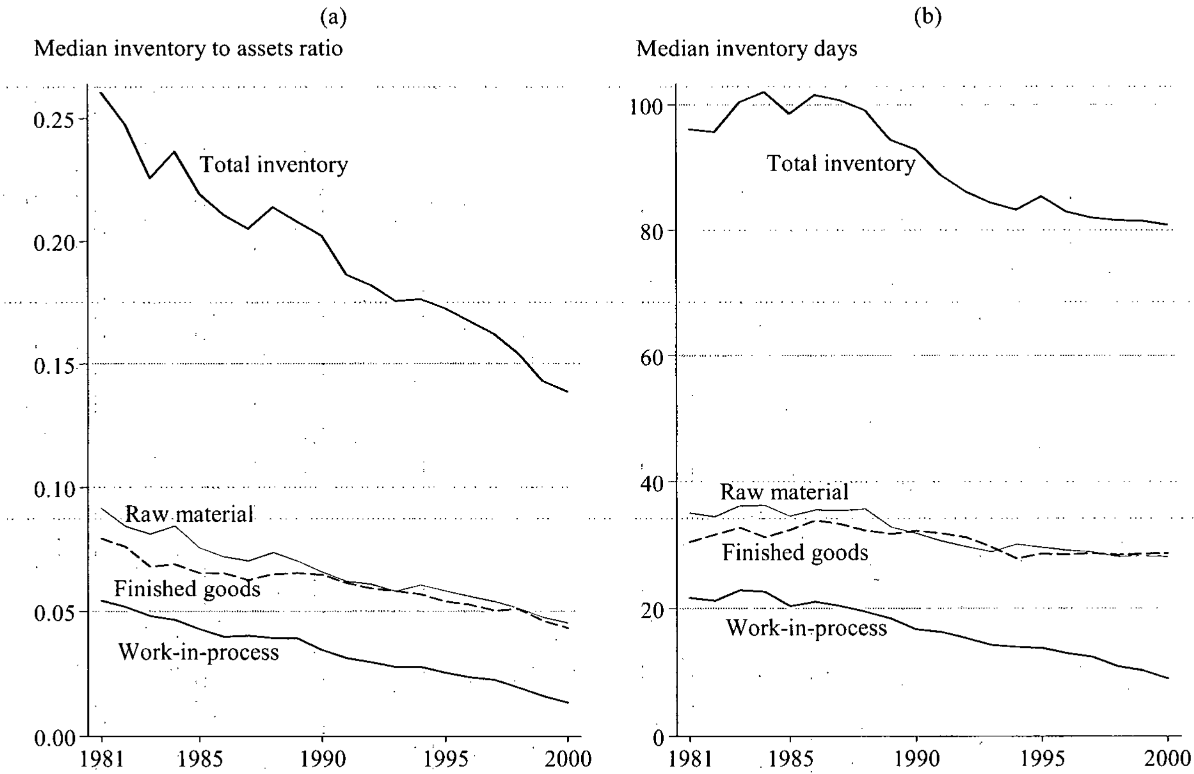
- R_f is the going interest rate. It is from H.15 Release – Federal Reserve Board of Governors.
- GGDP (Growth in Gross Domestic Product) is a macroeconomic growth rate. Let GDP be the real gross domestic product in 1996 dollars as reported by the Bureau of Economic Analysis at the U.S. Department of Commerce. Then $\mathrm{GGDP}_t = \ln(\mathrm{GDP}_t) - \ln(\mathrm{GDP}_{t-1})$.

[8]The original data are all available from the Federal Reserve Economic Data (FRED) at http://research.stlouisfed.org/fred2. The data series identifiers are MDISCRT, GDPCA, PPIACO and NAPM, respectively.

- Infl is the inflation rate. Let PPI be the "Producer Price Index: All Commodities" as reported by the Bureau of Labor Statistics at the U.S. Department of Labor. Then $\text{Infl}_t = \ln(\text{PPI}_t) - \ln(\text{PPI}_{t-1})$.

- PMI (Purchasing Managers Index) is a survey measure of the optimism of corporate purchasing managers. A PMI reading above/below 50 percent indicates that in the opinion of the purchasing managers who were surveyed, the economy is generally expanding/declining. This information is from the Institute for Supply Management.

Figure 4.1: Median inventory measures of U.S. manufacturers



There are 7431 firms providing 62218 observations of inventory to assets ratios, and 7295 firms providing 61038 observations of inventory days. Not all firms report their inventory components. As a result, the numbers of firms and observations for the inventory components are slightly less.

4.4 Have Inventories Been Falling?

4.4.1 Basic Inventory Time Trend

Table 4.1 provides basic descriptive statistics. The drop in inventory-to-assets between 1981 and 2000 was dramatic. The declines in inventory days are less dramatic. While the medians of raw material, finished goods, and total inventory days drop, the means actually rise between 1981 and 2000. We focus on the medians rather than the means due to the familiar concern that the means may be influenced by outliers. The descriptive statistics show that decomposing inventory into

Table 4.1: Descriptive statistics

(a) Inventory to assets ratio (×100%)

Year	Inventory type	Num of obs.	Mean	25th percentile	Median	75th percentile
1981	Raw material	2068	10.83	5.57	9.16	14.27
	Work-in-process	1835	7.45	1.90	5.43	10.43
	Finished goods	1980	9.96	2.97	7.95	14.08
	Total	2645	26.74	17.13	26.06	35.77
2000	Raw material	2669	6.85	1.46	4.54	9.21
	Work-in-process	2580	3.25	0.00	1.35	4.06
	Finished goods	2644	7.11	1.14	4.34	9.79
	Total	3209	16.51	5.90	13.86	23.57
1981~2000	Raw material	51400	8.58	3.09	6.55	11.59
	Work-in-process	48175	5.15	0.57	3.09	7.09
	Finished goods	50522	8.38	1.97	5.92	11.66
	Total	62218	21.24	10.50	19.14	29.63

(b) Inventory days

Year	Inventory type	Num of obs.	Mean	25th percentile	Median	75th percentile
1981	Raw material	2054	52.6	20.2	35.1	60.7
	Work-in-process	1821	37.8	7.3	21.7	45.8
	Finished goods	1965	44.8	11.6	30.5	57.5
	Total	2624	133.9	59.4	96.1	149.6
2000	Raw material	2590	53.7	11.2	28.2	57.1
	Work-in-process	2501	25.0	0.0	9.1	26.9
	Finished goods	2566	66.7	9.0	28.8	59.0
	Total	3117	154.1	41.2	80.8	137.1
1981~2000	Raw material	50363	61.4	16.4	31.8	57.9
	Work-in-process	47152	54.5	3.4	15.9	37.3
	Finished goods	49499	59.2	11.3	30.7	59.6
	Total	61038	165.1	52.1	90.1	144.0

(c) Macroeconomic factors 1980~1999

Macroeconomic factors	Mean	Standard deviation	Min	Max
Interest rate (R_f)	6.67	2.83	3.00	13.42
Macroeconomic growth rate (GGDP)	3.21	1.81	-1.95	6.94
Inflation rate (Infl)	2.05	3.41	-3.37	11.10
Purchasing managers index (PMI)	51.15	5.18	38.48	59.30

stages may be important.

In 1981 the median inventory days was 96.1. By 2000 it had fallen to 80.8. By far the biggest drop among the components was observed for the work-in-process inventory, which dropped from 21.7 to 9.1 days. Raw material dropped from 35.1 to 28.2 days, while finished goods only dropped from 30.5 to 28.8 days.

Figure 4.1 illustrates the same trend. Figure 4.1(a) depicts the inventory-to-assets ratio. This drop reflects both changes in inventory policy as well as changes in the holdings of other types of assets by firms. Figure 4.1(b) depicts the inventory days, a measure that is immune to changes in the holdings of other assets. In each case the big picture remains the same – inventories dropped over the period.

It is worth noting that the decline is not caused by shift of inventories from public firms to private firms. We compared the inventory and assets positions of the entire U.S. economy as reported in the U.S. Flow of Funds Accounts[9] to the positions of the firms in COMPUSTAT. There is no evidence that such a shift accounts for our results. In fact, a greater fraction of the economy-wide U.S. inventory holdings was in public firms' hands by the end of the period than at the start. In 1981, the ratio of publicly traded firms' inventory to the total Flow of Funds inventory was 78.4%. By the year 2000, the ratio became 93.2%. The correlation between the public firm numbers and Flow of Funds numbers are very high. For inventory levels, the correlation is 0.995, and for change in inventory, the correlation is 0.742. The difference is due to the increasing number of publicly traded firms over the period.

The magnitudes of the inventory decline differs across the stages in the inventory cycle. Figure 4.1 shows that there was a very large drop for work-in-process inventory, but milder declines for raw materials and finished goods. The decreases are statistically significant in all cases, except for the finished goods inventory days. Much of the decline in inventory days took place during the period 1987 to about 1995.

4.4.2 Controlling For Other Factors

The descriptive evidence is striking, but it does not control for firm heterogeneity, macroeconomic conditions, or changes in industry composition. To address these concerns requires a statistical

[9]See http://www.federalreserve.gov/releases/.

model.

Our data takes the form of a panel with a great many firms and a much smaller number of years. With panel data, there are many models that permit various types of time trends. The models that we consider differ in how much similarity is assumed among firms and among industries. We report results from estimating the following models: I. a random effects model, II. a mixed effects model, III. a fixed effects model.[10] For the reported estimates we are explaining inventory days (ID). These models have also been estimated for the inventory-to-sales ratio and the inventory-to-assets ratio. Since the inferences are essentially the same, they are not reported.

Model I is a random effects model given by

$$\log(\mathrm{ID}_{it}) = a + u_i + (b + v_i)(t - 1981) + \varepsilon_{it}, \tag{4.1}$$

where u_i and v_i are random intercept and slope with zero means, and ε_{it} has zero mean conditional on u_i and v_i. To estimate this model using the maximum likelihood method, we further assume that u_i and v_i are jointly normally distributed and ε_{it} is normally distributed.

Let k index the macro factors with the coefficients denoted m_k and the macro factors F_k, where $F_k \in \{\mathrm{R}_f, \mathrm{GGDP}, \mathrm{Infl}, \mathrm{PMI}\}$. Model II is a mixed effects model given by

$$\log(\mathrm{ID}_{it}) = a + u_i + (b + v_i)(t - 1981) + \sum_k m_k F_{k,t-1} + \varepsilon_{it}, \tag{4.2}$$

where the same assumptions about the random effects and the errors are made as model I. To ensure that the macroeconomic factors are predetermined, one year lag is used for the macroeconomic factors.

Model III is a fixed effects model given by

$$\log(\mathrm{ID}_{it}) = a + u_i + b(t - 1981) + \sum_k m_k F_{k,t-1} + \varepsilon_{it}, \tag{4.3}$$

where u_i is the firm fixed effect contrasting to model I and II, and ε_{it} has zero mean conditional on u_i. We use Huber-White robust standard errors in this model.

Each of these models has advantages and disadvantages. Model I provides a good basic reflection of the time trends, but it leaves open the issue of the macro factors. A model with full random

[10] We also estimated individual firm regressions, but this results in a vast number of parameters, and given that many firms have only a few years of data, the parameter estimates are not as reliable. They do not alter the main conclusions, but they add a lot of noise. Accordingly we prefer to impose more structure – as in the reported models. We have considered a variety of alternative models which permit heteroscedasticity and autocorrelation in the errors. Since these alternative specifications do not change our inferences, we do not report the results separately.

effects and macro factors could in principle be estimated. But when we did this, the maximum likelihood method failed to converge. Hence, we use the simpler form of mixed effects model as our model II.

Model III is a conventional fixed effects model with firm-specific intercepts. This type of model is very frequently used in econometric practice. This model has both advantages and disadvantages relative to the earlier two models. A Hausman test (see Greene 2003 for a discussion) favors a fixed effects specification over a random effects specification.[11] This argues in favor of model III. The results from Models I and II suggest that the variance of intercepts is much larger than the slope, which argues for a more elaborate model than model III to reflect the differing slopes. In principle we can allow for both firm-specific intercepts and slopes. But then we are back to the problem of having far too many parameters to estimate and interpret.

In our judgement, all of these considerations are pertinent. Reasonable people can disagree on how heavily to weigh each consideration. As a result, we report models I-III and we focus on results that are consistent across model specifications.

In Table 4.2 we see that all of the models give very similar parameter estimates for the key parameters of interest (i.e., intercept and time trend). The macro control factors also generally have similar effects across specifications. This suggests that the effects that interest us are not very sensitive to the choice among these three types of models.

In model I, the fitted inventory days can be expressed as $\mathrm{ID}_t = e^{a+b(t-1981)}$. Thus, the fitted inventory days in 1981 is e^a, and the yearly percentage change of the fitted inventory days is $(\mathrm{ID}_{t+1} - \mathrm{ID}_t)/\mathrm{ID}_t = 100(e^b - 1)\% \approx 100\,b\%$. The values of the time trends reported in Table 4.2 can be interpreted as the yearly percentage change of inventory days.

Model I estimates that the total inventory days in 1981 was $e^{4.24} \approx 70$ days and declined about 2% per year; raw material was $e^{3.24} \approx 26$ days in 1981 and declined about 3% per year; work-in-process was $e^{2.18} \approx 9$ days in 1981 and declined about 6% per year; finished goods was $e^{2.55} \approx 13$ days in 1981 and has no significant decline or increase.

[11] Baltagi (2001) thoroughly discusses the fact that more than just a Hausman test is needed before one can select a 'correct' model. Rather than make a definitive choice among the specifications, we show that the basic inventory trends are the same under these alternative frameworks.

Table 4.2: Inventory days of all U.S. manufacturers 1981 – 2000

	Raw material			Work-in-process			Finished goods			Total		
Model	I	II	III	I	II	III	I	II	III	I	II	III
Intercept	3.24 ***	3.18 ***	3.14 ***	2.18 ***	2.47 ***	2.58 ***	2.55 ***	2.42 ***	2.42 ***	4.24 ***	4.24 ***	4.22 ***
	(96.0)	(45.4)	(42.6)	(44.2)	(25.7)	(25.2)	(60.8)	(27.0)	(25.3)	(152)	(70.5)	(68.2)
Time trend	-3.47 ***	-3.65 ***	-1.83 ***	-6.11 ***	-6.98 ***	-5.25 ***	-0.07	-0.17	0.84 ***	-1.97 ***	-2.23 ***	-1.02 ***
	(-12.9)	(-12.7)	(-10.5)	(-17.2)	(-18.5)	(-21.9)	(-0.23)	(-0.50)	(3.93)	(-8.91)	(-9.35)	(-7.17)
Standard deviation of random intercept u	2.20	2.20		3.19	3.19		2.72	2.72		1.95	1.96	
Standard deviation of random time trend v	0.17	0.17		0.21	0.21		0.18	0.18		0.15	0.15	
Correlation of u and v	-0.59	-0.59		-0.57	-0.57		-0.59	-0.59		-0.58	-0.58	
Coef. of R_f		-0.19	0.38		-2.49 ***	-2.52 ***		0.34	1.23 ***		-0.44	-0.02
		(-0.58)	(1.06)		(-5.83)	(-5.09)		(0.84)	(2.68)		(-1.61)	(-0.06)
Coef. of GGDP		0.65 ***	0.02		-0.46	-0.36 *		0.08	-0.44 **		0.22	-0.15
		(2.76)	(0.12)		(-1.46)	(-1.80)		(0.26)	(-2.33)		(1.09)	(-1.27)
Coef. of Infl		0.01	0.84 ***		-0.30 *	-0.47		-0.52 ***	0.37		-0.17	0.38 *
		(0.08)	(3.15)		(-1.84)	(-1.23)		(-3.36)	(1.03)		(-1.60)	(1.71)
Coef. of PMI		0.12	0.14		-0.08	-0.12		0.24 **	0.28 **		0.09	0.10
		(1.40)	(1.42)		(-0.68)	(-0.86)		(2.12)	(2.08)		(1.16)	(1.18)
R^2			0.81			0.86			0.81			0.80
Log likelihood	-70931.6	-70921.7		-79269.6	-79247.5		-81108.1	-81100.4		-82610.8	-82602.2	
Number of firms		6348			6098			6306			7295	
Number of obs.		50363			47152			49499			61038	

Model I is a random intercept and time trend model. Model II is a mixed effects model with random intercept, random time trend and fixed macro factors effects. Model III has fixed firm-specific intercepts, fixed time trend and fixed macro factors effects. The models are described in equation (1), (2) and (3), respectively. Intercept and time trend correspond to a and b in equation (1), (2) and (3).

Model I and II are estimated using the maximum likelihood method, and the log likelihood is reported. To estimate model III, we create firm dummies and apply OLS with Huber-White robust estimator of standard error (statistical packages usually offer commands to handle fixed-effects regression automatically so that creating dummies is unnecessary).

Time trends and the coefficients of macroeconomic factors are reported 100 times larger than their original values. *, ** and *** denote statistical significance at 10, 5 and 1 percent level, respectively. t-statistics are reported in brackets.

Consider the results for total inventory days, raw material and work-in-process in Table 4.2. We see that in all models, the coefficient on time is negative and statistically significant. It is the most negative for work-in-process. Total inventories declined under all models and work-in-process declined most significantly. This is a very robust result.

Raw materials, work-in-process, and finished goods inventories play quite different roles in a firm's operations. Raw materials relate to the firm's interactions with suppliers. Work-in-process reflects the efficiency of the firm's own operations. Finished goods relate to the firm's interactions with customers.

There is very strong evidence that the manufacturing firms we study improved their interactions with suppliers and their own internal operations. However, there is no corresponding drop in finished goods inventory. Indeed, if one prefers the firm fixed effects model (model III), there is even some evidence that finished goods may have increased.

It is worth noting that there is a strong evidence that product variety has increased dramatically, see Fisher, Hammond, Obermeyer and Raman (1994). The increased variety leads to increased demand variability. At the same time, manufacturing firms might have focused more on improving customer service levels through product availability. Both changes may have contributed to the need to increase finished goods inventory, and may have cancelled the JIT efforts in reducing finished goods inventory.

The macro factors generally perform sensibly. Interest rates have a negative effect on work-in-process inventory holdings. Macroeconomic growth has a somewhat positive effect on raw materials and a somewhat negative effect elsewhere. But these effects are not all that robust to alternative specifications. Inflation seems to be associated with an increase in raw materials inventory and a drop in finished goods inventory. This makes some sense if firms recognize the inflation and are adjusting to it. When the purchasing managers expectations are good, there is an increase in the finished goods inventory. Presumably this reflects the firm's preparation for the expected strong demand.

Next, we decompose the effects into 25 individual industries. We use mostly 2-digit SIC industries. For some industries there are a sufficient number of firms to permit further subdivision into 3-digit SIC industries. There are also seemingly problematic cases that we adjusted.[12] Figure 4.2

[12] The chemical industry (SIC 28) is divided into non-drug and drug industry. Machinery and computer industry (SIC 35) is divided into two. Electronic and electrical equipment industry (SIC 36) is divided into two. Transportation

shows the median inventory days over the years for each industry.

Figure 4.2: Median inventory days of U.S. manufacturing industries

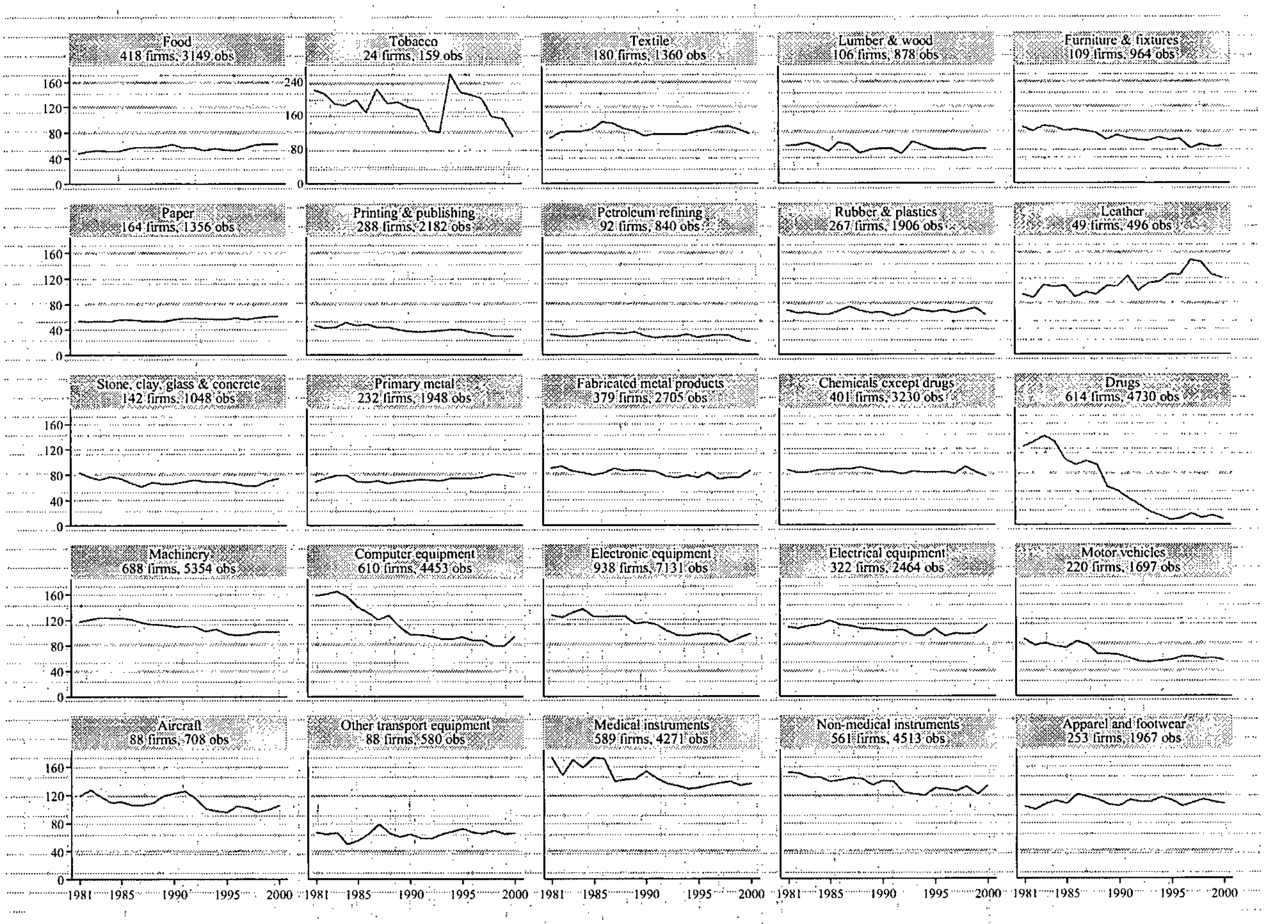


Table 4.3 decomposes models I-III into individual industries. To save space we do not report coefficients on the macro factors. The decline in total inventory days is not limited to just one or two industries. Rather, it is a pervasive phenomenon observed in many industries. There are 8 out of 25 industries whose decline in total inventory days is significant at conventional significance levels, under any model specification. No industry exhibits an increasing inventory trend that is robust across model specifications.

equipment industry (SIC 37) is divided into motor vehicle, aircraft and others. Lab instruments industry (SIC 38) is divided into medical and non-medical instruments. Apparel and footwear industry includes knitting (SIC 225), cutting and sewing (SIC 23 except 239), rubber and plastics footwear (SIC 302), leather footwear and gloves (SIC 313-315), costume jewelry, novelties and buttons (SIC 396). Textile industry includes SIC 22 and 239.

Table 4.3: Inventory days of U.S. manufacturing industries 1981–2000

Industry	Total inventory days							Finished goods inventory days						
	Num of firms	I		II		III		Num of firms	I		II		III	
	Num of obs.	Intercept	Time trend	Intercept	Time trend	Intercept	Time trend	Num of obs.	Intercept	Time trend	Intercept	Time trend	Intercept	Time trend
Food	418	3.81 ***	1.15 **	3.95 ***	0.49	3.90 ***	0.15	303	2.88 ***	2.42 ***	3.18 ***	1.22	3.14 ***	1.15**
	3149	(53.5)	(2.22)	(21.6)	(0.83)	(20.9)	(0.40)	2072	(28.3)	(3.17)	(12.8)	(1.45)	(12.3)	(2.36)
Tobacco	24	5.27 ***	-1.30	4.19 ***	0.54	4.26 ***	1.20	20	2.34 ***	9.12 **	1.62 ***	9.36 **	2.36 ***	3.29*
	159	(43.6)	(-1.01)	(7.8)	(0.32)	(8.4)	(0.98)	114	(2.9)	(2.09)	(1.4)	(2.05)	(3.0)	(1.83)
Textile	180	4.37 ***	-0.35	4.36 ***	-1.44	4.30 ***	-0.01	169	3.32 ***	1.58 *	3.55 ***	0.03	3.52 ***	0.45
	1360	(65.6)	(-0.37)	(16.9)	(-1.28)	(21.1)	(-0.03)	1223	(33.8)	(1.86)	(10.4)	(0.03)	(11.9)	(0.82)
Lumber and wood	106	4.29 ***	-2.65	4.35 ***	-3.30	4.29 ***	-2.94**	86	2.45 ***	-1.23	1.90 ***	0.54	1.86 ***	-0.37
	878	(15.5)	(-1.15)	(11.4)	(-1.39)	(12.3)	(-2.52)	553	(8.9)	(-0.46)	(2.2)	(0.17)	(2.0)	(-0.13)
Furniture and fixtures	109	4.46 ***	-2.25 ***	4.71 ***	-2.77 ***	4.74 ***	-2.67***	96	2.85 ***	0.86	2.96 ***	0.73	3.25 ***	-1.14
	964	(62.2)	(-4.69)	(32.5)	(-5.12)	(30.8)	(-7.02)	875	(12.6)	(0.67)	(7.8)	(0.52)	(7.9)	(-1.11)
Paper	164	4.05 ***	0.00	4.36 ***	-0.81	4.36 ***	-0.82*	130	2.72 ***	2.79	2.95 ***	1.88	3.02 ***	1.18
	1356	(45.9)	(0.00)	(16.8)	(-0.80)	(16.9)	(-1.67)	809	(8.7)	(1.39)	(5.5)	(0.88)	(6.5)	(1.09)
Printing and publishing	288	3.49 ***	-4.88 **	3.75 ***	-5.41 **	4.11 ***	-4.32***	197	0.68 ***	-0.49	2.40 ***	-3.43	2.55 ***	1.21
	2182	(14.6)	(-2.29)	(12.1)	(-2.50)	(15.3)	(-8.24)	1454	(1.8)	(-0.18)	(3.7)	(-1.17)	(3.7)	(0.88)
Petroleum refining	92	3.51 ***	-1.08	5.02 ***	-4.90 ***	4.54 ***	-3.99***	29	1.13 ***	5.24 ***	2.31 ***	1.79	3.37 ***	-0.49
	840	(21.3)	(-0.68)	(11.1)	(-2.73)	(14.3)	(-5.16)	166	(1.3)	(2.64)	(2.0)	(0.68)	(4.6)	(-0.29)
Rubber and plastics	267	4.19 ***	-0.35	4.60 ***	-1.49 *	4.64 ***	-1.03**	236	3.01 ***	0.97	3.52 ***	0.13	3.37 ***	0.53
	1906	(49.9)	(-0.43)	(20.3)	(-1.65)	(19.9)	(-2.03)	1632	(22.3)	(0.75)	(10.3)	(0.09)	(9.2)	(0.81)
Leather	49	4.53 ***	1.00	4.16 ***	1.42	4.42 ***	1.09*	42	3.88 ***	2.40 **	2.83 ***	4.48 ***	2.99 ***	3.58***
	496	(37.2)	(1.32)	(11.4)	(1.47)	(18.0)	(1.80)	390	(20.8)	(2.29)	(5.9)	(3.36)	(7.2)	(3.69)
Stone, clay, glass and concrete	142	4.27 ***	-0.64	4.35 ***	-0.43	4.23 ***	-0.12	113	3.22 ***	-2.07	3.17 ***	-1.20	3.03 ***	0.98
	1048	(47.5)	(-0.76)	(14.6)	(-0.42)	(21.7)	(-0.25)	788	(18.7)	(-1.57)	(6.6)	(-0.74)	(6.3)	(0.84)
Primary metal	232	4.30 ***	-0.93 *	4.16 ***	-0.45	3.85 ***	0.05	172	2.70 ***	-0.09	2.24 ***	-0.41	2.05 ***	0.38
	1948	(56.1)	(-1.70)	(18.1)	(-0.70)	(17.6)	(0.10)	1271	(8.6)	(-0.04)	(4.5)	(-0.18)	(3.9)	(0.40)
Fabricated metal products	379	4.38 ***	-0.13	4.79 ***	-0.56	4.67 ***	-0.97*	321	2.62 ***	0.16	2.17 ***	1.18	2.34 ***	-1.55
	2705	(76.7)	(-0.22)	(23.5)	(-0.77)	(22.0)	(-1.93)	2086	(15.2)	(0.11)	(5.2)	(0.71)	(5.1)	(-1.38)
Chemicals except drugs	401	4.32 ***	-0.47	4.23 ***	-0.57	4.14 ***	1.26***	347	3.35 ***	0.81	3.33 ***	0.81	3.44 ***	1.80***
	3230	(53.5)	(-0.87)	(20.1)	(-0.90)	(19.7)	(2.87)	2596	(34.3)	(1.51)	(12.0)	(1.18)	(13.2)	(3.26)
Drugs	614	1.35 ***	-1.06	1.14 ***	-0.77	1.82 ***	1.75*	568	0.69 ***	-3.25 **	0.30 ***	-3.02 *	0.95 ***	1.49
	4730	(5.7)	(-0.68)	(2.5)	(-0.48)	(3.7)	(1.74)	4087	(3.0)	(-2.05)	(0.7)	(-1.86)	(2.0)	(1.51)
Machinery	688	4.60 ***	-1.47 **	4.57 ***	-1.17 *	4.54 ***	-1.09***	602	2.72 ***	-1.11	2.86 ***	-1.51	2.57 ***	0.18
	5354	(68.2)	(-2.49)	(29.4)	(-1.81)	(29.7)	(-3.43)	4370	(20.6)	(-1.23)	(10.1)	(-1.50)	(8.9)	(0.26)
Computer equipment	610	5.04 ***	-5.33 ***	5.42 ***	-6.54 ***	5.16 ***	-6.69***	553	2.77 ***	0.82	3.15 ***	-1.10	3.07 ***	-2.69***
	4453	(76.6)	(-8.33)	(24.7)	(-8.79)	(21.3)	(-10.21)	4000	(22.1)	(0.80)	(9.4)	(-0.95)	(8.3)	(-3.08)
Electronic equipment	938	4.74 ***	-3.05 ***	4.73 ***	-3.42 ***	4.75 ***	-1.61***	861	2.23 ***	0.53	1.70 ***	1.19	1.62 ***	2.97***
	7131	(77.4)	(-5.97)	(29.4)	(-6.12)	(29.3)	(-4.60)	6239	(18.0)	(0.63)	(5.9)	(1.29)	(5.2)	(4.36)
Electrical equipment	322	4.51 ***	-0.09	4.44 ***	-0.63	4.59 ***	-0.77	293	2.74 ***	0.61	3.21 ***	-0.73	3.46 ***	-1.01
	2464	(42.3)	(-0.10)	(16.8)	(-0.66)	(16.4)	(-1.58)	2139	(16.8)	(0.44)	(8.5)	(-0.48)	(9.1)	(-1.19)
Motor vehicles	220	4.43 ***	-1.85 **	4.26 ***	-2.05 **	4.10 ***	-1.82***	188	2.83 ***	0.17	2.85 ***	0.08	2.49 ***	1.32
	1697	(42.0)	(-2.38)	(16.9)	(-2.38)	(16.9)	(-3.33)	1382	(17.0)	(0.14)	(6.8)	(0.06)	(5.7)	(1.31)
Aircraft	88	4.79 ***	-2.11 *	5.07 ***	-2.51 *	4.97 ***	-2.06***	67	1.19 ***	-0.28	-0.61 ***	4.45	-0.75 ***	2.60
	708	(71.9)	(-1.84)	(16.8)	(-1.89)	(17.9)	(-3.30)	512	(2.4)	(-0.08)	(-0.5)	(1.05)	(-0.8)	(0.95)
Other transportation equipment	88	4.10 ***	0.65	4.46 ***	-0.21	4.42 ***	-1.13*	72	1.21 ***	4.10	0.69 ***	6.69	0.07 ***	2.75
	580	(34.9)	(0.52)	(13.4)	(-0.16)	(13.3)	(-1.85)	454	(1.7)	(0.82)	(0.6)	(1.28)	(0.0)	(0.89)
Medical instruments	589	4.09 ***	2.67 **	3.99 ***	2.54 **	4.00 ***	1.06	545	2.72 ***	3.64 ***	2.13 ***	4.09 ***	2.07 ***	3.51***
	4271	(24.8)	(2.52)	(12.9)	(2.30)	(12.8)	(1.51)	3800	(15.3)	(3.07)	(5.6)	(3.26)	(5.2)	(3.91)
Non-medical instruments	561	4.86 ***	-2.09 ***	4.69 ***	-1.90 ***	4.54 ***	-1.48***	515	2.53 ***	0.80	1.81 ***	2.11 *	1.75 ***	0.53
	4513	(79.8)	(-3.36)	(25.9)	(-2.70)	(25.7)	(-3.79)	3890	(16.9)	(0.72)	(5.8)	(1.73)	(5.0)	(0.64)
Apparel and footwear	253	4.67 ***	-0.37	4.82 ***	-1.57 **	4.78 ***	-1.50**	236	3.99 ***	0.86	3.96 ***	-0.15	3.91 ***	-0.22
	1967	(81.7)	(-0.65)	(16.9)	(-2.12)	(18.6)	(-2.56)	1756	(42.0)	(0.80)	(13.1)	(-0.13)	(13.7)	(-0.33)
Total	7295	4.24 ***	-1.97 ***	4.24 ***	-2.23 ***	4.22 ***	-1.02***	6306	2.55 ***	-0.07	2.42 ***	-0.17	2.42 ***	0.84***
	61038	(151.9)	(-8.91)	(70.5)	(-9.35)	(68.2)	(-7.17)	49499	(60.8)	(-0.23)	(27.0)	(-0.50)	(25.3)	(3.93)

Intercept and time trend correspond to a and b in equation (1), (2) and (3), respectively. Time trends are reported 100 times larger than their original values. *, ** and *** denote statistical significance at 10, 5 and 1 percent level, respectively. t-statistics are reported in brackets.

Focusing on the coefficients from model I for illustration, we find the most rapid decline was in computer equipment (5% per year), printing and publishing (5%), and electronic equipment (3%). These are significant under all specifications. Five industries have 1%-2% decline per year that is significant across all estimated models. These industries are furniture and fixtures, machinery, motor vehicles, aircraft, and non-medical instruments industries.

In Table 4.3, we report the finished goods results separately, but not the raw materials and work-in-process. We do this because the finished goods pattern differs from the total inventory results, while raw materials and work-in-process do not differ significantly from the total.

There are 14 out of 25 industries that exhibit no significant finished goods time trend under any model specification. No industry exhibits a declining finished goods time trend that is robust across the specifications. Robust increasing trends are found in tobacco (10% per year), medical instruments (4%), and leather (2%) industries. Six industries have less robust evidence of finished goods inventory increases, while two industries (drugs and computer equipment) have less robust evidence of decline.

For work-in-process, there are 15 industries that exhibit a declining trend which is robust across model specifications. The most rapid declines are found in computer equipment (13% per year), leather (12%), and apparel and footwear (10%) industries. For raw material, there are 12 industries that exhibit robust decline. The three fastest are leather (10%), printing and publishing (8%), and computer equipment (7%).

As in any estimation, there are occasional anomalies observed. The drug industry is the only industry that has a serious discrepancy between median measure and the panel data models. The drug industry shows almost no change in its total inventory days under models I-III, but the median inventory days declined dramatically. This is because there was a great number of new firms that entered. There were only 74 firms in 1981, and almost 400 firms in later 90s. There were 540 new firms entering the industry from 1982-2000, but 304 of them had zero inventory in the entering year. These firms typically entered with low or zero inventory, which brought down the median, but that had little effect on the time trend in the panel data model. Many of these firms were essentially publicly traded research projects.

Table 4.3 also serves to reinforce the fact that finished goods inventory performed quite differently from raw materials and work-in-process. A simple way to describe the evidence is to say that the manufacturing firms have reduced the inefficiencies in their interactions with their suppliers and in their own internal operations. At the same time they have become more customer focused in that they have more finished goods ready for delivery, or they have to keep more finished goods due to increased product variety that implies the increased demand variability.[13]

4.5 Financial Impact of Inventories

4.5.1 Are Low Inventory Firms More Highly Valued in the Cross-Section?

A critical argument on behalf of inventory reduction is the claim that it will improve the financial position of firms. If this claim is true, then the market should value firms that have already reduced their inventories more highly than they value firms that have not reduced their inventories. Is this argument empirically valid? A common way to answer this type of question is to ask whether the factor of interest is associated with the market-to-book ratio or Tobin's q, which is defined as: (the market value of equity + book value of debt) / book value of total assets. Both lead us to the same inferences about the market valuation of inventory.

This is tested with a simple regression,

$$\text{Tobin's } q \quad = \quad a \; + \; b \,\text{AbI} \; + \; \varepsilon. \tag{4.4}$$

The result from (4.4) is Tobin's $q = 2.156 - 0.0558$ AbI. The t-statistics are 63.73 on the intercept and -1.61 (i.e., insignificant) on the slope. Adding the macroeconomic factors as regressors has almost no effect on the slope. Use of more complex functional form specifications and lagged specifications leads to the same basic conclusion. In this type of test there is no evidence of a significant impact of inventories on Tobin's q. Replacing Tobin's q by the market-to-book ratio (as defined in Fama and French 1993) does not change the conclusion.

[13]Ideally we would have liked to identify the deeper factors that permitted these trends to occur. Many possible factors could be at work, such as increased computerization, better delivery systems due to trucking deregulation and the rise of FedEx, improved scheduling software, an increase in the number of products produced by each firm, better understanding of potential drawbacks to holding inventory, etc. There are many such plausible factors. We have been unable to find reasonable empirical measures of these factors. Accordingly we are not in a position to judge the relative importance of each of these plausible factors.

4.5.2 Inventories and Longer Term Stock Returns

The cross-sectional analysis raises the possibility that the markets are not concerned about inventories. If that is correct, then firms with abnormally large inventories should have just as strong long term stock market performance as do other firms. To study this question we follow the popular methodology developed by Jegadeesh and Titman (1993). The method has become common in finance since Jegadeesh and Titman (1993) used it to study stock market momentum. A particularly nicely presented example of the method can be found in Gompers, Ishii and Metrick (2003), in which the same method is used to study the effects of corporate governance.

We start by sorting the firms into deciles according to their AbI in 1980. At the beginning of 1981, we invest \$1 in each AbI decile portfolio. The money is invested equally in all stocks in each decile. At the start of 1982 we sell all the stocks, and re-sort all firms according to their AbI in 1981. Then, all the money that came from decile 1 is reinvested equally in the current decile 1 stocks. The same type of reinvestment is done for each decile. This process is repeated year by year. In each year t we sell all of the stocks, and then re-sort the firms according to the AbI in year $t-1$. We take the money generated by decile i from year $t-1$ and reinvest it equally in the new decile i for year t. We repeat this procedure for each decile over the 20 years. We term the lowest AbI decile as decile 1, and the highest AbI decile is decile 10.

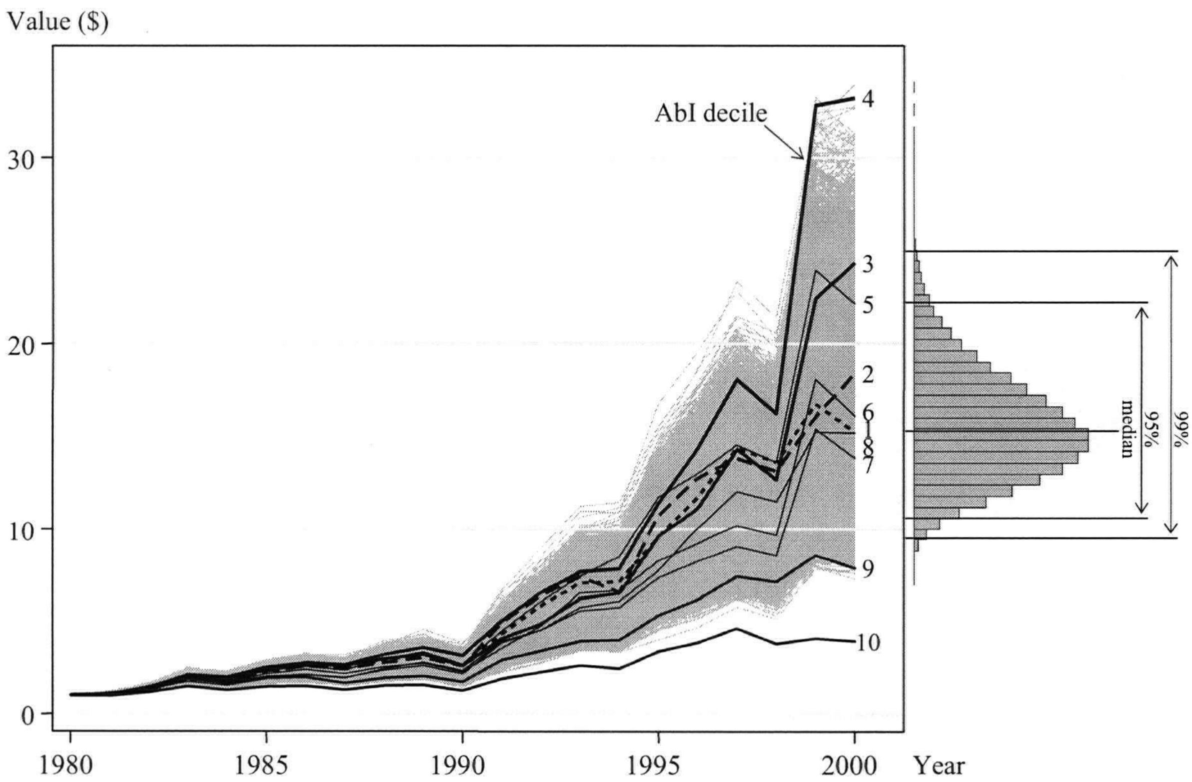
If low inventory is good, then the lowest AbI decile portfolio will have an abnormally high return. This is the prediction from romantic JIT. On the other hand, if romantic JIT is false, then it is also quite possible to find high returns in some other portfolios.

In order to decide if the returns are abnormal we need to determine the normal range. Suppose that AbI is really just noise that has nothing at all to do with stock returns. Then by chance it will sometimes happen to look as if it matters. But this will be rare. We mimic this process by using a random number generator to produce random portfolios. Having created a large number of such portfolios, we then see whether the observed returns on that AbI portfolios lie in the tails of the distribution.

To be specific, at the beginning of 1981, we randomly select 10% of the stocks, and invest \$1 equally in the selected stocks. In each of the following years, we take the money generated from previous years and reinvest it equally in a newly randomly selected portfolio in that year. We do this many times so that in the end we have created 100,000 of these random portfolios. If AbI is really significant then it should generate returns that are in the tails of the distribution created by the random portfolios. Empirically, each of the AbI decile portfolios and each random portfolios contains about 210 stocks on average. This changes over time, ranging from about 180 stocks in the early 1980s to about 250 stocks in the late 1990s.

The median final value for the 100,000 random portfolios over the full 20-year period is 15.29. We measure the variation using an empirical two-tailed p-values for each decile portfolio i. It is defined by $p_i = \min(n_i, N - n_i) \times 2/N$, where N = 100,000, and n_i is the number of random portfolios that have higher final values than portfolio i. With this definition, the 95% confidence interval is $[10.63, 22.21]$.

Figure 4.3: AbI portfolio returns and random portfolios



Each trajectory represents the value of a portfolio over time. All the portfolios start from $1 at the beginning of 1981. The AbI deciles portfolios are represented by black trajectories. The 100,000 random portfolios are shown in gray. The histogram on the right is the value distribution of the random portfolios at the end of year 2000. The median of 100,000 portfolio values is 15.29. The interval where 95% of the values lie is [10.63, 22.21]. The interval where 99% of the values lie is [9.52, 24.91].

The results are shown in Figure 4.3. Time is indicated along the horizontal axis, while portfolio values are indicated along the vertical axis. The distribution of portfolio values, year by year, are shown. On the right hand side of Figure 4.3 we plot the distribution of the final values for the 100,000 random portfolios being tracked.

Table 4.4: Final values of AbI decile portfolios

AbI decile portfolio	All firms			Firm size groups: 1981 – 2000		
	1981 – 2000	1981 – 1990	1991 – 2000	Small	Medium	Large
1	15.25 (0.991)	2.62 (0.189)	5.83 (0.352)	11.53 (0.390)	18.78 (0.205)	11.65 (0.535)
2	18.39 (0.330)	2.57 (0.247)	7.15 (0.695)	13.52 (0.639)	26.80 ** (0.013)	13.28 (0.993)
3	24.33 *** (0.014)	2.60 (0.211)	9.35 ** (0.035)	29.15 (0.166)	24.81 ** (0.027)	15.46 (0.477)
4	33.17 *** (0.000)	3.16 *** (0.003)	10.50 *** (0.005)	46.39 ** (0.013)	34.45 *** (0.001)	19.08 * (0.098)
5	22.13 * (0.052)	2.68 (0.126)	8.25 (0.193)	27.26 (0.221)	11.33 (0.655)	24.26 *** (0.007)
6	16.15 (0.774)	2.25 (0.941)	7.17 (0.680)	21.54 (0.510)	8.19 (0.115)	20.15 * (0.059)
7	13.88 (0.604)	2.32 (0.832)	5.98 (0.447)	10.43 (0.270)	15.24 (0.572)	14.02 (0.794)
8	15.23 (0.985)	2.16 (0.631)	7.07 (0.748)	34.74 * (0.073)	8.55 (0.155)	12.26 (0.708)
9	7.89 *** (0.000)	1.72 *** (0.008)	4.58 ** (0.012)	8.66 (0.120)	6.26 ** (0.012)	7.90 ** (0.011)
10	3.91 *** (0.000)	1.24 *** (0.000)	3.15 *** (0.000)	3.40 *** (0.000)	3.51 *** (0.000)	5.03 *** (0.000)
Median firm size (million dollars)	67.82	54.53	81.85	11.08	67.72	669.77
Number of obs.	41658	19494	22164	13878	13875	13905

*, ** and *** denote statistical significance at 10, 5 and 1 percent level, respectively.
p-values are reported in brackets. The p-value of portfolio i is given by $\min(n_i, N - n_i) \times 2 / N$, where N =100,000 and n_i is the number of random portfolios that have higher final values than portfolio i. It is a two-tailed p-value.

The final results are also reported numerically in Table 4.4. In addition to the overall results, we check for robustness to time period by providing results for the 1980s and the 1990s separately. Finally, Table 4.4 goes beyond Figure 4.3 by showing the effect conditioning on firm size.

Figure 4.3 and Table 4.4 show that high AbI (decile 9 and 10) is associated with unusually bad stock returns. This is true for the entire 1980-2000 period, and also for each of the decades considered separately. Deciles 9 and 10 have abnormally poor returns, while deciles 3 and 4 have abnormally high returns. Using continuously compounded returns, decile 4 has a return of 19.1% per year, which is 4.5% above the median portfolio.

Table 4.4 also reports portfolio results that are conditioned on firm size. The abnormally poor returns are found in decile 10 across all firm sizes, and in decile 9 for medium and large firms. The abnormally high returns are observed in deciles 4 across all firm sizes, and in deciles 2, 3 and 5 for some firm size groups.

To summarize, the evidence strongly rejects the idea that firms with the lowest levels of inventory perform best. Instead, consistent with pragmatic JIT, 'low but not too low' inventory seems to have done particularly well. Firms with bloated inventory perform poorly.

4.5.3 Is AbI a Proxy for Risk?

The portfolio analysis shows that high AbI is associated with low stock returns. But, according to standard financial theory, in a stock market equilibrium, different stocks will have different average returns depending on how much nondiversifiable risk they expose their shareholders to. Thus, AbI really could be serving as a proxy for a known risk factor.

To investigate this, we adopt a standard empirical asset pricing framework which is due to Fama and French (1993). Let i be the portfolio index. We run an expected return regression:

$$\mathrm{R}_i - \mathrm{R}_f \;=\; a + \; b(\mathrm{R}_m - \mathrm{R}_f) + \; c\,\mathrm{SMB} + \; d\,\mathrm{HML} + \varepsilon \tag{4.5}$$

The financial risk factors are: R_m (common market factor), R_f (risk-free rate), SMB (firm size factor), and HML (market-to-book ratio factor). In principle, many things could be included as potential risk factors. Empirically, as shown by Fama and French (1993), this relatively small set of factors performs very reliably. We run regression (4.5) for each AbI decile portfolio. The coefficients (b, c, and d) measure how sensitive a given portfolio returns are to the respective risk factors.

If the standard risk factors explain the returns, then the intercept a should equal zero. A value of a that differs significantly from zero is an indication of a return that is not explained by the standard factors. For JIT theory it then becomes interesting to see whether the abnormal values of a are found in the lowest decile as expected under romantic JIT.

Table 4.5(a) shows that abnormally high returns are found from deciles 3 through 7, and ab-

Table 4.5: Fama-French regressions for AbI decile portfolios

(a) All firms

AbI decile portfolio	Intercept	R_m-R_f	SMB	HML	R^2
1	0.17 (0.07)	1.08*** (8.12)	1.05*** (6.64)	0.19 (1.69)	0.90
2	0.76 (0.14)	1.16*** (3.90)	1.09*** (3.08)	0.26 (1.01)	0.66
3	8.14** (2.31)	0.86*** (4.41)	1.45*** (6.26)	-0.16 (-0.96)	0.84
4	12.76*** (3.15)	0.83*** (3.69)	1.68*** (6.29)	-0.42** (-2.19)	0.84
5	9.86*** (3.38)	0.77*** (4.77)	1.64*** (8.55)	-0.32** (-2.30)	0.90
6	9.74*** (2.94)	0.67*** (3.66)	1.72*** (7.88)	-0.42** (-2.70)	0.87
7	8.32*** (3.38)	0.74*** (5.45)	1.55*** (9.54)	-0.49*** (-4.16)	0.92
8	3.13 (1.04)	0.92*** (5.53)	1.28*** (6.49)	0.02 (0.12)	0.86
9	-2.98 (-0.98)	1.06*** (6.28)	1.09*** (5.40)	0.15 (1.07)	0.85
10	-8.47*** (-3.10)	1.04*** (6.91)	0.85*** (4.72)	0.34** (2.59)	0.84

(b) Firm size groups

AbI decile portfolio	Small					Medium					Large				
	Intercept	R_m-R_f	SMB	HML	R^2	Intercept	R_m-R_f	SMB	HML	R^2	Intercept	R_m-R_f	SMB	HML	R^2
1	1.94 (0.32)	1.17*** (3.53)	1.56*** (3.93)	-0.09 (-0.30)	0.71	1.42 (0.58)	1.02*** (7.48)	0.98*** (6.07)	0.24* (2.03)	0.88	-3.63 (-1.13)	1.07*** (6.06)	0.53** (2.49)	0.38** (2.52)	0.76
2	-0.19 (-0.02)	1.44** (2.79)	1.30* (2.11)	0.10 (0.22)	0.50	4.76 (0.86)	1.03*** (3.37)	1.44*** (3.93)	0.29 (1.09)	0.67	-1.81 (-0.84)	0.92*** (7.71)	0.59*** (4.16)	0.39*** (3.77)	0.85
3	10.74 (1.54)	1.06** (2.76)	1.99*** (4.35)	-0.23 (-0.69)	0.69	9.74* (2.03)	0.90*** (3.40)	1.69*** (5.35)	-0.27 (-1.20)	0.78	2.34 (1.26)	0.69*** (6.74)	0.60*** (4.93)	0.15 (1.68)	0.85
4	21.68** (2.89)	0.90** (2.18)	2.68*** (5.43)	-0.91** (-2.57)	0.77	9.21** (2.27)	0.98*** (4.39)	1.49*** (5.57)	-0.12 (-0.64)	0.81	6.98** (2.17)	0.67*** (3.79)	0.96*** (4.52)	-0.18 (-1.19)	0.77
5	16.49** (2.88)	1.03*** (3.27)	2.59*** (6.88)	-0.81*** (-2.97)	0.85	3.75 (1.08)	0.71*** (3.73)	1.50*** (6.59)	0.04 (0.22)	0.82	7.14** (2.50)	0.64*** (4.04)	0.67*** (3.55)	-0.06 (-0.45)	0.72
6	19.41*** (3.35)	0.62* (1.95)	2.69*** (7.06)	-0.96*** (-3.51)	0.84	1.88 (0.70)	0.79*** (5.38)	1.36*** (7.76)	-0.12 (-0.94)	0.88	7.22* (2.00)	0.60*** (3.03)	0.84*** (3.55)	-0.12 (-0.72)	0.67
7	9.41** (2.32)	0.86*** (3.86)	1.99*** (7.47)	-0.72*** (-3.76)	0.88	10.08** (2.31)	0.74*** (3.08)	1.91*** (6.65)	-0.48** (-2.31)	0.83	4.13** (2.80)	0.71*** (8.71)	0.66*** (6.86)	-0.17** (-2.49)	0.93
8	12.01* (1.86)	1.11*** (3.12)	2.18*** (5.14)	-0.29 (-0.94)	0.76	-2.00 (-0.69)	0.93*** (5.88)	1.10*** (5.82)	0.22 (1.62)	0.84	1.24 (0.50)	0.75*** (5.49)	0.77*** (4.72)	0.11 (0.93)	0.81
9	-1.33 (-0.22)	1.23*** (3.67)	1.29*** (3.24)	0.02 (0.08)	0.67	-2.52 (-0.76)	1.06*** (5.82)	1.63*** (7.52)	0.12 (0.78)	0.88	-3.83 (-1.63)	0.88*** (6.81)	0.44** (2.82)	0.25** (2.24)	0.80
10	-8.67** (-2.45)	1.12*** (5.72)	1.07*** (4.57)	0.28 (1.68)	0.81	-6.90** (-2.46)	1.01*** (6.53)	1.05*** (5.69)	0.13 (0.95)	0.86	-8.37** (-2.84)	0.89*** (5.46)	0.37* (1.90)	0.57*** (4.05)	0.72

Intercept, R_m-R_f, SMB and HML refer to the coefficients *a*, *b*, *c* and *d* in equation (5).

*, ** and *** denote statistical significance at 10, 5 and 1 percent level, respectively. *t*-statistics are reported in brackets.

normally low return is found in decile 10. The role of firm size is a potential source of concern. Accordingly, in Table 4.5(b) we divide the portfolios in thirds according to firm size. We then carry out the analysis separately for small firms, medium firms and large firms. The statistical significance of low, but not extremely low, AbI deciles seems to be somewhat stronger for the large firms. The bad performance of extremely low AbI deciles is found for all firm sizes.

The results in Table 4.5 are consistent with the portfolio findings. The fact that similar results are found using such different methods and such different conditioning factors is reassuring. The results appear to be quite robust.

The results show that inventory provides information that is relevant for stock returns. This information is public, and it is not reflected in the standard model (Fama and French 1993) of stock returns. Even adding in the popular momentum factor does not account for the results. How can this happen?

Public information can lead to seemingly excess returns in three well-known ways. First, high returns are compatible with market efficiency if the returns would lead to very high trading costs that would remove what appears to be excess returns. Second, high returns are compatible with market efficiency if the returns are associated with a type of risk that the investors care about, but that is not otherwise reflected in the model. Third, high returns can be obtained if the stock market is not efficient.

Whenever a factor is shown to be associated with high returns, each of these points of view can be proposed. Significant debate has been ongoing in the literature over the relative merits of each interpretation. Fama (1998) provides a helpful overview from an efficient markets perspective.

Transactions cost declined significantly over the past twenty years. But we find that the excess return is both in the 1980s and 1990s data. So we are not inclined to favor the first interpretation. The results in Table 4.5 show that if a risk factor is driving the results, it is not a type of risk that is reflected in the conventional model. In panel (b) of Table 4.5 we see that the abnormal returns in decile 4 are found for all firm size categories. This shows that the effect is not simply a firm size effect. However the effect is numerically largest for the small firm category. Such firms are often deemed to be relatively risky. Thus, the idea that inventory is reflecting a risk factor that is otherwise missing seems plausible to us.

There is, of course, no way to prove that inventory itself is the driving force. Another omitted factor that is suitably correlated with our inventory measure could be the true driving force. For example, our inventory measure (AbI) could be serving as a general proxy for "unexpectedly well run firms." We are only in a position to argue that the evidence is reflecting something that matters. We are not in a position to prove causality.

In an effort to ensure that these results were not being driven by some other omitted factor, we studied other aspects of the corporate balance sheets and income statements. Despite many tests, we were not able to identify any such factors. To save space, we do not report these negative results in any detail.

4.6 Conclusion

This chapter establishes two basic empirical points about the inventory holdings of U.S. manufacturing firms over the 1981-2000 period. First, we show that the broad population of manufacturing firms in the U.S. did significantly reduce their inventories. This reduction was particularly marked for work-in-process inventory. This reduction is not explained by macroeconomic effects, nor by a shift of inventory from public firms towards private firms.

Second, we examined the association between abnormal inventory and stock market performance. In the cross-section, abnormal inventory has no effect on the market-to-book ratio or Tobin's q. Over the longer term, inventory does seem to matter. Firms with abnormally high inventory have poor long term stock market performance. Firms with low, but not extremely low, inventory have unusually good long term stock market performance. However, firms with the lowest levels of inventory have only ordinary performance. These stock market returns are not accounted for by the conventional financial factors of Fama and French (1993).

The skeptical idea that nothing of substance has changed, apart from the macroeconomic conditions, is clearly rejected. However, there is evidence that the macroeconomic conditions affect inventories. Interest rates are negatively related to work-in-process inventory. Inflation is associated with an increase in the holdings of raw materials. Apparently this reflects an effort to buy goods before the prices rise. When managers expect improving economic performance they increase their inventory of finished goods. These macroeconomic factors have sensible impacts, but there is no evidence that they can account for the main long term trend of declining inventory.

In the early 1980s many argued that American manufacturing firms needed to dramatically reduce their inventories. Of course, real firms did not achieve the zero inventory that was advocated by some of the gurus. However, quite respectable reductions did take place. Total inventory declined by about 2% a year on average over the 20 years. Work-in-process has had a remarkable performance with an average annual drop of approximately 6%. Notably immune to the drop was finished goods inventory which was largely unchanged. While this might not have been the kind of inventory revolution envisioned by some in the early 1980s, the improvements that took place are actually quite respectable.

More recently there have been calls for supply chain management researchers to focus on the coordination between suppliers and retailers. Anecdotal evidence of best practices suggests that manufacturing firms can reduce their finished goods inventory through, for example, vendor managed inventory and information sharing. The fact that finished goods inventories did not decline suggests that there may be room for improvement on that front. However, it is likely to take several years before it will be possible to study whether such effects on finished goods inventory for a large number of firms is currently taking place.

References

Balakrishnan, R., T.J. Linsmeier, M. Venkatachalam. 1996. Financial benefits from JIT adoption: Effects of customer concentration and cost structure. *Accounting Review.* **71**(2) 183-205.

Baltagi, B.H. 2001. *Econometric Analysis of Panel Data*, 2nd ed. John Wiley & Sons, New York, NY.

Fama, E.F. 1998. Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics.* **49**(3) 283-306.

Fama E.F., K.R. French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics.* **33**(1) 3-56.

Fisher, M.L., J.H. Hammond, W.R. Obermeyer, A. Raman, 1994. Making supply meet demand in an uncertain world. *Harvard Business Review.* **72**(3) 83-93.

Gompers, P.A., J.L. Ishii, A. Metrick, 2003. Corporate governance and equity prices. *Quarterly Journal of Economics.* **118**(1) 107-155.

Greene, W.H. 2003. *Econometric Analysis*, 5th ed. Prentice Hall, Upper Saddle River, NJ.

Hall, R.W. 1983. *Zero Inventories.* Dow Jones-Irwin, Homewood, IL.

Hendricks, K.B., V.R. Singhal. 2003. The effect of supply chain glitches on shareholder wealth. *Journal of Operations Management* **21**(5) 501-522.

Huson, M., D. Nanda. 1995. The impact of Just-In-Time manufacturing on firm performance in the US. *Journal of Operations Management.* **12** 297-310.

Jegadeesh, N., S. Titman. 1993. Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance.* **48**(1) 65-91.

Rajagopalan, S., A. Malhotra. 2001. Have U.S. manufacturing inventories really decreased? An empirical study. *Manufacturing and Service Operations Management.* **3**(1) 14-24.

Sakakibara, S., B.B. Flynn, R.G. Schroeder, W.T. Morris. 1997. The impact of Just-in-Time manufacturing and its infrastructure on manufacturing performance. *Management Science.* **43**(9) 1246-1257.

Schonberger, R.J. 1982. *Japanese Manufacturing Techniques: Nine Hidden Lessons in Simplicity.* Free Press, New York, NY.

White, G.I., A.C. Sondhi, D. Fried. 1994. *The Analysis and Use of Financial Statements.* John Wiley & Sons, New York, NY.

Zipkin P.H. 1991. Does manufacturing need a JIT revolution? *Harvard Business Review.* **69**(1) 40-50.

Appendix

A.1 Proofs of Lemmas, Propositions, and Theorems in Chapter 2

Derivation of Equation (2.6). Here we derive the Hamilton-Jacobi-Bellman equation for a more general optimal control problem with state-dependent discount rate. As in a canonical optimal control problem, the state variable $\mathbf{x}$ is a diffusion in $\Re^n$, with drift and diffusion depending on the control $\pi \in \mathcal{U}$:

$$d\mathbf{x}_t = \boldsymbol{\mu}(\mathbf{x}_t, \pi_t)dt + \boldsymbol{\sigma}(\mathbf{x}_t, \pi_t)d\mathbf{w}_t.$$

The running profit $f : \Re^n \times \mathcal{U} \to \Re$ is a continuous function of $(\mathbf{x}, \pi)$. The discount rate $\rho : \Re^n \to \Re$ is a continuous function of the state $\mathbf{x}$, and hence the cumulative discount factor is $r_t = \int_0^t \rho(\mathbf{x}_s)ds$. The problem is

$$V(\mathbf{x}) = \sup_{\{\pi_t \in \mathcal{U}\}} \mathsf{E}_0^\pi \int_0^\infty e^{-r_t} f(\mathbf{x}_t, \pi_t)dt,$$

where the state process starts at $\mathbf{x}_0 = \mathbf{x}$. Consider an auxiliary problem that includes $\widetilde{r}_t = r + r_t$ as another state variable starting from r:

$$W(\mathbf{x}, r) = \sup_{\{\pi_t \in \mathcal{U}\}} \mathsf{E}_0^\pi \int_0^\infty e^{-\widetilde{r}_t} f(\mathbf{x}_t, \pi_t)dt.$$

With the augmented state space, the above problem is in canonical form. The HJB equation (see, for example, Fleming and Soner 1993) for $W(\mathbf{x}, r)$ is

$$\sup_{\pi \in \mathcal{U}} \left\{ e^{-r} f(\mathbf{x}, \pi) + \frac{\partial W}{\partial r}\rho(\mathbf{x}) + \sum_{i=1}^n \frac{\partial W}{\partial x_i}\mu_i(\mathbf{x}, \pi) + \sum_{i,j=1}^n b_{ij}(\mathbf{x}, \pi)\frac{\partial^2 W}{\partial x_i \partial x_j} \right\} = 0,$$

where $b_{ij} = \frac{1}{2}(\sigma\sigma^T)_{ij}$. Obviously, $W(\mathbf{x}, r) = e^{-r}V(\mathbf{x})$. Substituting this relation into the above equation, we have

$$\sup_{\pi \in \mathcal{U}} \left\{ f(\mathbf{x}, \pi) + \sum_{i=1}^n \frac{\partial V}{\partial x_i}\mu_i(\mathbf{x}, \pi) + \sum_{i,j=1}^n b_{ij}(\mathbf{x}, \pi)\frac{\partial^2 V}{\partial x_i \partial x_j} \right\} = V(\mathbf{x})\rho(\mathbf{x}).$$

∎

The proof of Proposition 2.2 and 2.3 will use the following lemma.

Lemma A.1.1 *Let $x^0 \leq x^1$ and $y^0 \leq y^1$. Let $x^\alpha = (1-\alpha)x^0 + \alpha x^1$, $y^\beta = (1-\beta)y^0 + \beta y^1$.*

(i) If $f(x,y)$ is convex and supermodular on $\Re^2$, then

$$f(x^0,y^0) + f(x^1,y^1) \geq f(x^\alpha, y^\beta) + f(x^{1-\alpha}, y^{1-\beta}), \quad \forall \alpha, \beta \in [0,1].$$

(ii) If $f(x,y)$ is convex and submodular on $\Re^2$, then

$$f(x^0,y^1) + f(x^1,y^0) \geq f(x^\alpha, y^\beta) + f(x^{1-\alpha}, y^{1-\beta}), \quad \forall \alpha, \beta \in [0,1].$$

Proof. (i) First note that $x^{\alpha_1} \leq x^{\alpha_2}$ if and only if $\alpha_1 \leq \alpha_2$. Thus, $\min\{x^{\alpha_1}, x^{\alpha_2}\} = x^{\min\{\alpha_1,\alpha_2\}}$. Then, the supermodularity property implies that

$$f(x^{\min\{\alpha,1-\alpha\}}, y^{\min\{\beta,1-\beta\}}) + f(x^{\max\{\alpha,1-\alpha\}}, y^{\max\{\beta,1-\beta\}}) \geq f(x^\alpha, y^\beta) + f(x^{1-\alpha}, y^{1-\beta}).$$

Hence, to prove that the inequality in the lemma holds for $(\alpha,\beta) \in [0,1]^2$, it suffices to consider only $(\alpha,\beta) \in [0,\frac{1}{2}]^2$. Without loss of generality, we consider $0 \leq \alpha \leq \beta \leq \frac{1}{2}$. Under this condition, we have

$$\begin{aligned} f(x^\alpha, y^\beta) - f(x^\alpha, y^\alpha) &\leq f(x^{1-\alpha}, y^\beta) - f(x^{1-\alpha}, y^\alpha) \\ &\leq f(x^{1-\alpha}, y^{1-\alpha}) - f(x^{1-\alpha}, y^{1-\beta}), \end{aligned}$$

where the first inequality follows from supermodularity and the second inequality is due to the convexity in y. Rearranging terms,

$$f(x^\alpha, y^\alpha) + f(x^{1-\alpha}, y^{1-\alpha}) \geq f(x^\alpha, y^\beta) + f(x^{1-\alpha}, y^{1-\beta}).$$

By the convexity of f,

$$f(x^0,y^0) + f(x^1,y^1) \geq f(x^\alpha, y^\alpha) + f(x^{1-\alpha}, y^{1-\alpha}).$$

The above two inequalities lead to the desired inequality.

(ii) By similar argument as in part (i), it suffices to consider $0 \leq \alpha \leq 1-\beta \leq \frac{1}{2}$. Under this condition, we have

$$\begin{aligned} f(x^{1-\alpha}, y^{1-\beta}) - f(x^{1-\alpha}, y^\alpha) &\leq f(x^{1-\alpha}, y^{1-\alpha}) - f(x^{1-\alpha}, y^\beta) \\ &\leq f(x^\alpha, y^{1-\alpha}) - f(x^\alpha, y^\beta), \end{aligned}$$

where the first inequality follows from the concavity in y and the second inequality is due to the submodularity. Rearranging terms,

$$f(x^{\alpha}, y^{1-\alpha}) + f(x^{1-\alpha}, y^{\alpha}) \geq f(x^{\alpha}, y^{\beta}) + f(x^{1-\alpha}, y^{1-\beta}).$$

By convexity,

$$f(x^0, y^1) + f(x^1, y^0) \geq f(x^{\alpha}, y^{1-\alpha}) + f(x^{1-\alpha}, y^{\alpha}).$$

The above two inequalities lead to the desired inequality in part (ii). ∎

Proof of Proposition 2.2. Consider the initial inventory levels (x_1^a, x_2^a) and (x_1^b, x_2^b) with $x_1^a < x_1^b$ and $x_2^a < x_2^b$. Let $\boldsymbol{\pi}^{aa} = \{\pi_t^{aa} = (\lambda_t^{aa}, q_t^{aa}, s_t^{aa}) : t \geq 0\}$ and $\boldsymbol{\pi}^{bb} = \{\pi_t^{bb} = (\lambda_t^{bb}, q_t^{bb}, s_t^{bb}) : t \geq 0\}$ be the corresponding optimal controls, and let $\{\mathbf{x}_t^{aa} : t \geq 0\}$ and $\{\mathbf{x}_t^{bb} : t \geq 0\}$ be the corresponding optimal inventory processes. Clearly, $\mathbf{x}_0^{aa} = (x_1^a, x_2^a) < (x_1^b, x_2^b) = \mathbf{x}_0^{bb}$.

Consider initial inventory (x_1^a, x_2^b) and (x_1^b, x_2^a). We now construct admissible controls under which the controlled inventory processes can be expressed as convex combination of the optimal inventory processes. Then applying Lemma A.1.1 and using the convexity and the supermodularity of g and h lead to the submodularity of the value function.

Let us define

$$T := \inf\{\, t > 0 \;:\; x_{1t}^{aa} = x_{1t}^{bb} \quad \text{or} \quad x_{2t}^{aa} = x_{2t}^{bb}\}.$$

That is, T is the first time that $\mathbf{x}_t^{aa} < \mathbf{x}_t^{bb}$ does not hold.

For initial inventory (x_1^a, x_2^b) and (x_1^b, x_2^a), consider applying the following controls, respectively:

$$\begin{aligned} \boldsymbol{\pi}_{[0,T)}^{ab} &= \{\pi_t^{ab} = (\lambda_t^{aa}, \min\{q_t^{aa}, q_t^{bb}\}, s_t^{bb}) : t \in [0,T)\}, \\ \boldsymbol{\pi}_{[0,T)}^{ba} &= \{\pi_t^{ba} = (\lambda_t^{bb}, \max\{q_t^{aa}, q_t^{bb}\}, s_t^{aa}) : t \in [0,T)\}. \end{aligned} \tag{A.1}$$

Let $\{\mathbf{x}_t^{ab} : t \in [0,T)\}$ and $\{\mathbf{x}_t^{ba} : t \in [0,T)\}$ be the controlled inventory processes under the above policies. By the balance equations (2.3) and (A.1) we have

$$dx_{1t}^{ab} = (\lambda_t^{aa} - \min\{q_t^{aa}, q_t^{bb}\})dt \geq (\lambda_t^{aa} - q_t^{aa})dt = dx_{1t}^{aa}, \tag{A.2}$$

$$dx_{2t}^{ab} = (\min\{q_t^{aa}, q_t^{bb}\} - s_t^{bb})dt \leq (q_t^{bb} - s_t^{bb})dt = dx_{2t}^{bb}, \tag{A.3}$$

$$dx_{1t}^{ba} = (\lambda_t^{bb} - \max\{q_t^{aa}, q_t^{bb}\})dt \leq (\lambda_t^{bb} - q_t^{bb})dt = dx_{1t}^{bb}, \tag{A.4}$$

$$dx_{2t}^{ba} = (\max\{q_t^{aa}, q_t^{bb}\} - s_t^{aa})dt \geq (q_t^{aa} - s_t^{aa})dt = dx_{2t}^{aa}. \tag{A.5}$$

In the rest of the proof, for notational simplicity, we define

$$z^{\alpha} \quad := \quad (1-\alpha)z^{aa} + \alpha z^{bb},$$

where $\alpha \in \Re$, and z can be $\mathbf{x}_t$ or π_t or any component of them.

Since $\mathbf{x}_t^{aa} < \mathbf{x}_t^{bb}$ for $t \in [0,T)$, we can express $\mathbf{x}_t^{ab}$ in the following form:

$$\mathbf{x}_t^{ab} = \left(x_{1t}^{\alpha_t}, x_{2t}^{\beta_t}\right), \tag{A.6}$$

where α_t and β_t are some real values that can be uniquely determined. From (A.2-A.5) we have $\mathbf{x}_t^{ab} + \mathbf{x}_t^{ba} = \mathbf{x}_t^{aa} + \mathbf{x}_t^{bb}$. Subtracting (A.6) from this identity gives

$$\mathbf{x}_t^{ba} = \left(x_{1t}^{1-\alpha_t}, x_{2t}^{1-\beta_t}\right). \tag{A.7}$$

Note that inventory processes are continuous processes, so are $\{\alpha_t : t \in [0,T)\}$ and $\{\beta_t : t \in [0,T)\}$. The raw material inventory process $\{x_{1t}^{ab}\}$ starts from the same point as $\{x_{1t}^{aa}\}$, but then rises above it with increasing difference (due to (A.2)). Similarly, the finished goods inventory $\{x_{2t}^{ab}\}$ initially coincides with $\{x_{2t}^{bb}\}$, but then drops below it with increasing difference (due to (A.3)). These facts imply that $\alpha_0 = 0$, $\beta_0 = 1$, and $\alpha_t \geq 0$, $\beta_t \leq 1$ for $t \in [0,T)$. Define a stopping time:

$$s \quad := \quad \inf\{\, t \in [0,T) : \alpha_t = \beta_t\}. \tag{A.8}$$

As a convention, $s := \infty$ if $\alpha_t \neq \beta_t$ for all $t \in [0,T)$. By the continuity of α_t and β_t, we must have $\alpha_t, \beta_t \in [0,1]$ for $t \in [0,s]$, and consequently (A.6) and (A.7) implies that $\mathbf{x}_t^{ab}$ and $\mathbf{x}_t^{ba}$ stay within the box whose lower-left and upper-right corners are $\mathbf{x}_t^{aa}$ and $\mathbf{x}_t^{bb}$, respectively (see Figure A.1).

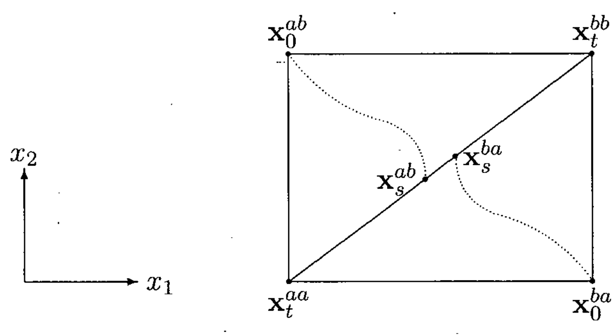


Figure A.1: Illustration of the controlled inventory processes (Proposition 2.2)

Depending on whether s or T is finite, we have three cases.

Case 1: $s < \infty$.

In this case, $\mathbf{x}_s^{ab}$ and $\mathbf{x}_s^{ba}$ are convex combinations of $\mathbf{x}_s^{aa}$ and $\mathbf{x}_s^{bb}$. That is, $\mathbf{x}_s^{ab} = \mathbf{x}_s^{\alpha_s}$ and $\mathbf{x}_s^{ba} = \mathbf{x}_s^{1-\alpha_s}$. From s onwards ($t \in [s, \infty)$), we apply the following controls:

$$\boldsymbol{\pi}_{[s,\infty)}^{ab} = \left\{\pi_t^{\alpha_s} : t \in [s, \infty)\right\}, \qquad \boldsymbol{\pi}_{[s,\infty)}^{ba} = \left\{\pi_t^{1-\alpha_s} : t \in [s, \infty)\right\}. \tag{A.9}$$

The above controls will maintain the inventory processes as convex combinations of the two optimal inventory processes from s onwards, that is, $\mathbf{x}_t^{ab} = \mathbf{x}_t^{\alpha_s}$ and $\mathbf{x}_t^{ba} = \mathbf{x}_t^{1-\alpha_s}$, $\forall t \in [s, \infty)$.

Therefore, under the controls in (A.1) with the time ranges replaced by $[0, s)$ and the controls in (A.9) for $[s, \infty)$, there exist α_t and β_t in $[0, 1]$ such that

$$\mathbf{x}_t^{ab} = \left(x_{1t}^{\alpha_t}, x_{2t}^{\beta_t}\right), \qquad \mathbf{x}_t^{ba} = \left(x_{1t}^{1-\alpha_t}, x_{2t}^{1-\beta_t}\right), \qquad \forall t \in [0, \infty). \tag{A.10}$$

In addition, if we define $\alpha_t' = 0$ if $q_t^{ab} = q_t^{aa}$, and $\alpha_t' = 1$ if $q_t^{ab} = q_t^{bb}$, and $\alpha_t' = \alpha_s$ for $t \geq s$, then

$$q_t^{ab} = q_t^{\alpha_t'}, \qquad q_t^{ba} = q_t^{1-\alpha_t'}, \qquad \forall t \in [0, \infty). \tag{A.11}$$

And obviously,

$$\lambda_t^{ab} + \lambda_t^{ba} = \lambda_t^{aa} + \lambda_t^{bb}, \qquad s_t^{ab} + s_t^{ba} = s_t^{aa} + s_t^{bb}, \qquad \forall t \in [0, \infty). \tag{A.12}$$

Case 2: $s = \infty$ and $T = \infty$.

In this case, we just implement the policies defined in (A.1). Clearly, (A.10-A.12) still hold.

Case 3: $s = \infty$ and $T < \infty$.

In this case, the box in Figure A.1 collapses into a line or a dot at time T, while $\mathbf{x}_t^{ab}$ or $\mathbf{x}_t^{ba}$ never hits the diagonal line before T.

Without loss of generality, assume $x_{1T}^{aa} = x_{1T}^{bb}$. We claim that $\alpha_t = 0$ for $t \in [0, T)$ in this case. To see this, suppose $\alpha_{t_0} > 0$ for some $t_0 \in [0, T)$, then $\alpha_t = (x_{1t}^{ab} - x_{1t}^{aa})/(x_{1t}^{bb} - x_{1t}^{aa}) \to \infty$ as $t \to T$, because the numerator is positive and increasing while the denominator shrinks to zero. Thus there exists $t_1 < T$ such that $\alpha_{t_1} = \beta_{t_1}$, which is contradictory to the assumption that $s = \infty$.

Now define $\beta_T = (x_{2T}^{ab} - x_{2T}^{aa})/(x_{2T}^{bb} - x_{2T}^{aa})$, and for $t \in [T, \infty)$ apply policies (A.9) with α_s replaced by β_T here. It can be easily verified that (A.10-A.12) still hold.

In all three cases, for $t < \min\{s, T\}$, we have $\mathbf{x}_t^{aa} < \mathbf{x}_t^{bb}$. Then it follows from Lemma A.1.1 and (A.10) that

$$f(\mathbf{x}_t^{aa}) + f(\mathbf{x}_t^{bb}) \quad \geq \quad f(\mathbf{x}_t^{ab}) + f(\mathbf{x}_t^{ba}), \tag{A.13}$$

where $f(\cdot)$ can be $g_2(\cdot, \mathbf{k})$ or $h(\cdot, \mathbf{k})$ for any $\mathbf{k}$. While for $t \geq \min\{s, T\}$, $\mathbf{x}_t^{ab}$ and $\mathbf{x}_t^{ba}$ are convex combination of $\mathbf{x}_t^{aa}$ and $\mathbf{x}_t^{bb}$, and hence, (A.13) still holds.

To prove the submodularity, we note that

$$
\begin{aligned}
& V(x_1^a, x_2^b, \mathbf{p}, \mathbf{k}) + V(x_1^b, x_2^a, \mathbf{p}, \mathbf{k}) \\
\geq\ & \mathsf{E}_0 \int_0^\infty e^{-R_t} \Big[(s_t^{ab} + s_t^{ba}) p_{2t} - (\lambda_t^{ab} + \lambda_t^{ba}) p_{1t} - g_1(q_t^{ab}, \mathbf{k}_t) - g_2(\mathbf{x}_t^{ab}, \mathbf{k}_t) - g_1(q_t^{ba}, \mathbf{k}_t) \\
& - g_2(\mathbf{x}_t^{ba}, \mathbf{k}_t) - h(\mathbf{x}_t^{ab}, \mathbf{k}_t) - h(\mathbf{x}_t^{ba}, \mathbf{k}_t) \Big] dt \\
\geq\ & \mathsf{E}_0 \int_0^\infty e^{-R_t} \Big[(s_t^{aa} + s_t^{bb}) p_{2t} - (\lambda_t^{aa} + \lambda_t^{bb}) p_{1t} - g_1(q_t^{aa}, \mathbf{k}_t) - g_2(\mathbf{x}_t^{aa}, \mathbf{k}_t) - g_1(q_t^{bb}, \mathbf{k}_t) \\
& - g_2(\mathbf{x}_t^{bb}, \mathbf{k}_t) - h(\mathbf{x}_t^{aa}, \mathbf{k}_t) - h(\mathbf{x}_t^{bb}, \mathbf{k}_t) \Big] dt \\
=\ & V(x_1^a, x_2^a, \mathbf{p}, \mathbf{k}) + V(x_1^b, x_2^b, \mathbf{p}, \mathbf{k}),
\end{aligned}
$$

where the first inequality follows from the fact that the controls constructed above are feasible but not necessarily optimal, while the second inequality follows directly from (A.11-A.13) and the convexity of g_1 in q. ∎

Proof of Proposition 2.3. This proof is similar to the proof of Proposition 2.2, and therefore, is abridged. We only prove the increasing substitution in x_1.

Consider the initial inventory levels (x_1^a, x_2^a) and (x_1^b, x_2^b) with $x_1^b - x_1^a > x_2^a - x_2^b \geq \delta$. Let $\boldsymbol{\pi}^{aa} = \{\pi_t^{aa} : t \geq 0\}$ and $\boldsymbol{\pi}^{bb} = \{\pi_t^{bb} : t \geq 0\}$ be the corresponding optimal controls, and let $\{\mathbf{x}_t^{aa} : t \geq 0\}$ and $\{\mathbf{x}_t^{bb} : t \geq 0\}$ be the corresponding optimal inventory processes.

Consider initial inventory $(x_1^a + \delta, x_2^b)$ and $(x_1^b - \delta, x_2^a)$. We now construct admissible controls under which the inventory processes stay within the parallelogram, shown in Figure A.2.

Let $T := \inf\{t > 0 : x_{1t}^{aa} + x_{2t}^{aa} = x_{1t}^{bb} + x_{2t}^{bb} \text{ or } x_{2t}^{aa} = x_{2t}^{bb}\}$ be the first time that the parallelogram collapses into a line or a point. For initial inventory $(x_1^a + \delta, x_2^b)$ and $(x_1^b - \delta, x_2^a)$, consider applying the following controls, respectively:

$$
\begin{aligned}
\boldsymbol{\pi}_{[0,T)}^{ab} &= \{\pi_t^{ab} = (\lambda_t^{aa}, \max\{q_t^{aa}, q_t^{bb}\}, \min\{s_t^{aa}, s_t^{bb}\}) : t \in [0, T)\}, \\
\boldsymbol{\pi}_{[0,T)}^{ba} &= \{\pi_t^{ba} = (\lambda_t^{bb}, \min\{q_t^{aa}, q_t^{bb}\}, \max\{s_t^{aa}, s_t^{bb}\}) : t \in [0, T)\}.
\end{aligned}
\tag{A.14}
$$

Let $\{\mathbf{x}_t^{ab} : t \in [0, T)\}$ and $\{\mathbf{x}_t^{ba} : t \in [0, T)\}$ be the controlled inventory processes under the above policies. By the balance equations (2.3) and (A.14) we have

$$
dx_{2t}^{ab} = (\max\{q_t^{aa}, q_t^{bb}\} - \min\{s_t^{aa}, s_t^{bb}\})dt \geq dx_{2t}^{bb}, \tag{A.15}
$$

$$
dx_{2t}^{ba} = (\min\{q_t^{aa}, q_t^{bb}\} - \max\{s_t^{aa}, s_t^{bb}\})dt \leq dx_{2t}^{aa}, \tag{A.16}
$$

$$dx_{1t}^{ab} + dx_{2t}^{ab} = (\lambda_t^{aa} - \min\{s_t^{aa}, s_t^{bb}\})dt \qquad \geq dx_{1t}^{aa} + dx_{2t}^{aa}, \tag{A.17}$$

$$dx_{1t}^{ba} + dx_{2t}^{ba} = (\lambda_t^{bb} - \max\{s_t^{aa}, s_t^{bb}\})dt \qquad \leq dx_{1t}^{bb} + dx_{2t}^{bb}. \tag{A.18}$$

From (A.15-A.18), the relative positions of $\mathbf{x}_t^{ab}$ and $\mathbf{x}_t^{ba}$ with respect to the optimally controlled processes are shown in Figure A.2.

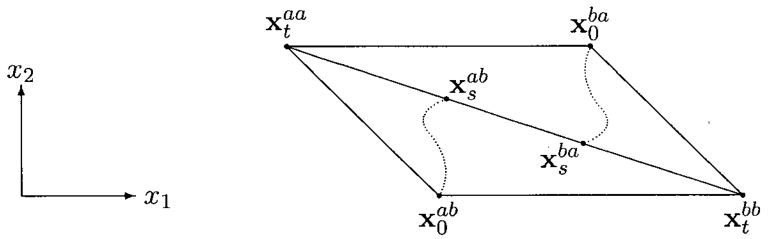


Figure A.2: Illustration of the controlled inventory processes (Proposition 2.3)

Since $x_{1t}^{aa} < x_{1t}^{bb}$ and $x_{2t}^{aa} > x_{2t}^{bb}$ for $t \in [0, T)$, we can find unique α_t and β_t such that (A.6)-(A.7) hold. The relative positions of $\mathbf{x}_t^{ab}$ and $\mathbf{x}_t^{ba}$ with respect to $\mathbf{x}_t^{aa}$ and $\mathbf{x}_t^{bb}$ imply that $\alpha_0 \in (0, 1), \beta_0 = 1$ and $\alpha_t \geq 0, \beta_t \leq 1$ for $t \in [0, T)$.

Define the same stopping time as in (A.8). If $s < \infty$, it is the time when $\mathbf{x}_t^{ab}$ and $\mathbf{x}_t^{ba}$ hit the longer diagonal, as shown in Figure A.2. From s onwards, we apply the same controls as in (A.9). This keeps the controlled inventory processes staying as convex combination of $\mathbf{x}_t^{aa}$ and $\mathbf{x}_t^{bb}$ for $t \in [s, \infty)$.

Similar to the proof of Proposition 2.2, we need to consider the other two cases. Since the technical details are almost the same, they are not presented in any detail here.

In all three cases, there exist α_t and β_t in $[0, 1]$, such that $\mathbf{x}_t^{ab} = (x_{1t}^{\alpha_t}, x_{2t}^{\beta_t})$ and $\mathbf{x}_t^{ba} = (x_{1t}^{1-\alpha_t}, x_{2t}^{1-\beta_t})$ for all $t \in [0, \infty)$. For $t < \min\{s, T\}$, these processes stay symmetrically within the parallelogram. If $f(x_1, x_2)$ is convex and has decreasing substitution in x_1, then $\widehat{f}(y, x_2) := f(y - x_2, x_2)$ is convex and submodular in (y, x_2), and it follows from Lemma A.1.1(ii) that

$$\widehat{f}(x_{1t}^{aa} + x_{2t}^{aa}, x_{2t}^{aa}) + \widehat{f}(x_{1t}^{bb} + x_{2t}^{bb}, x_{2t}^{bb}) \;\; \geq \;\; \widehat{f}(x_{1t}^{ab} + x_{2t}^{ab}, x_{2t}^{ab}) + \widehat{f}(x_{1t}^{ba} + x_{2t}^{ba}, x_{2t}^{ba}),$$

or equivalently,

$$f(\mathbf{x}_t^{aa}) + f(\mathbf{x}_t^{bb}) \;\; \geq \;\; f(\mathbf{x}_t^{ab}) + f(\mathbf{x}_t^{ba}), \tag{A.19}$$

where $f(\cdot)$ can be $g_2(\cdot, \mathbf{k})$ or $h(\cdot, \mathbf{k})$ for any $\mathbf{k}$. While for $t \geq \min\{s, T\}$, $\mathbf{x}_t^{ab}$ and $\mathbf{x}_t^{ba}$ are convex combination of $\mathbf{x}_t^{aa}$ and $\mathbf{x}_t^{bb}$, and hence (A.19) still holds. The rest of the proof is completely analogous to that of Proposition 2.2. ∎

The proof of Theorem 2.2 will use the following lemma.

Lemma A.1.2 *Let* $\mathbf{F} = \mathbf{CA} + \frac{\rho^2}{4}\mathbf{I}$ *as defined in Theorem 2.2. Then,*

(i) $\mathsf{tr}[\mathbf{F}] > \frac{\rho^2}{2}$;

(ii) $\mathsf{det}[\mathbf{F}] > \frac{\rho^2}{4}\mathsf{tr}[\mathbf{F}] - \frac{\rho^4}{16} > \frac{\rho^4}{16}$;

(iii) $\mathsf{tr}[\mathbf{F}]^2 - 4\mathsf{det}[\mathbf{F}] \geq 0$, *where equality holds if and only if* $\mathbf{F}$ *is a diagonal matrix.*

Proof. By definition,

$$\mathbf{F} = \begin{bmatrix} c_{11}A_1 + \frac{c_{11}-c_{12}}{a_2} + \frac{\rho^2}{4} & c_{12}A_2 - \frac{c_{11}-c_{12}}{a_2} \\ c_{12}A_1 - \frac{c_{22}-c_{12}}{a_2} & c_{22}A_2 + \frac{c_{22}-c_{12}}{a_2} + \frac{\rho^2}{4} \end{bmatrix} \equiv \begin{bmatrix} f_{11} & f_{12} \\ f_{21} & f_{22} \end{bmatrix}.$$

(i) $\mathsf{tr}[\mathbf{F}] = c_{11}A_1 + c_{22}A_2 + \frac{c_{11}+c_{22}-2c_{12}}{a_2} + \frac{\rho^2}{2} > \frac{\rho^2}{2}$, where we used $c_{11}A_1 + c_{22}A_2 > 0$, $c_{12} < \sqrt{c_{11}c_{22}}$ and $c_{11} + c_{22} - 2\sqrt{c_{11}c_{22}} \geq 0$.

(ii) $\mathsf{det}[\mathbf{F}] = \mathsf{det}[\mathbf{CA} + \frac{\rho^2}{4}\mathbf{I}] = \mathsf{det}[\mathbf{CA}] + \frac{\rho^2}{4}\left(\mathsf{tr}[\mathbf{F}] - \frac{\rho^2}{2}\right) + \frac{\rho^4}{16} > \frac{\rho^2}{4}\mathsf{tr}[\mathbf{F}] - \frac{\rho^4}{16}$, where the inequality follows from $\mathsf{det}[\mathbf{C}] > 0$, $\mathsf{det}[\mathbf{A}] > 0$ and part (i).

(iii) When $\mathbf{F}$ is a diagonal matrix, $f_{12} = f_{21} = 0$, we have $A_1 = \frac{c_{22}-c_{12}}{c_{12}a_2}$ and $A_2 = \frac{c_{11}-c_{12}}{c_{12}a_2}$, which imply that $f_{11} = f_{22} = \frac{c_{11}c_{22}-c_{12}^2}{c_{12}a_2} + \frac{\rho^2}{4}$. Then, $\mathsf{tr}[\mathbf{F}]^2 - 4\mathsf{det}[\mathbf{F}] = 0$.

When $\mathbf{F}$ is not a diagonal matrix, we show that $\mathsf{tr}[\mathbf{F}]^2 - 4\mathsf{det}[\mathbf{F}] > 0$, or equivalently,

$$(f_{11} - f_{22})^2 > -4f_{12}f_{21}. \tag{A.20}$$

(a) $f_{12}f_{21} > 0$. It is clear that (A.20) holds in this case.

(b) $f_{12}f_{21} = 0$. Since $\mathbf{F}$ is not diagonal, exactly one of the two equalities $A_2 = \frac{c_{11}-c_{12}}{c_{12}a_2}$ and $A_1 = \frac{c_{22}-c_{12}}{c_{12}a_2}$ holds, implying that $c_{11}A_1 - c_{22}A_2 + \frac{c_{11}-c_{22}}{a_2} \neq 0$ or $f_{11} - f_{22} \neq 0$. Hence (A.20) holds.

(c) $f_{21} > 0$ and $f_{12} < 0$, or

$$A_1 > \frac{c_{22}-c_{12}}{c_{12}a_2} \quad \text{and} \quad A_2 < \frac{c_{11}-c_{12}}{c_{12}a_2}. \tag{A.21}$$

The second inequality in (A.21) implies $c_{11} > c_{12}$ because $A_2 > 0$. We first prove an inequality

$$(c_{11} - c_{12})A_1 > (c_{22} - c_{12})A_2. \tag{A.22}$$

If $c_{22} \leq c_{12}$, then (A.22) clearly holds. While if $c_{22} > c_{12}$, (A.21) impiles that $(c_{11}-c_{12})A_1 > \frac{(c_{11}-c_{12})(c_{22}-c_{12})}{c_{12}a_2} > (c_{22} - c_{12})A_2$. Thus,

$$c_{11}A_1 - c_{22}A_2 + \tfrac{c_{11}-c_{22}}{a_2} \;>\; c_{12}A_1 - c_{12}A_2 + \tfrac{c_{11}-c_{22}}{a_2} \;>\; \tfrac{c_{22}-c_{12}+c_{12}-c_{11}+c_{11}-c_{22}}{a_2} \;=\; 0,$$

where the first inequality follows from (A.22) and the second inequality follows from (A.21). Hence,

$$\begin{aligned}(f_{11} - f_{22})^2 &= \Big(c_{11}A_1 - c_{22}A_2 + \frac{c_{11} - c_{22}}{a_2}\Big)^2 \\ &> \Big(c_{12}A_1 - c_{12}A_2 + \frac{c_{11} - c_{22}}{a_2}\Big)^2 \;=\; (f_{21} - f_{12})^2 \;\geq\; -4f_{12}f_{21}.\end{aligned}$$

(d) $f_{21} < 0$ and $f_{12} > 0$. The argument is completely analogous to case (c).

∎

Proof of Theorem 2.2. There are six steps.

1. We first show that $\mathbf{F}$ has positive real eigenvalues and linearly independent eigenvectors. Let ξ_1 and ξ_2 denote the eigenvalues of $\mathbf{F}$, which are the roots to the equation:

$$\mathsf{det}[\mathbf{F} - \xi\mathbf{I}] \;=\; \xi^2 - \mathsf{tr}[\mathbf{F}]\xi + \mathsf{det}[\mathbf{F}] \;=\; 0.$$

When $\mathbf{F}$ is a diagonal matrix, the proof of Lemma A.1.2(iii) has shown that $\xi_1 = \xi_2 = \frac{c_{11}c_{22}-c_{12}^2}{c_{12}a_2} + \frac{\rho^2}{4} > 0$, and $\mathbf{F}$ has two linearly independent eigenvectors, e.g., $[0, 1]$ and $[1, 0]$.

When $\mathbf{F}$ is not a diagonal matrix, from Lemma A.1.2(iii), $\mathsf{tr}[\mathbf{F}]^2 - 4\mathsf{det}[\mathbf{F}] > 0$, meaning that $\mathbf{F}$ has two different real eigenvalues, and therefore, $\mathbf{F}$ has two linearly independent eigenvectors. Furthermore, from Lemma A.1.2, the eigenvalues must be greater than $\frac{\rho^2}{4}$:

$$\min\{\xi_1, \xi_2\} \;=\; \tfrac{\mathsf{tr}[\mathbf{F}]-\sqrt{\mathsf{tr}[\mathbf{F}]^2-4\mathsf{det}[\mathbf{F}]}}{2} \;>\; \tfrac{\mathsf{tr}[\mathbf{F}]-\sqrt{\mathsf{tr}[\mathbf{F}]^2-\mathsf{tr}[\mathbf{F}]\rho^2+\frac{\rho^4}{4}}}{2} \;=\; \tfrac{\mathsf{tr}[\mathbf{F}]-(\mathsf{tr}[\mathbf{F}]-\frac{\rho^2}{2})}{2} \;=\; \tfrac{\rho^2}{4}. \tag{A.23}$$

2. Next, we show that there exists a solution $\mathbf{B}$ to $\mathbf{B}^2 - \rho\mathbf{B} = \mathbf{CA}$ and (2.39), and the solution $\mathbf{B}$ has negative real eigenvalues. Let $\Xi = \mathsf{diag}[\xi_1, \xi_2]$ and $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2]$, where $\mathbf{v}_i$ is the eigenvector corresponding to ξ_i, and $\mathbf{v}_1$ and $\mathbf{v}_2$ are linearly independent. Then, we have $\mathbf{F} = \mathbf{V}\Xi\mathbf{V}^{-1}$. We define $\sqrt{\Xi} := \mathsf{diag}[\sqrt{\xi_1}, \sqrt{\xi_2}]$, and $\sqrt{\mathbf{F}} := \mathbf{V}\sqrt{\Xi}\mathbf{V}^{-1}$. Note that $\mathsf{tr}[\sqrt{\mathbf{F}}] = \mathsf{tr}[\sqrt{\Xi}]$ and $\mathsf{det}[\sqrt{\mathbf{F}}] = \mathsf{det}[\sqrt{\Xi}]$.

We show that $\mathbf{B} = \frac{\rho}{2}\mathbf{I} - \sqrt{\mathbf{F}}$ is a solution to $\mathbf{B}^2 - \rho\mathbf{B} = \mathbf{CA}$:

$$\mathbf{B}^2 - \rho\mathbf{B} \;=\; \tfrac{\rho^2}{4}\mathbf{I} + \mathbf{F} - \rho\sqrt{\mathbf{F}} - \tfrac{\rho^2}{2}\mathbf{I} + \rho\sqrt{\mathbf{F}} \;=\; -\tfrac{\rho^2}{4}\mathbf{I} + \mathbf{F} \;=\; \mathbf{CA}.$$

Furthermore, $\mathbf{B}$ satisfies (2.39) because

$$\mathsf{tr}[\mathbf{B}] \;=\; \mathsf{tr}\big[\tfrac{\rho}{2}\mathbf{I} - \sqrt{\mathbf{F}}\big] \;=\; \rho - \mathsf{tr}[\sqrt{\Xi}] \;=\; \rho - \sqrt{\xi_1 + \xi_2 + 2\sqrt{\xi_1\xi_2}} \;<\; \rho - \sqrt{\tfrac{\rho^2}{2} + 2\sqrt{\tfrac{\rho^4}{16}}} = 0,$$

where the last inequality follows from Lemma A.1.2, and

$$\mathsf{det}[\mathbf{B}] \;=\; \mathsf{det}[\sqrt{\mathbf{F}}] - \tfrac{\rho}{2}\mathsf{tr}[\sqrt{\mathbf{F}}] + \tfrac{\rho^2}{4} \;=\; \mathsf{det}[\sqrt{\Xi}] - \tfrac{\rho}{2}\mathsf{tr}[\sqrt{\Xi}] + \tfrac{\rho^2}{4} \;=\; (\sqrt{\xi_1} - \tfrac{\rho}{2})(\sqrt{\xi_2} - \tfrac{\rho}{2}) \;>\; 0,$$

where the last inequality follows from (A.23). The above conditions imply that the eigenvalues of $\mathbf{B}$ have negative real parts. The eigenvalues are actually real, because

$$\begin{aligned}
\mathsf{tr}[\mathbf{B}]^2 - 4\mathsf{det}[\mathbf{B}] \;&=\; (\rho - \mathsf{tr}[\sqrt{\Xi}])^2 - 4\mathsf{det}[\sqrt{\Xi}] + 2\rho\mathsf{tr}[\sqrt{\Xi}] - \rho^2 = \mathsf{tr}[\sqrt{\Xi}]^2 - 4\mathsf{det}[\sqrt{\Xi}] \\
&=\; (\sqrt{\xi_1} + \sqrt{\xi_2})^2 - 4\sqrt{\xi_1\xi_2} \;\geq\; 0.
\end{aligned}$$

When $\xi_1 \neq \xi_2$, $\mathbf{B}$ has linearly independent eigenvectors; when $\xi_1 = \xi_2$, from Lemma A.1.2(iii), $\mathbf{F}$ and $\mathbf{B}$ are diagonal matrices, and $\mathbf{B}$ also has linearly independent eigenvectors.

We show in passing that $\mathbf{B}_1 = (\mathbf{B} - \rho\mathbf{I})^{-1}$ also has negative real eigenvalues. This will be used later. First, by (2.39),

$$\mathsf{det}[\mathbf{B}_1] = \big(\mathsf{det}[\mathbf{B} - \rho\mathbf{I}]\big)^{-1} = \big(\rho^2 - \mathsf{tr}[\mathbf{B}]\rho + \mathsf{det}[\mathbf{B}]\big)^{-1} > 0, \tag{A.24}$$

Secondly,

$$\mathsf{tr}[\mathbf{B}_1] = \mathsf{tr}[(\mathbf{B} - \rho\mathbf{I})^{-1}] = \frac{\mathsf{tr}[\mathbf{B} - \rho\mathbf{I}]}{\mathsf{det}[\mathbf{B} - \rho\mathbf{I}]} = \frac{\mathsf{tr}[\mathbf{B}] - 2\rho}{\mathsf{det}[\mathbf{B} - \rho\mathbf{I}]} < 0, \tag{A.25}$$

where the second equality is because the matrices here are all two by two matrices. Thirdly,

$$\mathsf{tr}[\mathbf{B}_1]^2 - 4\mathsf{det}[\mathbf{B}_1] \;=\; \frac{\big(\mathsf{tr}[\mathbf{B}] - 2\rho\big)^2 - 4\big(\rho^2 - \mathsf{tr}[\mathbf{B}]\rho + \mathsf{det}[\mathbf{B}]\big)}{\mathsf{det}[\mathbf{B} - \rho\mathbf{I}]^2} \;=\; \frac{\mathsf{tr}[\mathbf{B}]^2 - 4\mathsf{det}[\mathbf{B}]}{\mathsf{det}[\mathbf{B} - \rho\mathbf{I}]^2} \geq 0.$$

Hence, both $\mathbf{B}$ and $\mathbf{B}_1$ have negative real eigenvalues.

3. Now, we prove that $\mathbf{B}^2 - \rho\mathbf{B} = \mathbf{CA}$ has a unique solution that satisfies (2.39). Suppose $\mathbf{B}_a$ and $\mathbf{B}_b$ are two such solutions. Then,

$$\mathbf{CA} = \; \mathbf{B}_b^2 - \rho\mathbf{B}_b = \; \mathbf{B}_b(\mathbf{B}_b^2 - \rho\mathbf{B}_b)\mathbf{B}_b^{-1} = \; \mathbf{B}_b(\mathbf{B}_a^2 - \rho\mathbf{B}_a)\mathbf{B}_b^{-1} = \; \big(\mathbf{B}_b\mathbf{B}_a\mathbf{B}_b^{-1}\big)^2 - \rho\mathbf{B}_b\mathbf{B}_a\mathbf{B}_b^{-1}.$$

Thus, $\mathbf{B}_c \equiv \mathbf{B}_b\mathbf{B}_a\mathbf{B}_b^{-1}$ is also a solution to $\mathbf{B}^2 - \rho\mathbf{B} = \mathbf{CA}$ and it also satisfies (2.39): $\mathsf{tr}(\mathbf{B}_c) =$

$\mathrm{tr}(\mathbf{B}_a) < 0$ and $\det(\mathbf{B}_c) = \det(\mathbf{B}_a) > 0$. Write $\mathbf{B}^2 - \rho\mathbf{B} = \mathbf{CA} \equiv \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}$ in component form:

$$\begin{aligned} b_{11}^2 - \rho b_{11} + b_{12}b_{21} &= C_{11}, & b_{22}^2 - \rho b_{22} + b_{12}b_{21} &= C_{22}, \\ b_{12}(\mathrm{tr}[\mathbf{B}] - \rho) &= C_{12}, & b_{21}(\mathrm{tr}[\mathbf{B}] - \rho) &= C_{21}. \end{aligned} \tag{A.26}$$

The last equation in (A.26) implies that $b_{c21}(\mathrm{tr}[\mathbf{B}_c] - \rho) = b_{a21}(\mathrm{tr}[\mathbf{B}_a] - \rho)$. Since $\mathrm{tr}(\mathbf{B}_c) = \mathrm{tr}(\mathbf{B}_a) < 0$, we must have $b_{c21} = b_{a21}$. Similarly, we have $b_{c12} = b_{a12}$. Next we show that the diagonal elements of $\mathbf{B}_c$ and $\mathbf{B}_a$ are equal as well.

Suppose $b_{c11} \neq b_{a11}$ are two different solutions to the first equation in (A.26). As $\mathrm{tr}(\mathbf{B}_c) = \mathrm{tr}(\mathbf{B}_a)$, we have $b_{c22} \neq b_{a22}$, which are two different solutions to the second equation in (A.26). Thus, $b_{c11} + b_{a11} = \frac{\rho}{2}$ and $b_{c22} + b_{a22} = \frac{\rho}{2}$. Since $\mathrm{tr}[\mathbf{B}_c] = \mathrm{tr}[\mathbf{B}_a]$, we must have $b_{c11} = b_{a22}$ and $b_{a11} = b_{c22}$. Consequently, the first two equations in (A.26) have the same set of roots, implying $C_{11} = C_{22}$. This leads to $b_{a11}^2 - \rho b_{a11} = b_{a22}^2 - \rho b_{a22}$ or $(\mathrm{tr}[\mathbf{B}_a] - \rho)(b_{a11} - b_{a22}) = 0$. Since $\mathrm{tr}[\mathbf{B}_a] < 0$, we must have $b_{a11} = b_{a22}$. Similarly, we have $b_{c11} = b_{c22}$, but then $b_{c11} \neq b_{a11}$ implies $\mathrm{tr}[\mathbf{B}_c] \neq \mathrm{tr}[\mathbf{B}_a]$, a contradiction.

Hence, we have $\mathbf{B}_a = \mathbf{B}_c \equiv \mathbf{B}_b\mathbf{B}_a\mathbf{B}_b^{-1}$. This implies that $\mathbf{B}_a$ and $\mathbf{B}_b$ commute. Therefore,

$$e^{\mathbf{B}_a t}e^{\mathbf{B}_b t} = e^{(\mathbf{B}_a + \mathbf{B}_b)t}, \qquad \forall t \geq 0.$$

Since both $\mathbf{B}_a$ and $\mathbf{B}_b$ are stable matrices, the above quantity approaches to zero matrix as $t \longrightarrow \infty$, implying $\mathbf{B}_a + \mathbf{B}_b$ is also stable. Consequently, $\mathbf{B}_a + \mathbf{B}_b - \rho\mathbf{I}$ is a non-singular matrix.

Now notice that $\mathbf{0} = \mathbf{B}_a^2 - \rho\mathbf{B}_a - \mathbf{B}_b^2 + \rho\mathbf{B}_b = (\mathbf{B}_a - \mathbf{B}_b)(\mathbf{B}_a + \mathbf{B}_b - \rho\mathbf{I})$. As $\mathbf{B}_a + \mathbf{B}_b - \rho\mathbf{I}$ is non-singular, we must have $\mathbf{B}_a - \mathbf{B}_b = \mathbf{0}$. This proves the uniqueness of $\mathbf{B}$.

4. We prove that $\mathbf{D} = \mathbf{C}_1\mathbf{I}_1 - \mathbf{B}_1\mathbf{DK}$ uniquely determines $\mathbf{D}$. Horn and Johnson (1991) Chapter 4 presents the Kronecker products representation for the matrix equation. Thereby we can rewrite the above matrix equation in vector form:

$$\mathrm{vec}[\mathbf{D}] = \mathrm{vec}[\mathbf{C}_1\mathbf{I}_1] - (\mathbf{K}^\top \otimes \mathbf{B}_1)\mathrm{vec}[\mathbf{D}].$$

Let ζ_1 and $\zeta_2 < 0$ be the eigenvalues of $\mathbf{B}_1$, and let λ_1 and λ_2 be the eigenvalues of $\mathbf{K}$. Then the eigenvalues of $\mathbf{K}^\top \otimes \mathbf{B}_1$ are $\zeta_1\lambda_1$, $\zeta_1\lambda_2$, $\zeta_2\lambda_1$ and $\zeta_2\lambda_2$ (see Horn and Johnson 1991). We have shown at the end of step 2 that $\zeta_1 < 0$ and $\zeta_2 < 0$. We assumed in the specification of the economy that the real parts of λ_1 and λ_2 are negative. Thus, all the eigenvalues of $\mathbf{K}^\top \otimes \mathbf{B}_1$ have positive real parts, and therefore $\det[\mathbf{K}^\top \otimes \mathbf{B}_1 + \mathbf{I}] \neq 0$. Hence, $\mathbf{D}$ is uniquely determined from the above equation.

5. We now derive the equilibrium processes. The equilibrium price process is already given in (2.38) with $\mathbf{B}$ and $\mathbf{D}$ solved above. The first equation of (2.40) implies

$$\mathbf{p}_t = \mathbf{C}_1\big(\mathbf{X}_t + \mathbf{C}^{-1}(\mathbf{B}\mathbf{m} - \mathbf{c})\big) - \mathbf{D}_1\mathbf{k}_t. \tag{A.27}$$

Substituting (A.27) into (2.41), we have

$$\frac{d\mathbf{X}_t}{dt} = \mathbf{A}\mathbf{C}_1\big(\mathbf{X}_t + \mathbf{C}^{-1}(\mathbf{B}\mathbf{m} - \mathbf{c})\big) + \mathbf{a} + (\mathbf{I}_1 - \mathbf{A}\mathbf{D}_1)\mathbf{k}_t.$$

The first equation in (2.43) is equivalent to $\mathbf{B} = \mathbf{C}_1\mathbf{A}$. From Lemma A.1.3(i), $\mathbf{C}_1$ is symmetric, and therefore, $\mathbf{B}^\top = \mathbf{A}\mathbf{C}_1$. Using this relation, we have $\mathbf{I} = \mathbf{A}\mathbf{C}_1\mathbf{C}_1^{-1}\mathbf{A}^{-1} = \mathbf{B}^\top\mathbf{C}^{-1}(\mathbf{B} - \rho\mathbf{I})\mathbf{A}^{-1}$. Then, the equilibrium inventory process becomes:

$$\begin{aligned}\frac{d\mathbf{X}_t}{dt} &= \mathbf{B}^\top\big(\mathbf{X}_t - \mathbf{C}^{-1}(\mathbf{c} - \mathbf{B}\mathbf{m})\big) - \mathbf{B}^\top\mathbf{C}^{-1}(\mathbf{B} - \rho\mathbf{I})\mathbf{m} + (\mathbf{I}_1 - \mathbf{A}\mathbf{D}_1)\mathbf{k}_t. \\ &= \mathbf{B}^\top\big(\mathbf{X}_t - \mathbf{m}_X\big) + (\mathbf{I}_1 - \mathbf{A}\mathbf{D}_1)\mathbf{k}_t,\end{aligned}$$

where $\mathbf{m}_X = \mathbf{C}^{-1}(\mathbf{c} - \rho\mathbf{m})$. This, together with (A.27), implies

$$\begin{aligned}\mathbf{p}_t &= \mathbf{C}_1\big(\mathbf{X}_t - \mathbf{C}^{-1}(\mathbf{c} - \rho\mathbf{m}) + \mathbf{C}^{-1}(\mathbf{B} - \rho\mathbf{I})\mathbf{m}\big) - \mathbf{D}_1\mathbf{k}_t \\ &= \mathbf{C}_1\big(\mathbf{X}_t - \mathbf{m}_X\big) + \mathbf{m} - \mathbf{D}_1\mathbf{k}_t.\end{aligned}$$

The equilibrium production, input and output rates can be derived directly from (2.23), (2.24) and the second equation of (2.40).

6. Finally, we show that the equilibrium $(\mathbf{p}_t, \mathbf{X}_t, \mathbf{\Pi}_t)$ is uniformly bounded for all $t \geq 0$. Consider the process $\mathbf{z}_t = \mathbf{p}_t + \mathbf{D}_1\mathbf{k}_t$. From (2.35) and (2.42), $\mathbf{z}$ is given by

$$d\mathbf{z}_t = \big[\mathbf{B}(\mathbf{z}_t - \mathbf{m}) + (\mathbf{C}_1\mathbf{I}_1 - \mathbf{B}\mathbf{D}_1)\mathbf{k}_t\big]dt.$$

The solution for $\mathbf{z}_t$ is:

$$\mathbf{z}_t = e^{\mathbf{B}t}\mathbf{z}_0 + \int_0^t e^{\mathbf{B}(t-u)}(-\mathbf{B}\mathbf{m} + (\mathbf{C}_1\mathbf{I}_1 - \mathbf{B}\mathbf{D}_1)\mathbf{k}_u)du,$$

where $\mathbf{z}_0 = \mathbf{p}_0 + \mathbf{D}_1\mathbf{k}_0$. Since $\mathbf{k}_t$ is uniformly bounded, if we can show that $\mathbf{z}_t$ is uniformly bounded, then $\mathbf{p}_t$ is uniformly bounded as well.

Let $\|\mathbf{v}\|$ denote the super norm of $\mathbf{v} \in \mathcal{Z}^2$, the complex space, and let $\|\mathbf{Q}\| = \sup\{\|\mathbf{Q}\mathbf{v}\| :$

$\|\mathbf{v}\| \leq 1, \mathbf{v} \in \mathcal{Z}^2\}$ be the norm of linear transformation $\mathbf{Q}$ on on $\mathcal{Z}^2$. Then, we have

$$\begin{aligned}
\|\mathbf{z}_t\| &\leq \|e^{\mathbf{B}t}\mathbf{z}_0\| + \left\| \int_0^t e^{\mathbf{B}(t-u)}(-\mathbf{Bm} + (\mathbf{C}_1\mathbf{I}_1 - \mathbf{BD}_1)\mathbf{k}_u)du \right\| \\
&\leq \|e^{\mathbf{B}t}\|\|\mathbf{z}_0\| + \int_0^t \|e^{\mathbf{B}(t-u)}\|\| - \mathbf{Bm} + (\mathbf{C}_1\mathbf{I}_1 - \mathbf{BD}_1)\mathbf{k}_u\|du \\
&\leq \|e^{\mathbf{B}t}\|\|\mathbf{z}_0\| + C_3 \int_0^t \|e^{\mathbf{B}u}\|du
\end{aligned}$$

where C_3 is a constant such that $\| - \mathbf{Bm} + (\mathbf{C}_1\mathbf{I}_1 - \mathbf{BD}_1)\mathbf{k}_t\| \leq C_3$ for all $t \geq 0$.

We show in step 2 that $\mathbf{B}$ has two linearly independent eigenvectors $\widetilde{\mathbf{V}} = [\widetilde{\mathbf{v}}_1, \widetilde{\mathbf{v}}_2]$, with corresponding eigenvalues $\widetilde{\Xi} = \mathsf{diag}[\widetilde{\xi}_1, \widetilde{\xi}_2]$. Then, $e^{\mathbf{B}t} = \widetilde{\mathbf{V}}e^{\widetilde{\Xi}t}\widetilde{\mathbf{V}}^{-1}$, and we have

$$\int_0^t \|e^{\mathbf{B}u}\|du \leq \int_0^t \|\widetilde{\mathbf{V}}\|\|e^{\widetilde{\Xi}u}\|\|\widetilde{\mathbf{V}}^{-1}\|du \leq \|\widetilde{\mathbf{V}}\|\|\widetilde{\mathbf{V}}^{-1}\| \int_0^t \left|e^{\xi_1 u}\right| + \left|e^{\xi_2 u}\right|du \leq C_4, \quad \forall t \geq 0.$$

There also exists C_5 such that $\|e^{\mathbf{B}t}\| \leq C_5$ for all $t \geq 0$. Hence,

$$\|\mathbf{z}_t\| \leq C_5\|\mathbf{z}_0\| + C_3C_4, \qquad \forall t \geq 0.$$

∎

The proof of Theorem 2.2 used part (i) of the following lemma. The rest of the lemma will be used in the proof of Theorem 2.3.

Lemma A.1.3 *Under a rational expectations equilibrium,*

(i) $\mathbf{C}_1 = \mathbf{C}_1^\mathsf{T}$, $\mathsf{det}[\mathbf{C}_1] > 0$, $\mathbf{C}_1 < 0$;

(ii) $(\mathbf{B}_1 - r\mathbf{I})^{-1}\mathbf{C}_1 > 0$ *for any* $r > 0$;

(iii) $\mathbf{B}_1(\mathbf{B}_1 - r\mathbf{I})^{-1}\mathbf{C}_1 < 0$ *for any* $r > 0$;

(iv) $\mathbf{BC}_1$ *and* $\mathbf{BC}$ *are symmetric matrices;*

(v) If $\mathbf{B}$ *has negative diagonal elements, then* $\mathbf{C}^{-1}(\mathbf{B}_1 - r\mathbf{I})^{-1}\mathbf{C}_1$ *has positive diagonal elements for any* $r > 0$.

Proof.

(i) The first condition in (2.43) is equivalent to $\mathbf{B} = \mathbf{C}_1\mathbf{A}$. By the definitions of $\mathbf{B}_1$ and $\mathbf{C}_1$, we have $\mathbf{B} = \mathbf{C}\mathbf{C}_1^{-1} + \rho\mathbf{I}$. Thus, we have $\mathbf{C}_1\mathbf{A} = \mathbf{C}\mathbf{C}_1^{-1} + \rho\mathbf{I}$. Let $\mathbf{C}_1 = [\widetilde{c}_{ij}]$, then this equation can be written as

$$\begin{bmatrix} \widetilde{c}_{11} & \widetilde{c}_{12} \\ \widetilde{c}_{21} & \widetilde{c}_{22} \end{bmatrix} \mathbf{A} = \frac{1}{\mathsf{det}[\mathbf{C}_1]}\mathbf{C} \begin{bmatrix} \widetilde{c}_{22} & -\widetilde{c}_{12} \\ -\widetilde{c}_{21} & \widetilde{c}_{11} \end{bmatrix} + \rho\mathbf{I}.$$

Regarding $\det[\mathbf{C}_1]$ as a parameter, we can solve the above linear system for $\widetilde{c}_{ij}$:

$$\begin{aligned}\widetilde{c}_{11} &= \left(c_{11} + (A_2 + \tfrac{1}{a_2})\det[\mathbf{C}_1]\right)\rho\Delta,\\ \widetilde{c}_{22} &= \left(c_{22} + (A_1 + \tfrac{1}{a_2})\det[\mathbf{C}_1]\right)\rho\Delta,\\ \widetilde{c}_{12} = \widetilde{c}_{21} &= \left(c_{12} + \tfrac{1}{a_2}\det[\mathbf{C}_1]\right)\rho\Delta,\end{aligned}$$

where $\Delta = \left(\det[\mathbf{C}_1]\det[\mathbf{A}] - \det[\mathbf{C}]\det[\mathbf{C}_1]^{-1}\right)^{-1}$. Thus, $\mathbf{C}_1$ is symmetric. Notice that $\Delta^{-1} = \det[\mathbf{B}] - \det[\mathbf{B}^{-1}] = \det[\mathbf{B}] - \det[\mathbf{B} - \rho\mathbf{I}] = -\rho^2 + \mathrm{tr}[\mathbf{B}]\rho < 0$, as $\mathrm{tr}[\mathbf{B}] < 0$ due to (2.39). We also have $\det[\mathbf{C}_1] = \det[\mathbf{B}_1]\det[\mathbf{C}] > 0$, which follows from (A.24). As all the other parameters in the above expressions for $\widetilde{c}_{ij}$ are all positive, we conclude that $\mathbf{C}_1 < 0$.

(ii) For two by two matrix, we have $(\mathbf{B}_1 - r\mathbf{I})^{-1} = \dfrac{\mathbf{B}_1^{-1}\det[\mathbf{B}_1] - r\mathbf{I}}{\det[\mathbf{B}_1 - r\mathbf{I}]}$. Then,

$$(\mathbf{B}_1 - r\mathbf{I})^{-1}\mathbf{C}_1 = \frac{\det[\mathbf{B}_1]\mathbf{C} - r\mathbf{C}_1}{r^2 - \mathrm{tr}[\mathbf{B}_1]r + \det[\mathbf{B}_1]} > 0,$$

where the last inequality follows from (A.24), (A.25), $\mathbf{C} > 0$ and $\mathbf{C}_1 < 0$.

(iii) By the definition of $\mathbf{B}_1$ and using equation $\mathbf{B} = \mathbf{C}_1\mathbf{A}$, we have

$$\mathbf{B}_1(\mathbf{B}_1 - r\mathbf{I})^{-1}\mathbf{C}_1 = \left(\mathbf{C}_1^{-1} - r\mathbf{C}_1^{-1}\mathbf{B}_1^{-1}\right)^{-1} = \left((1+r\rho)\mathbf{C}_1^{-1} - r\mathbf{A}\right)^{-1}.$$

Part (i) implies that the elements of $\mathbf{C}_1^{-1}$ have signs $\begin{bmatrix} - & + \\ + & - \end{bmatrix}$. Then the elements of $(1 + r\rho)\mathbf{C}_1^{-1} - r\mathbf{A}$ also have signs $\begin{bmatrix} - & + \\ + & - \end{bmatrix}$. Furthermore, $\det[(1+r\rho)\mathbf{C}_1^{-1} - r\mathbf{A}] = \det[\mathbf{C}_1^{-1}]\det[(1+r\rho)\mathbf{I} - r\mathbf{B}] > 0$ as $\det[\mathbf{C}_1] > 0$ and $\det[\mathbf{B} - r\mathbf{I}] > 0$ for any $r > 0$. Hence, $\mathbf{B}_1(\mathbf{B}_1 - r\mathbf{I})^{-1}\mathbf{C}_1 < 0$.

(iv) $\mathbf{B}\mathbf{C}_1 = (\mathbf{B}_1^{-1} + \rho\mathbf{I})\mathbf{C}_1 = \mathbf{C} + \rho\mathbf{C}_1 = (\mathbf{B}_1^{-1}\mathbf{C}_1)^\top + \rho\mathbf{C}_1 = \mathbf{C}_1(\mathbf{B}^\top - \rho\mathbf{I}) + \rho\mathbf{C}_1 = \mathbf{C}_1\mathbf{B}^\top$.
$\mathbf{B}\mathbf{C} = \mathbf{B}\mathbf{B}_1^{-1}\mathbf{C}_1 = \mathbf{B}_1^{-1}\mathbf{B}\mathbf{C}_1 = \mathbf{B}_1^{-1}\mathbf{C}_1\mathbf{B}^\top = \mathbf{C}\mathbf{B}^\top$.

(v) Using the definition of $\mathbf{C}_1$ and (iv), we have $\mathbf{C}^{-1}(\mathbf{B}_1 - r\mathbf{I})^{-1}\mathbf{C}_1 = \left(\mathbf{C}_1^{-1}(\mathbf{B}_1 - r\mathbf{I})\mathbf{C}\right)^{-1} = \left(\mathbf{I} - r\mathbf{C}^{-1}\mathbf{B}_1^{-1}\mathbf{C}\right)^{-1} = \left((1+r\rho)\mathbf{I} - r\mathbf{C}^{-1}\mathbf{B}\mathbf{C}\right)^{-1} = \left((1+r\rho)\mathbf{I} - r\mathbf{B}^\top\right)^{-1}$. Now if $\mathbf{B}$ has negative diagonal elements, then $(1 + r\rho)\mathbf{I} - r\mathbf{B}^\top$ has positive diagonal elements. Since it has positive determinant, $\left((1 + r\rho)\mathbf{I} - r\mathbf{B}^\top\right)^{-1}$ has positive diagonal elements.

∎

Proof of Theorem 2.3. Let $\mathbf{C}_1 = [\widetilde{\mathbf{c}}_1, \widetilde{\mathbf{c}}_2]$. Then, the solution for $\mathbf{D}$ in Theorem 2.2(ii) becomes:

$$\mathrm{vec}[\mathbf{D}] = \begin{bmatrix} -\kappa_1\mathbf{B}_1 + \mathbf{I} & 0 \\ 0 & -\kappa_2\mathbf{B}_1 + \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \widetilde{\mathbf{c}}_1 \\ -\widetilde{\mathbf{c}}_2 \end{bmatrix} = \begin{bmatrix} -(\kappa_1\mathbf{B}_1 - \mathbf{I})^{-1}\widetilde{\mathbf{c}}_1 \\ (\kappa_2\mathbf{B}_1 - \mathbf{I})^{-1}\widetilde{\mathbf{c}}_2 \end{bmatrix}.$$

Using Lemma A.1.3(ii), the elements of $\mathbf{D}$ have signs $\begin{bmatrix} - & + \\ - & + \end{bmatrix}$. Using Lemma A.1.3(iii) and $\mathbf{D}_1 = \mathbf{B}_1\mathbf{D}$, the elements of $\mathbf{D}_1$ have signs $\begin{bmatrix} + & - \\ + & - \end{bmatrix}$. Using Lemma A.1.3(v), the elements of $\mathbf{C}^{-1}\mathbf{D}$ have signs $\begin{bmatrix} - & \\ & + \end{bmatrix}$.

(i) Let $\Delta_1 > 0$ and $\Delta_2 > 0$ denote positive supply and demand shocks, respectively. Let $\mathbf{D}_1 = [\widetilde{d}_{ij}]$. Then, the impact on raw material price is $-\widetilde{d}_{11}\Delta_1 - \widetilde{d}_{12}\Delta_2$ and the impact on finished goods price is $-\widetilde{d}_{21}\Delta_1 - \widetilde{d}_{22}\Delta_2$. Part (i) follow immediately from the signs of $\widetilde{d}_{ij}$.

(ii) The sign of $\det[\mathbf{D}_1]$ can be either positive or negative (numerical result). When $\det[\mathbf{D}_1] < 0$, or equivalently $-\frac{\widetilde{d}_{12}}{\widetilde{d}_{11}} < -\frac{\widetilde{d}_{22}}{\widetilde{d}_{21}}$, and the demand and supply shock satisfies $-\frac{\widetilde{d}_{12}}{\widetilde{d}_{11}} < \frac{\Delta_1}{\Delta_2} < -\frac{\widetilde{d}_{22}}{\widetilde{d}_{21}}$, we have $-\widetilde{d}_{11}\Delta_1 - \widetilde{d}_{12}\Delta_2 < 0$ and $-\widetilde{d}_{21}\Delta_1 - \widetilde{d}_{22}\Delta_2 > 0$.

(iii) The result follows immediately from the signs of $\mathbf{D}$.

(iv) Using the equilibrium condition (2.43), we have $\mathbf{I}_1 - \mathbf{A}\mathbf{D}_1 = \mathbf{C}_1^{-1}(\mathbf{C}_1\mathbf{I}_1 - \mathbf{B}\mathbf{D}_1) = \mathbf{C}_1^{-1}((\mathbf{B}_1^{-1} - \mathbf{B})\mathbf{D}_1 + \mathbf{D}_1\mathbf{K}) = \mathbf{C}_1^{-1}\mathbf{D}_1(\mathbf{K} - \rho\mathbf{I}) = \mathbf{C}^{-1}\mathbf{D}(\mathbf{K} - \rho\mathbf{I})$. $\mathbf{K} - \rho\mathbf{I}$ is a diagonal matrix with negative diagonal elements. Thus, the signs of $\mathbf{C}^{-1}\mathbf{D}$ implies that the elements of $\mathbf{I}_1 - \mathbf{A}\mathbf{D}_1$ have signs $\begin{bmatrix} + & \\ & - \end{bmatrix}$.

∎

A.2 Proofs of Lemmas and Propositions in Chapter 3

Proof of Lemma 3.1. (This lemma follows from well-known results regarding optional stopping applied to Brownian motion; refer to, e.g., Karlin and Taylor (1975), and Ross (1996). The proof here is included for completeness.) Omit the subscript n, and write $T := \tau_n$, $x := S/N$. We have

$$\lambda T + \sigma B(T) = x. \tag{A.28}$$

This, combined with $\mathsf{E}B(T) = 0$, which follows from applying optional stopping to $B(t)$, leads to $\mathsf{E}[T] = x/\lambda$.

To derive the second moment of T, we have

$$\sigma^2[B^2(T) - T] = (x - \lambda T)^2 - \sigma^2 T.$$

Applying optional stopping again, to the martingale $B^2(t) - t$, we have

$$\mathsf{E}[(x - \lambda T)^2] = \sigma^2 \mathsf{E}[T],$$

which can be expressed as

$$\lambda^2 \mathsf{Var}(T) = \sigma^2 \mathsf{E}[T].$$

This establishes

$$\mathsf{E}[T^2] \;=\; \frac{\sigma^2}{\lambda^2}\mathsf{E}[T] + \mathsf{E}^2[T] \;=\; \frac{\sigma^2 x}{\lambda^3} + \frac{x^2}{\lambda^2}.$$

Now consider $\mathsf{E}[e^{-\gamma T}]$, where $\gamma > 0$ is a constant (discount rate). Write

$$-\gamma T = aB(T) - \frac{1}{2}a^2 T - bx, \tag{A.29}$$

where a and b are parameters to be determined. Then, applying optional stopping to the exponential martingale, $e^{aB(t) - \frac{1}{2}a^2 t}$, we have

$$\mathsf{E}[e^{-\gamma T}] = e^{-bx} \cdot \mathsf{E}[e^{aB(T) - \frac{1}{2}a^2 T}] = e^{-bx}.$$

The power b can be derived as follows. Substituting (A.28) into the right hand side of (A.29), we have

$$-\gamma T = aB(T) - \frac{1}{2}a^2 T - b\sigma B(T) - b\lambda T.$$

Equating the coefficients of T and $B(T)$ on both sides yields:

$$a = b\sigma, \qquad \frac{1}{2}a^2 + b\lambda = \gamma,$$

leading to

$$\sigma a^2 + 2\lambda a - 2\gamma\sigma = 0.$$

Hence,

$$a = \frac{\sqrt{\lambda^2 + 2\sigma^2\gamma} - \lambda}{\sigma};$$

and

$$\mathsf{E}[e^{-\gamma T}] = e^{-bx} = \exp[-\frac{\sqrt{\lambda^2 + 2\sigma^2\gamma} - \lambda}{\sigma^2}x].$$

∎

Proof of Lemma 3.2. First, assume interior optimum. The first-order conditions in (3.11) hold, which imply that

$$\frac{d\widetilde{v}_m(\mu_m)}{d\mu_m} = \frac{d\widetilde{v}_n(\mu_n)}{d\mu_n},$$

or equivalently,

$$\tfrac{dp}{d\mu}(\tfrac{1}{\mu_m^*}) - \tfrac{dp}{d\mu}(\tfrac{1}{\mu_n^*}) + \tfrac{hS}{N}(m-n) + h\sigma^2(\mu_n^* - \mu_m^*) = 0. \tag{A.30}$$

By Theorem 3.1, $\mu_n^* \geq \mu_m^*$ for $n > m$. Next, applying the mean-value theorem and using the definition of B, we have

$$\tfrac{dp}{d\mu}(\tfrac{1}{\mu_m^*}) - \tfrac{dp}{d\mu}(\tfrac{1}{\mu_n^*}) = -\tfrac{d^2p}{d\mu^2}(\tfrac{1}{\zeta})(\mu_n^* - \mu_m^*) \geq B(\mu_n^* - \mu_m^*), \tag{A.31}$$

where $\zeta \in [\mu_m^*, \mu_n^*]$. Combining (A.30) and (A.31), we have

$$\tfrac{hS}{N}(m-n) + (h\sigma^2 + B)(\mu_n^* - \mu_m^*) \leq 0,$$

or

$$\mu_n^* - \mu_m^* \leq \frac{S(n-m)}{N\left(\sigma^2 + \frac{B}{h}\right)}.$$

Next, we show the above result holds even when the optimum is on the boundary. If $\mu_m^* = \mu_n^*$, then the desired inequality obviously holds. If $\mu_m^* < \mu_n^*$, consider the following alternative pricing policy: $(\mu_1^*, \ldots, \mu_m^* + \delta, \ldots, \mu_n^* - \delta, \ldots, \mu_N^*)$, where $0 < \delta < \frac{\mu_n^* - \mu_m^*}{2}$. Since this alternative cannot be better than the optimum, we have

$$\begin{aligned} & p(\tfrac{1}{\mu_m^*+\delta}) + p(\tfrac{1}{\mu_n^*-\delta}) - \tfrac{hS}{N}\left[(N-m+\tfrac{1}{2})(\mu_m^*+\delta) + (N-n+\tfrac{1}{2})(\mu_n^*-\delta)\right] \\ & -\tfrac{h\sigma^2}{2}\left[(\mu_m^*+\delta)^2 + (\mu_n^*-\delta)^2\right] \\ \leq\ & p(\tfrac{1}{\mu_m^*}) + p(\tfrac{1}{\mu_n^*}) - \tfrac{hS}{N}\left[(N-m+\tfrac{1}{2})\mu_m^* + (N-n+\tfrac{1}{2})\mu_n^*\right] - \tfrac{h\sigma^2}{2}\left[\mu_m^{*2} + \mu_n^{*2}\right]. \end{aligned}$$

Taylor's expansion with straightforward algebra simplifies the inequality to

$$\tfrac{dp}{d\mu}(\tfrac{1}{\mu_m^*+\zeta_1\delta})\delta - \tfrac{dp}{d\mu}(\tfrac{1}{\mu_n^*-\zeta_2\delta})\delta + \tfrac{hS}{N}\delta(m-n) + h\sigma^2\delta(\mu_n^* - \mu_m^* - \delta) \leq 0,$$

where $\zeta_1, \zeta_2 \in [0,1]$. Dividing both sides by δ, and then letting $\delta \to 0$, we have

$$\tfrac{dp}{d\mu}(\tfrac{1}{\mu_m^*}) - \tfrac{dp}{d\mu}(\tfrac{1}{\mu_n^*}) + \tfrac{hS}{N}(m-n) + h\sigma^2(\mu_n^* - \mu_m^*) \leq 0. \tag{A.32}$$

Combining (A.31) and (A.32) yields the desired inequality. ■

Proof of Proposition 3.3. We shall write $\boldsymbol{\mu}^*(\sigma)$, $\boldsymbol{\mu}^*(h)$ and $\boldsymbol{\mu}^*(S)$ to emphasize the dependence of the maximizer of (3.10) on these parameters. First we prove the following lemma.

Lemma A.2.1 *Assuming interior optimum,*

(i) $\boldsymbol{\mu}^*(\sigma)$ *is differentiable and* $\frac{d\mu_n^*(\sigma)}{d\sigma} < 0$ *if and only if*

$$\sum_{k=1}^{N} \mu_k^*(\sigma)^2 - 2\mu_n^*(\sigma) \sum_{k=1}^{N} \mu_k^*(\sigma) \quad < \quad 0; \tag{A.33}$$

(ii) $\boldsymbol{\mu}^*(h)$ *is differentiable and* $\frac{d\mu_n^*(h)}{dh} < 0$ *if and only if*

$$\frac{S}{N} \sum_{k=1}^{N} (n-k)\mu_k^* + \sigma^2 \sum_{k=1}^{N} \left(\frac{\mu_k^{*2}}{2} - \mu_n^* \mu_k^* \right) \quad < \quad 0; \tag{A.34}$$

(iii) $\boldsymbol{\mu}^*(S)$ *is differentiable and* $\frac{d\mu_n^*(S)}{dS} < 0$ *if and only if*

$$\frac{h}{N^2} \sum_{k=1}^{N} (n-k)\mu_k^*(S) \quad < \quad -a'(S). \tag{A.35}$$

Proof. We only prove part (i); the proof of part (ii) and (iii) is completely analogous. Consider an equivalent problem to the fractional problem (3.10):

$$\max_{\boldsymbol{\mu}} \quad \sum_{n=1}^{N} \left[p\left(\frac{1}{\mu_n}\right) - \frac{hS(N-n+\frac{1}{2})}{N}\mu_n - \frac{h\sigma^2}{2}\mu_n^2 - a(S) \right] - \eta(\sigma) \sum_{n=1}^{N} \mu_n,$$

where $\eta(\sigma)$ is the optimal value of (3.10) when the demand variability is σ and all the other parameters are fixed. (This equivalence is discussed in Section 3.3.3.3.) To solve the above problem, we can maximize each μ_n separately:

$$\max_{\mu_n} \quad g_n(\mu_n, \sigma) \;=\; p\left(\frac{1}{\mu_n}\right) - \frac{hS(N-n+\frac{1}{2})}{N}\mu_n - \frac{h\sigma^2}{2}\mu_n^2 - \eta(\sigma)\mu_n.$$

Since $\boldsymbol{\mu}^*(\sigma)$ is in the interior, applying the envelope theorem, we have

$$\frac{d\eta}{d\sigma} \;=\; -\frac{h\sigma \sum_{k=1}^{N} \mu_k^{*2}}{\sum_{k=1}^{N} \mu_k^*}.$$

To establish the desired monotonicity, it suffices to establish the submodularity of $g_n(\mu_n, \sigma)$ (refer to Topkis 1978):

$$\begin{aligned}
\frac{\partial^2 g_n(\mu_n^*, \sigma)}{\partial \mu_n \partial \sigma} &= -2h\sigma\mu_n^* - \frac{\partial \eta}{\partial \sigma} \\
&= -2h\sigma\mu_n^* + \frac{h\sigma \sum_{k=1}^{N} \mu_k^{*2}}{\sum_{k=1}^{N} \mu_k^*}. \\
&= \frac{h\sigma}{\sum_{k=1}^{N} \mu_k^*} \left(\sum_{k=1}^{N} \mu_k^{*2} - 2\mu_n^* \sum_{k=1}^{N} \mu_k^* \right),
\end{aligned}$$

If $\sum \mu_k^{*2} - 2\mu_n^* \sum \mu_k^* < 0$, then the objective $g_n(\mu_n, \sigma)$ is submodular in (μ_n, σ) in a neighborhood of (μ_n^*, σ). Thus, μ_n^* is (locally) decreasing in σ. If the converse is true, μ_n^* is (locally) increasing in σ. ∎

Proof of Proposition 3.3. (i) By Lemma A.2.1, the inequality in (A.33) always holds for $n = N$, as μ_N^* is the largest component of $\boldsymbol{\mu}$ following Theorem 3.1. Hence μ_N^* is decreasing in σ.

(ii) Denote the left side of (A.34) by L_n. We will show that $L_1 + L_N < 0$ and $L_1 < L_N$, which then implies $L_1 < 0$.

$$\begin{aligned} L_1 + L_N &= \frac{S}{N}\sum_{k=1}^{N}(N+1-2k)\mu_k^* + \sigma^2 \sum_{k=1}^{N}\left[\mu_k^{*2} - (\mu_1^* + \mu_N^*)\mu_k^*\right] \\ &= \frac{S}{N}\left[(N-1)(\mu_1^* - \mu_N^*) + (N-3)(\mu_2^* - \mu_{N-1}^*) + \cdots + (\mu^*_{\lfloor \frac{N}{2} \rfloor} - \mu^*_{\lfloor \frac{N+3}{2} \rfloor})\right] \\ &\quad + \sigma^2 \sum_{k=1}^{N}\left[\mu_k^*(\mu_k^* - \mu_N^*) - \mu_1^*\mu_k^*\right] \\ &< 0, \end{aligned}$$

where the last inequality follows from μ_n^* increasing in n. Next, note that

$$L_1 - L_N = \left[\frac{S(1-N)}{N} + \sigma^2(\mu_N^* - \mu_1^*)\right]\sum_{k=1}^{N}\mu_k^* \leq 0,$$

where the last inequality follows from Lemma 3.2. Hence $L_1 < 0$, and $p_1^*(h)$ is decreasing in h.

(iii) By Lemma A.2.1, if $a'(S) < 0$ and $n = 1$, the inequality in (A.35) always holds. Hence μ_1^* is decreasing in S.

(iv) This part is obvious. ∎

Proof of Proposition 3.4. From Example 3.1, the optimal replenishment level for each λ is $S^*(\lambda) = \sqrt{2K\lambda/h}$. We replace S in the objective by $S^*(\lambda)$, and then find the optimal λ. Let

$$W(\lambda) := V(S^*(\lambda), \lambda) = (\alpha - \lambda)\lambda/\beta - c\lambda - \sqrt{2Kh\lambda} - \frac{h\sigma^2}{2\lambda}.$$

We prove that λ^* satisfying $W'(\lambda^*) = 0$ and $W(\lambda^*) > 0$ is the global maximizer.

Let $f(\lambda) := (\alpha - \lambda)\lambda/\beta - c\lambda - \sqrt{2Kh\lambda}$. It is easy to see that $f(0) = 0$, $f'(0) = -\infty$ and $f''(\lambda) = -\frac{2}{\beta} + \sqrt{\frac{hK}{8\lambda^3}}$. Let $\widehat{\lambda}$ be the inflection point, i.e., $f''(\widehat{\lambda}) = 0$. Then, $f(\lambda)$ starts decreasing from zero, is convex in $[0, \widehat{\lambda}]$ and concave in $[\widehat{\lambda}, \infty)$.

Suppose $f'(\widehat{\lambda}) \leq 0$, then by convexity/concavity, $f'(\lambda) \leq f'(\widehat{\lambda}) \leq 0$ for all $\lambda \geq 0$. Therefore,

$W(\lambda) < f(\lambda) \leq f(0) = 0$ for all $\lambda \geq 0$, so there exists no maximizer with positive profit in this case. Hence, we must have $f'(\widehat{\lambda}) > 0$.

Next, let λ_1 be the local minimizer for $f(\lambda)$ in the convex part of the function. Obviously, $\lambda_1 < \widehat{\lambda}$. First, $\lambda^* \notin [0, \lambda_1]$ because in this region $W(\lambda) < f(\lambda) \leq 0$. Second, $\lambda^* \notin [\lambda_1, \widehat{\lambda}]$ because in this region $W'(\lambda) = f'(\lambda) + \frac{h\sigma^2}{2\lambda^2} > 0$. Hence, $\lambda^* \in [\widehat{\lambda}, \alpha]$. ($\widehat{\lambda} < \alpha$ in order for λ^* to exist.) Since $W(\lambda) = f(\lambda) - \frac{h\sigma^2}{2\lambda}$ is strictly concave in $[\widehat{\lambda}, \alpha]$, λ^* is the unique maximizer for $W(\lambda)$ in $[\widehat{\lambda}, \alpha]$. It is the global maximizer since all other local maximizers (if any) must be within $[0, \lambda_1]$, the region where $W(\lambda) < 0$. ∎

Proof of Proposition 3.5. (i) The joint optimization problem can be solved sequentially. We first optimize S for each fixed $\lambda > 0$ (same as Section 3.3.1). From Proposition 3.1, the optimal S is invariant to σ, and we denote it by $S^*(\lambda)$. Then, the optimal λ can be found by

$$\max_{\lambda \in \mathcal{L}} \quad \widetilde{V}(\lambda, \sigma) \ := \ r(\lambda) - \frac{hS^*(\lambda)}{2} - \lambda a(S^*(\lambda)) - \frac{h\sigma^2}{2\lambda}.$$

Now, $\widetilde{V}(\lambda, \sigma)$ is supermodular in (λ, σ) since

$$\frac{\partial^2 \widetilde{V}}{\partial \lambda \partial \sigma} = \frac{h\sigma}{\lambda^2} \geq 0,$$

and therefore $\lambda^*(\sigma)$ is increasing in σ, or equivalently, the optimal price is decreasing in σ.

(ii) That S^* is increasing in σ follows immediately from (i) and Proposition 3.1 (i). To examine the effect of h on S^*, we first optimize λ for each fixed $S > 0$ (same as Section 3.3.2), and denote the maximizer by $\lambda^*(S, h)$. Then, the optimal S is determined by

$$\max_{S > 0} \quad \widetilde{V}(S, h) \ := \ r(\lambda^*(S, h)) - \frac{hS}{2} - \lambda^*(S, h) a(S) - \frac{h\sigma^2}{2\lambda^*(S, h)}.$$

Let λ^*_S, λ^*_h and λ^*_{Sh} denote the partial derivatives. We have

$$\begin{aligned}
\frac{\partial^2 \widetilde{V}}{\partial S \partial h} &= r''\lambda^*_S\lambda^*_h + r'\lambda^*_{Sh} - \frac{1}{2} - a\lambda^*_{Sh} - a'\lambda^*_h + \frac{\sigma^2}{2\lambda^{*2}}\lambda^*_S - \frac{h\sigma^2}{\lambda^{*3}}\lambda^*_S\lambda^*_h + \frac{h\sigma^2}{2\lambda^{*2}}\lambda^*_{Sh} \\
&= \lambda^*_S\Big(r''\lambda^*_h + \frac{\sigma^2}{2\lambda^{*2}} - \frac{h\sigma^2}{\lambda^{*3}}\lambda^*_h\Big) - \frac{1}{2} - a'\lambda^*_h + \Big(r' - a + \frac{h\sigma^2}{2\lambda^{*2}}\Big)\lambda^*_{Sh}
\end{aligned}$$

Since $\lambda^*(S, h)$ is uniquely determined by (3.9), the last term in the above is zero, and

$$\lambda^*_S = \frac{a'}{r'' - \frac{h\sigma^2}{\lambda^{*3}}}, \qquad \lambda^*_h = \frac{-\frac{\sigma^2}{2\lambda^{*2}}}{r'' - \frac{h\sigma^2}{\lambda^{*3}}},$$

Then, the first term in the above is also zero, and we have

$$\frac{\partial^2 \widetilde{V}}{\partial S \partial h} = -\frac{1}{2} - a' \lambda_h^* = -\frac{1}{2} + \frac{a' \frac{\sigma^2}{2\lambda^{*2}}}{r'' - \frac{h\sigma^2}{\lambda^{*3}}}$$

Now we show that the above is less than zero if evaluated at S^*. This is because S^* satisfies (3.6), which implies $a'(S^*) = -\frac{h}{2\lambda^*}$, and

$$\frac{\partial^2 \widetilde{V}}{\partial S \partial h} = -\frac{1}{2} + \frac{\frac{h\sigma^2}{4\lambda^{*3}}}{-r'' + \frac{h\sigma^2}{\lambda^{*3}}} < -\frac{1}{2} + \frac{\frac{h\sigma^2}{4\lambda^{*3}}}{\frac{h\sigma^2}{\lambda^{*3}}} < -\frac{1}{4}.$$

Hence, $\widetilde{V}(S, h)$ is submodular in (S, h) in the neighborhood of the optima, and therefore, $S^*(h)$ is decreasing in h.

(iii). This part is obvious. ∎

A.3 Extensions to the Model in Chapter 3

A.3.1 Discounted Objective

In this section, we consider our model under a discounted objective. The demand model, cost parameters and policy specifications are described in Section 3.2 of the chapter. Parallel to the analysis in Section 3.3, we first derive the profit functions, and then consider optimal replenishment and pricing decisions, with the other decision variable fixed or optimized jointly.

We will focus on the contrasts rather than the similarities between these two classes of models. So as to present the results with minimal distraction, all proofs are relegated to Section A.3.3.

Recall that period n refers to the period in which the price p_n applies and the inventory drops from $S_{n-1} = \frac{(N-n+1)S}{N}$ to $S_n = \frac{(N-n)S}{N}$. Let $\gamma > 0$ be the discount rate. Let $v_{n,\gamma}(\lambda_n)$ denote the expected profit over period n discounted to the beginning of the period. We have

$$\begin{aligned} v_{n,\gamma}(\lambda_n) &= \mathsf{E}\left[\int_0^{\tau_n} e^{-\gamma t}\Big[p(\lambda_n)dD(t+T_{n-1}) - hX(t+T_{n-1})dt\Big]\right] \\ &= \mathsf{E}\left[\int_0^{\tau_n} e^{-\gamma t}\Big(p(\lambda_n) + \frac{h}{\gamma}\Big)dD(t+T_{n-1}) + \frac{he^{-\gamma\tau_n}S_n}{\gamma} - \frac{hS_{n-1}}{\gamma}\right] \\ &= \Big(p(\lambda_n) + \frac{h}{\gamma}\Big)\frac{\lambda_n}{\gamma}\Big(1 - \mathsf{E}[e^{-\gamma\tau_n}]\Big) + \frac{h\mathsf{E}[e^{-\gamma\tau_n}]S_n}{\gamma} - \frac{hS_{n-1}}{\gamma} \\ &= f(\lambda_n) + \frac{hS(1-\Theta(\lambda_n))}{N\gamma}n, \end{aligned} \tag{A.36}$$

where,

$$\begin{aligned} f(\lambda) &= \Big(p(\lambda)+\frac{h}{\gamma}\Big)\frac{\lambda}{\gamma}\Big(1-\Theta(\lambda)\Big)+\frac{h\Theta(\lambda)S}{\gamma}-\frac{hS(N+1)}{N\gamma}, \\ \Theta(\lambda) &= \mathsf{E}[e^{-\gamma\tau_n(\lambda)}] = e^{-b(\lambda)S/N}, \\ b(\lambda) &= \frac{\sqrt{\lambda^2+2\sigma^2\gamma}-\lambda}{\sigma^2} = \frac{2\gamma}{\sqrt{\lambda^2+2\sigma^2\gamma}+\lambda}. \end{aligned}$$

Here Lemma 3.1 is applied to derive the expression for $\Theta(\lambda)$. Assuming that the optimal λ_n is finite, then $\Theta(\lambda_n) < 1$. Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_N) \in \mathcal{L}^N$, and let $V_\gamma(S, \boldsymbol{\lambda})$ denote the discounted profit starting from zero inventory under the policy $(S, \boldsymbol{\lambda})$. Then,

$$V_\gamma(S,\boldsymbol{\lambda}) = \frac{v_\gamma(S,\boldsymbol{\lambda})}{1-\Theta_\gamma(S,\boldsymbol{\lambda})}, \tag{A.37}$$

where,

$$\begin{aligned} v_\gamma(S,\boldsymbol{\lambda}) &= v_{1,\gamma}(\lambda_1)+\Theta(\lambda_1)v_{2,\gamma}(\lambda_2)+\cdots+\Theta(\lambda_1)\ldots\Theta(\lambda_{N-1})v_{N,\gamma}(\lambda_N)-c(S), \\ \Theta_\gamma(S,\boldsymbol{\lambda}) &= \Theta(\lambda_1)\ldots\Theta(\lambda_N) < 1. \end{aligned}$$

Note in the above expressions $v_{n,\gamma}$ and Θ depend on S as well, but we have suppressed S to simplify notation. For the same reason, we shall omit S in $v_\gamma(S, \boldsymbol{\lambda})$ and $\Theta_\gamma(S, \boldsymbol{\lambda})$ when S is not a decision variable.

For a single price, (A.37) becomes

$$V_\gamma(S,\lambda) = \frac{r(\lambda)}{\gamma} - \frac{S(h/\gamma+a(S))}{1-e^{-b(\lambda)S}} + \frac{h\lambda}{\gamma^2}. \tag{A.38}$$

Optimal Replenishment

The problem is

$$\max_{S>0} V_\gamma(S,\lambda).$$

The first-order condition is

$$[h/\gamma + a(S) + Sa'(S)](e^{bS}-1) = bS[h/\gamma + a(S)]. \tag{A.39}$$

Proposition A.3.1 *$V_\gamma(S, \lambda)$ is strictly concave in S if $a'(S) < 0$ and $c''(S) \geq 0$.*

The condition $a'(S) < 0$ holds if S satisfies (A.39). Hence, the first-order condition in (A.39) is sufficient for optimality if $c''(S) \geq 0$. We have also identified examples of multiple stationary points when $c''(S) < 0$. (This example is $c(S) = 2 + 10S - S^2$ for $S \in [0, 5]$, $\lambda(p) = 10 - p$, $\sigma = 0.5$, $h = 1$ and $\gamma = 0.3$. Two stationary points can be found.)

Proposition A.3.2 *With price fixed,*
(i) the optimal replenishment level S^*_γ is increasing in σ, decreasing in h and p;
(ii) the optimal profit is decreasing in σ and h.

When there are multiple optima, in lieu of increasing and decreasing in part (i), the relevant properties are ascending and descending, respectively.

Recall, under the average objective and fixed price, the demand variability has no effect on the optimal replenishment level. In contrast, the demand variability will raise the replenishment level under the discounted objective. (See the example provided in Section 3.4.1.) Other monotonicity properties in Proposition 3.1 continue to hold here.

The Single-Price Problem

The problem is

$$\max_{\lambda \in \mathcal{L}} V_\gamma(S, \lambda)$$

The first-order condition is

$$r'(\lambda) + \frac{hS + \gamma c(S)}{(1 - e^{-b(\lambda)S})^2} S e^{-b(\lambda)S} b'(\lambda) + \frac{h}{\gamma} \quad = \quad 0. \tag{A.40}$$

Proposition A.3.3 *$V_\gamma(S, \lambda)$ is strictly concave in λ.*

Hence the optimal single price is uniquely determined by the first-order condition.

Proposition A.3.4 *With the replenishment level fixed at S, let the optimal price be p^*. Then*
(i) p^* is decreasing in h;
(ii) p^* is decreasing in S for S satisfying $a'(S) \leq 0$;
(iii) the optimal profit is decreasing in σ and h.

Most results in Proposition 3.2 continue to hold here. The only difference is that the optimal price is decreasing in demand variability under the average objective, while here the effect of demand variability is rather unclear. Numerically, we have observed that in most cases the optimal price is still decreasing in σ under the discounted objective, but exceptions do exist.

The N-Price Problem

The problem is

$$\max_{\boldsymbol{\lambda} \in \mathcal{L}^N} V_\gamma(S, \boldsymbol{\lambda}) \tag{A.41}$$

This problem has an equivalent dynamic programming formulation. Let $V_{n,\gamma}$ denote the optimal discounted profit-to-go starting from the beginning of period n. Following the principle of optimality, we have the following recursions:

$$\begin{aligned} V_{1,\gamma} &= \max_{\lambda \in \mathcal{L}} \{-c(S) + v_{1,\gamma}(\lambda) + \Theta(\lambda) V_{2,\gamma}\}, \\ V_{n,\gamma} &= \max_{\lambda \in \mathcal{L}} \{v_{n,\gamma}(\lambda) + \Theta(\lambda) V_{n+1,\gamma}\}, \qquad n = 2, \ldots, N-1, \\ V_{N,\gamma} &= \max_{\lambda \in \mathcal{L}} \{v_{N,\gamma}(\lambda) + \Theta(\lambda) V_{1,\gamma}\}. \end{aligned} \tag{A.42}$$

The following result parallels Theorem 3.1 for the average objective in the chapter.

Proposition A.3.5 *For any fixed replenishment level S, the optimal prices under the discounted objective are increasing over the periods, i.e., $p_1^* \leq p_2^* \leq \ldots \leq p_N^*$.*

Instead of searching in $\mathcal{L}^N$ for the optimal $\boldsymbol{\lambda}$, we provide an iterative algorithm with each step solving a one-dimensional problem in $\mathcal{L}$.

`A Fixed-Point Algorithm for the dynamic program (A.42)`

1. Set an arbitrary initial value $V_{1,\gamma}^o$, and set $\epsilon > 0$ small.
2. Use $V_{1,\gamma}^o$ in the last equation of (A.42), and solve for $V_{N,\gamma}, V_{N-1,\gamma}, \ldots, V_{1,\gamma}$ recursively following (A.42).
3. If $|V_{1,\gamma}^o - V_{1,\gamma}| < \epsilon$, stop. Otherwise, let $V_{1,\gamma}^o \leftarrow V_{1,\gamma}$, and go to Step 2.

To explore the convergence of the algorithm, we substitute the equations for $V_{2,\gamma}, \ldots, V_{N,\gamma}$ in (A.42) recursively into the first equation and obtain the following:

$$V_{1,\gamma} = \max_{\boldsymbol{\lambda} \in \mathcal{L}^N} \left\{ v_\gamma(\boldsymbol{\lambda}) + \Theta_\gamma(\boldsymbol{\lambda}) V_{1,\gamma} \right\} \equiv \max_{\boldsymbol{\lambda} \in \mathcal{L}^N} \psi(\boldsymbol{\lambda}, V_{1,\gamma}) \equiv \Psi(V_{1,\gamma}). \tag{A.43}$$

Proposition A.3.6 *Let ψ and Ψ be defined as above, then*

(i) the mapping $\Psi(V)$ is increasing and Lipschitz continuous with modulus 1;

(ii) there exists a unique fixed point satisfying $V^* = \Psi(V^*) = \psi(\boldsymbol{\lambda}^*, V^*)$. Furthermore, $\boldsymbol{\lambda}^*$ solves (A.41) with the maximum discounted profit being V^*; and

(iii) The above algorithm for solving (A.42) converges to the optimal solution from any initial value $V^o_{1,\gamma}$.

Joint Pricing-Replenishment

In general, the pricing-replenishment joint optimal decisions lack the monotonicity properties with respect to the parameters σ and h even under the single price.

When N prices are optimized jointly with the replenishment level S, we can use the iterative algorithm developed above for each fixed S, and then do a line search to find the optimal S.

Convergence to the Average Objective

The discounted profit function converges to the average profit function in the following sense.

Proposition A.3.7 *Let $v_{n,\gamma}$ and V_γ be defined as (A.36) and (A.37), and let v_n and V follow (3.3) and (3.4) under the average objective. Then,*

$$\lim_{\gamma\to 0} v_{n,\gamma} = v_n \qquad \textit{and} \qquad \lim_{\gamma\to 0} \gamma V_\gamma = V.$$

The optimal pricing-replenishment decisions under the discounted objective converge to those under the average objective, notwithstanding the contrasts highlighted above.

Proposition A.3.8 *Let (S^*_γ, p^*_γ) and (S^*, p^*) denote, respectively, the optimal solution under the discounted and the average objectives. Then, we have*

$$\lim_{\gamma\to 0} S^*_\gamma = S^* \qquad \textit{and} \qquad \lim_{\gamma\to 0} p^*_\gamma = p^*.$$

A.3.2 Price-Sensitive Demand Variability

In Section 3.4.3, the Brownian model (3.1) is extended to a more general model (3.19), which allows a price-dependent diffusion coefficient. Using the same way as that leading to (3.4) and (3.5), we can derive the average objective function for the price-sensitive demand model, shown in (3.21) and (3.22). In this section, we state the results extending Lemma 3.2, Theorem 3.2 and 3.3. Proofs are relegated to Section A.3.3.

The relevant maximization problem is

$$\max_{\boldsymbol{\mu}\in\mathcal{M}^N} \; V(S,\boldsymbol{\mu}) \;=\; \frac{\sum_{n=1}^N \left[p(\frac{1}{\mu_n}) - \frac{hS}{N}(N-n+\frac{1}{2})\mu_n - \frac{1}{2}h\sigma(\mu_n^{-1})^2\mu_n^2 - a(S)\right]}{\sum_{n=1}^N \mu_n}. \tag{A.44}$$

Assumption A.1 (on demand variability) *$C_v(\lambda) = \sigma(\lambda)^2/\lambda$ is convex in λ.*

This assumption implies that $\sigma(\mu^{-1})^2\mu^2$ is convex in μ. We again note that the holding cost attributed to the deterministic part of the demand is the *only* motive for varying prices over the periods. In the numerator of (A.44), the additional holding cost term $-\sum\frac{1}{2}h\sigma(\mu_n^{-1})^2\mu_n^2$ and the revenue term $\sum p(\frac{1}{\mu_n})$ are, on the contrary, suggesting not varying prices, since they are both concave in $\boldsymbol{\mu}$. These two terms limit the extend to which the optimal prices vary, and thus limit the potential profit improvement.

Lemma A.3.1 (extension of Lemma 3.2) *For fixed S, let $\boldsymbol{\mu}^*$ be the optimal N prices. Then for $1 \le m < n \le N$,*

$$\mu_n^* - \mu_m^* \;\le\; \frac{S(n-m)}{N\left(G+\frac{B}{h}\right)},$$

where $B = \inf\left\{ -\frac{d^2p(\frac{1}{\mu})}{d\mu^2} \;:\; \mu\in[\mu_1^*,\mu_N^*]\right\}$ and $G = \inf\left\{ \frac{d^2\left(\frac{1}{2}\sigma(\frac{1}{\mu})^2\mu^2\right)}{d\mu^2} \;:\; \mu\in[\mu_1^*,\mu_N^*]\right\}$.

Proposition A.3.9 (extension of Theorem 3.2) *For fixed S, let $\boldsymbol{\mu}^*$ be the optimal N prices, and let V_N^* be the optimal average profit defined in (A.44). Then,*

$$V_N^* - V_1^* \;\le\; \frac{hS^2\left(1-N^{-2}\right)}{12\bar{\mu}\left(G+\frac{B}{h}\right)},$$

where $\bar{\mu} = \frac{1}{N}\sum_{n=1}^N\mu_n^*$, $B = \inf\left\{ -\frac{d^2p(\frac{1}{\mu})}{d\mu^2} \;:\; \mu\in[\mu_1^*,\mu_N^*]\right\}$ and $G = \inf\left\{ \frac{d^2\left(\frac{1}{2}\sigma(\frac{1}{\mu})^2\mu^2\right)}{d\mu^2} \;:\; \mu\in [\mu_1^*,\mu_N^*]\right\}$.

Proposition A.3.10 (extension of Theorem 3.3) *Let $(S_N^*,\boldsymbol{\mu}^*)$ be the optimal joint pricing-replenishment decision, and V_N^* be the corresponding optimal profit. Then,*

$$V_N^* - V_1^* \;\le\; \frac{h{S_N^*}^2\left(1-N^{-2}\right)}{12\bar{\mu}\left(G+\frac{B}{h}\right)},$$

where $\bar{\mu}$, B, and G are the same as in Proposition A.3.9.

Proof. Same as the proof for Theorem 3.3 with σ^2 replaced by G. ∎

A.3.3 Proofs

Proof of Proposition A.3.1. We can derive

$$\frac{\partial^2 V_\gamma}{\partial S^2} = -\frac{(bSe^{-bS}+2e^{-bS}+bS-2)e^{-bS}b(\frac{h}{\gamma}+a(S))}{(1-e^{-bS})^3}+\frac{2bSe^{-bS}a'(S)}{(1-e^{-bS})^2}-\frac{c''(S)}{1-e^{-bS}}.$$

The first term on the right side is non-positive due to the fact that

$$h(x) := xe^{-x}+2e^{-x}+x-2 \geq 0 \qquad \text{for all } x \geq 0, \tag{A.45}$$

since $h(0) = h'(0) = 0$ and $h''(x) = xe^{-x} \geq 0$ for all $x \geq 0$; the second term is negative since $a'(S) < 0$; the third term is also non-positive if $c(S)$ is convex in S. Hence $V_\gamma(S,\lambda)$ is strictly concave in S if $a'(S) < 0$ and $c''(S) \geq 0$. ∎

Proof of Proposition A.3.2.

(i). To show that S^*_γ is decreasing in p, it suffices to verify that $V_\gamma(S,\lambda)$ is supermodular in (S,λ); i.e., $\frac{\partial^2 V}{\partial S \partial \lambda} \geq 0$. Since the optimal replenishment level must satisfy $a'(S) < 0$, it suffices to verify the supermodularity in the sub-lattice where $a'(S) < 0$.

We have

$$\frac{\partial^2 V_\gamma(S,\lambda)}{\partial S \partial \lambda} = g^2 e^{-bS} b' S\Big[\frac{h}{\gamma}+c'(S)-\Big(\frac{h}{\gamma}+a(S)\Big)\Big(bS\frac{1+e^{-bS}}{1-e^{-bS}}-1\Big)\Big].$$

On the right side of the above, we have $b'(\lambda) < 0$ and $\frac{h}{\gamma}+c'(S) < \frac{h}{\gamma}+a(S)$. (The latter is implied by $a'(S) < 0$).

Thus, $\frac{\partial^2 V_\gamma}{\partial S \partial \lambda} \geq 0$ can be established if we have

$$bS\frac{1+e^{-bS}}{1-e^{-bS}}-1 \;\geq\; 1.$$

But this last inequality follows immediately from (A.45).

The proof of S^* increasing in σ is completely analogous: we can follow the above to establish $\frac{\partial^2 V_\gamma}{\partial S \partial \sigma} \geq 0$ in the sub-lattice where $a'(S) < 0$; the only change is to replace $b'(\lambda)$ by $b'(\sigma)$. Note in particular that $b'(\sigma) < 0$ too.

To show that S^* is decreasing in h, note that V_γ is submodular in (S,h):

$$\frac{\partial^2 V_\gamma}{\partial S \partial h} = -\frac{1-e^{-bS}-bSe^{-bS}}{\gamma(1-e^{-bS})^2} \leq 0,$$

since $1-e^{-bS} \geq bSe^{-bS}$.

(ii). This part is obvious. ∎

Proof of Proposition A.3.3. It suffices to show that $g(\lambda) := \frac{1}{1-e^{-b(\lambda)S}}$ is convex in λ. To this end, we derive

$$\begin{aligned} g''(\lambda) &= g^2 e^{-bS} S\Big(b'^2 S(2ge^{-bS}+1) - b''\Big) \\ &= \frac{g^2 e^{-bS} S}{\lambda^2+2\gamma\sigma^2}\Big(b^2 S(2ge^{-bS}+1) - \frac{2\gamma}{\sqrt{\lambda^2+2\gamma\sigma^2}}\Big), \end{aligned}$$

where the last equality follows from $b' = -\frac{b}{\sqrt{\lambda^2+2\gamma\sigma^2}}$ and $b'' = \frac{2\gamma}{(\lambda^2+2\gamma\sigma^2)^{3/2}}$.

We have

$$\begin{aligned} b^2 S(2ge^{-bS}+1) - \frac{2\gamma}{\sqrt{\lambda^2+2\gamma\sigma^2}} &> b^2 S(2ge^{-bS}+1) - 2b \\ &= bg(bSe^{-bS} + bS + 2e^{-bS} - 2) \\ &\geq 0, \end{aligned}$$

where the last inequality is again due to (A.45). Hence, $g(\lambda)$ is convex in λ, and $V_\gamma(S,\lambda)$ is concave in λ. ∎

Proof of Proposition A.3.4.

(i) To prove the monotonicity in h, we show that V_γ is supermodular in (λ, h):

$$\begin{aligned} \frac{\partial^2 V}{\partial\lambda\partial h} &= \frac{S^2 e^{-bS} b'}{\gamma(1-e^{-bS})^2} + \frac{1}{\gamma^2} \\ &= \frac{1}{b\gamma^2}\left[-\frac{e^{-bS}S^2b^2}{(1-e^{-bS})^2}\frac{\gamma}{\sqrt{\lambda^2+2\sigma^2\gamma}} + b\right] \\ &\geq \frac{1}{b\gamma^2}\left[-\frac{\gamma}{\sqrt{\lambda^2+2\sigma^2\gamma}} + b\right] \\ &\geq 0, \end{aligned}$$

where the first inequality results from the following:

$$\frac{x^2 e^{-x}}{(1-e^{-x})^2} = \frac{x^2}{e^x + e^{-x} - 2} = \frac{x^2}{\sum_{n=1}^{\infty} 2x^{2n}/(2n)!} \leq 1,$$

and in the last inequality we used the fact that

$$\frac{\gamma}{\sqrt{\lambda^2+2\sigma^2\gamma}} \leq b = \frac{2\gamma}{\sqrt{\lambda^2+2\sigma^2\gamma}+\lambda} \leq \frac{\gamma}{\lambda}.$$

(ii) The monotonicity in S relies on the supermodularity of V_γ in (S,λ), which has been established in the proof of Proposition A.3.2 (i).

(iii). This part is obvious. ∎

Proof of Proposition A.3.5. Let the optimal profit-to-go be $V^*_{n,\gamma}$, for $n = 1, \dots, N$, which satisfy (A.42). For the ease of exposition, define $V^*_{N+1,\gamma} := V^*_{1,\gamma}$. In view of (A.42) and (A.36), we can write

$$\lambda^*_n \;\in\; \arg\max_{\lambda\in\mathcal{L}} \{F(n,\lambda)\}, \qquad \text{for } n = 1, \dots, N,$$

where

$$F(n,\lambda) \;=\; f(\lambda) + \frac{hS(1-\Theta(\lambda))}{N\gamma} n + \Theta(\lambda) V^*_{n+1,\gamma}.$$

Thus, λ^*_n is decreasing in n if we can show that $F(n,\lambda)$ is submodular in (n,λ). To this end, notice that

$$\begin{aligned} F(n,\lambda) - F(n-1,\lambda) \;&=\; \frac{hS(1-\Theta(\lambda))}{N\gamma} + \Theta(\lambda)(V^*_{n+1,\gamma} - V^*_{n,\gamma}) \\ &=\; \frac{hS}{N\gamma} + \Theta(\lambda)\Big(V^*_{n+1,\gamma} - V^*_{n,\gamma} - \frac{hS}{N\gamma}\Big), \qquad \text{for } n = 2, \dots, N. \end{aligned}$$

Since $\Theta(\lambda)$ is decreasing in λ, the right side of the last equation is decreasing in λ if $V^*_{n+1,\gamma} - V^*_{n,\gamma} \geq \frac{hS}{N\gamma}$. This is indeed the case. Consider starting from the inventory level $\frac{S(N-n+1)}{N}$ but following the optimal policy for the inventory starting at $\frac{S(N-n)}{N}$. This implies that $\frac{S}{N}$ units of inventory will be held for ever. Consequently, the profit associated with this strategy is $V^*_{n+1,\gamma} - \frac{hS}{N\gamma}$, where $\frac{hS}{N\gamma}$ is the discounted cost of holding the extra $\frac{S}{N}$ units. Since this strategy is sub-optimal, we have

$$V^*_{n+1,\gamma} - V^*_{n,\gamma} \;\geq\; V^*_{n+1,\gamma} - \Big(V^*_{n+1,\gamma} - \frac{hS}{N\gamma}\Big) \;=\; \frac{hS}{N\gamma}.$$

The submodularity of $F(n,p)$ follows, and hence λ^*_n is decreasing in n, or equivalently, p^*_n is increasing in n. ∎

Proof of Proposition A.3.6.

(i). For $V > \widetilde{V}$, let $\boldsymbol{\lambda}$ and $\widetilde{\boldsymbol{\lambda}}$ be the maximizers (not necessarily in the interior) of $\psi(\cdot, V)$ and $\psi(\widetilde{\boldsymbol{\lambda}}, \widetilde{V})$, respectively (refer to (A.43)). Then,

$$\Psi(V) - \Psi(\widetilde{V}) \;=\; \psi(\boldsymbol{\lambda}, V) - \psi(\widetilde{\boldsymbol{\lambda}}, \widetilde{V}) \;\geq\; \psi(\widetilde{\boldsymbol{\lambda}}, V) - \psi(\widetilde{\boldsymbol{\lambda}}, \widetilde{V}) \;=\; \Theta_\gamma(\widetilde{\boldsymbol{\lambda}})(V - \widetilde{V}) \;\geq\; 0,$$

which implies that $\Psi(V)$ is increasing in V. Furthermore,

$$\Psi(V) - \Psi(\widetilde{V}) \;=\; \psi(\boldsymbol{\lambda}, V) - \psi(\widetilde{\boldsymbol{\lambda}}, \widetilde{V}) \;\leq\; \psi(\boldsymbol{\lambda}, V) - \psi(\boldsymbol{\lambda}, \widetilde{V}) \;=\; \Theta_\gamma(\boldsymbol{\lambda})(V - \widetilde{V}). \qquad \text{(A.46)}$$

Since $\Theta_\gamma(\boldsymbol{\lambda}) \leq 1$, the above inequality implies that Ψ is Lipschitz continuous, with modulus no larger than 1. That is, $\Psi(V)$ is non-expansive.

(ii). Let $\boldsymbol{\lambda}^*$ be the optimal solution to (A.41) with optimal value V^*. We eliminate the uninteresting case where $\boldsymbol{\lambda}^* = \infty$. Thus, $\Theta_\gamma(\boldsymbol{\lambda}^*) < 1$ and $V^* = \frac{v_\gamma(\boldsymbol{\lambda}^*)}{1-\Theta_\gamma(\boldsymbol{\lambda}^*)}$, or equivalently, $V^* = \psi(\boldsymbol{\lambda}^*, V^*)$. On the other hand, $V^* \geq \frac{v_\gamma(\boldsymbol{\lambda})}{1-\Theta_\gamma(\boldsymbol{\lambda})}$, or equivalently, $V^* \geq \psi(\boldsymbol{\lambda}, V^*)$ for all $\boldsymbol{\lambda} \in \mathcal{L}^N$. Thus, $V^* = \max_{\boldsymbol{\lambda}\in\mathcal{L}^N} \psi(\boldsymbol{\lambda}, V^*) = \Psi(V^*)$; and hence, V^* is a fixed point of Ψ. This proves the existence.

Conversely, suppose V^* is a fixed point of $\Psi(V)$ with the corresponding maximizer $\boldsymbol{\lambda}^*$. Note that $\boldsymbol{\lambda}^* = \infty$ implies that all inventory can be sold out instantaneously after the replenishment (with the revenue covering exactly the replenishment cost). If we eliminate this uninteresting case, then $\Theta_\gamma(\boldsymbol{\lambda}^*) < 1$, and $V^* = \psi(\boldsymbol{\lambda}^*, V^*)$ or $V^* = \frac{v_\gamma(\boldsymbol{\lambda}^*)}{1-\Theta_\gamma(\boldsymbol{\lambda}^*)}$. At the same time, $V^* \geq \psi(\boldsymbol{\lambda}, V^*)$ or $V^* \geq \frac{v_\gamma(\boldsymbol{\lambda})}{1-\Theta_\gamma(\boldsymbol{\lambda})}$ for all $\boldsymbol{\lambda} \in \mathcal{L}^N$. Hence, $\boldsymbol{\lambda}^*$ solves the problem in (A.41), achieving the optimal value V^*.

Now, suppose there exist two fixed points $V^* > \widetilde{V}^*$ corresponding to the maximizers $\boldsymbol{\lambda}^*$ and $\widetilde{\boldsymbol{\lambda}}^*$, respectively. Then, both V^* and $\widetilde{V}^*$ are optimal values for the problem in (A.41), an obvious contradiction (against optimality). This establishes the uniqueness.

(iii). Let V^* be the unique fixed point. From (i), we have $\Psi(V) > V$ for $V < V^*$ and $\Psi(V) < V$ for $V > V^*$. Hence, if we start from an initial value $V_0 < V^*$, the sequence of values generated by the iteration $V_n = \Psi(V_{n-1})$ is increasing and bounded from above by V^*. So the sequence converges to a fixed point of the mapping Ψ, which must be V^*. A similar argument holds for an initial value $V_0 > V^*$. ∎

Proof of Proposition A.3.7. When $\gamma \to 0$, via Taylor expansion, we can derive the following:

$$b(\lambda) = \frac{\gamma}{\lambda} - \frac{\sigma^2\gamma^2}{2\lambda^3} + O(\gamma^3), \tag{A.47}$$

$$\Theta(\lambda) = e^{-b(\lambda)S/N} = 1 - \frac{S\gamma}{N\lambda} + \frac{\sigma^2 S\gamma^2}{2N\lambda^3} + \frac{S^2\gamma^2}{2N^2\lambda^2} + O(\gamma^3). \tag{A.48}$$

Applying (A.47) and (A.48), we have

$$\begin{aligned} v_{n,\gamma}(\lambda) &= \left(p(\lambda) + \frac{h}{\gamma}\right)\left(\frac{S}{N} - \frac{\sigma^2 S\gamma}{2N\lambda^2} - \frac{S^2\gamma}{2N^2\lambda}\right) + \frac{hS}{\gamma} - \frac{hS^2}{N\lambda} - \frac{hS(N+1)}{N\gamma} + \frac{hS^2 n}{N^2\lambda} + o(1) \\ &= \frac{p(\lambda)S}{N} - \frac{hS^2(N-n+\frac{1}{2})}{N^2\lambda} - \frac{h\sigma^2 S}{2N\lambda^2} + o(1) \\ &= v_n(\lambda) + o(1), \qquad \text{as } \gamma \to 0. \end{aligned}$$

Then,

$$\begin{aligned}\gamma V_\gamma(S,\boldsymbol{\lambda}) &= \frac{v_\gamma(\boldsymbol{\lambda})}{\gamma^{-1}(1-\Theta_\gamma(\boldsymbol{\lambda}))} = \frac{\sum_n v_n(\lambda_n) - c(S) + o(1)}{\gamma^{-1}(1-\prod_{n=1}^N(1-\frac{S\gamma}{N\lambda_n}))} \\ &= \frac{\sum_n v_n(\lambda_n) - c(S) + o(1)}{\frac{S}{N}\sum_n \lambda_n^{-1} + o(1)} = V(S,\boldsymbol{\mu}) + o(1), \qquad \text{as } \gamma \to 0.\end{aligned}$$

∎

Proof of Proposition A.3.8. The first-order condition for S_γ^* in (A.39) can be written as follows:

$$(h+\gamma c'(S))(\frac{S\gamma}{\lambda} - \frac{\sigma^2 S\gamma^2}{2\lambda^3} - \frac{S^2\gamma^2}{2\lambda^2}) = (1-\frac{S\gamma}{\lambda})(\frac{\gamma}{\lambda} - \frac{\sigma^2\gamma^2}{2\lambda^3})[hS + \gamma c(S)] + o(\gamma^2)$$

Ignoring higher-order terms, we can simplify the above to the equation in (3.6), which is the first-order optimality condition for S^*.

Using (A.47) and (A.48), when $\gamma \to 0$, the first-order condition in (A.40) can be approximated as

$$\begin{aligned} 0 &= r'(\lambda) + \frac{(h+\gamma a(S))S}{\frac{S^2\gamma^2}{\lambda^2}(1-\frac{\sigma^2\gamma}{2\lambda^2}-\frac{S\gamma}{2\lambda})^2} S\Big(1-\frac{S\gamma}{\lambda}\Big)\Big(-\frac{\gamma}{\lambda^2}+\frac{3\sigma^2\gamma^2}{2\lambda^4}\Big) + \frac{h}{\gamma} + o(1) \\ &= r'(\lambda) - \Big(\frac{h}{\gamma}+a(S)\Big)\Big(1+\frac{\sigma^2\gamma}{\lambda^2}+\frac{S\gamma}{\lambda}\Big)\Big(1-\frac{S\gamma}{\lambda}\Big)\Big(1-\frac{3\sigma^2\gamma}{2\lambda^2}\Big) + \frac{h}{\gamma} + o(1) \\ &= r'(\lambda) - \frac{h}{\gamma} - a(S) - \frac{h}{\gamma}\Big(\frac{\sigma^2\gamma}{\lambda^2}+\frac{S\gamma}{\lambda}-\frac{S\gamma}{\lambda}-\frac{3\sigma^2\gamma}{2\lambda^2}\Big) + \frac{h}{\gamma} + o(1) \\ &= r'(\lambda) - a(S) + \frac{h\sigma^2}{2\lambda^2} + o(1), \end{aligned}$$

which is exactly the same as the first-order condition in (3.9) for the average objective.

The convergence of the first-order conditions (A.39) and (A.40) to their counterparts in (3.6) and (3.9), respectively, implies that the decisions under the discounted objective converge to those under the average objective. ∎

Proof of Lemma A.3.1. Optimal prices are increasing over the periods: $\mu_n^* \geq \mu_m^*$ for $n > m$. If $\mu_m^* = \mu_n^*$, then the desired inequality obviously holds. If $\mu_m^* < \mu_n^*$, consider the following alternative pricing policy: $(\mu_1^*, \ldots, \mu_m^* + \delta, \ldots, \mu_n^* - \delta, \ldots, \mu_N^*)$, where $0 < \delta < \frac{\mu_n^* - \mu_m^*}{2}$. Since this alternative cannot be better than the optimum, we have

$$\begin{aligned}& p(\tfrac{1}{\mu_m^*+\delta}) + p(\tfrac{1}{\mu_n^*-\delta}) - \tfrac{hS}{N}\left[(N-m+\tfrac{1}{2})(\mu_m^*+\delta) + (N-n+\tfrac{1}{2})(\mu_n^*-\delta)\right] \\ & -h\left[f(\mu_m^*+\delta) + f(\mu_n^*-\delta)\right]\end{aligned}$$

$$\leq \quad p(\tfrac{1}{\mu_m^*}) + p(\tfrac{1}{\mu_n^*}) - \tfrac{hS}{N}\left[(N-m+\tfrac{1}{2})\mu_m^* + (N-n+\tfrac{1}{2})\mu_n^*\right] - h\left[f(\mu_m^*) + f(\mu_n^*)\right],$$

where $f(\mu) := \frac{1}{2}\sigma(\frac{1}{\mu})^2\mu^2$. Taylor's expansion with straightforward algebra simplifies the inequality to

$$\tfrac{dp}{d\mu}(\tfrac{1}{\mu_m^*+\zeta_1\delta})\delta - \tfrac{dp}{d\mu}(\tfrac{1}{\mu_n^*-\zeta_2\delta})\delta + \tfrac{hS}{N}\delta(m-n) + h\delta\Big(\tfrac{df}{d\mu}(\mu_n^* - \zeta_3\delta) - \tfrac{df}{d\mu}(\mu_m^* + \zeta_4\delta)\Big) \quad \leq \quad 0,$$

where $\zeta_1, \zeta_2, \zeta_3, \zeta_4 \in [0,1]$. Dividing both sides by δ, and then letting $\delta \to 0$, we have

$$\tfrac{dp}{d\mu}(\tfrac{1}{\mu_m^*}) - \tfrac{dp}{d\mu}(\tfrac{1}{\mu_n^*}) + \tfrac{hS}{N}(m-n) + h\Big(\tfrac{df}{d\mu}(\mu_n^*) - \tfrac{df}{d\mu}(\mu_m^*)\Big) \quad \leq \quad 0. \tag{A.49}$$

Applying the mean-value theorem and using the definition of B and G, we have

$$\tfrac{dp}{d\mu}(\tfrac{1}{\mu_m^*}) - \tfrac{dp}{d\mu}(\tfrac{1}{\mu_n^*}) \quad = \quad -\tfrac{d^2p}{d\mu^2}(\tfrac{1}{\zeta_5})(\mu_n^* - \mu_m^*) \ \geq \ B(\mu_n^* - \mu_m^*), \tag{A.50}$$

$$\tfrac{df}{d\mu}(\mu_n^*) - \tfrac{df}{d\mu}(\mu_m^*) \quad = \quad \tfrac{d^2f}{d\mu^2}(\zeta_6)(\mu_n^* - \mu_m^*) \ \geq \ G(\mu_n^* - \mu_m^*) \tag{A.51}$$

where $\zeta_5, \zeta_6 \in [\mu_m^*, \mu_n^*]$. Combining (A.49)-(A.51) yields the desired inequality. ∎

Proof of Proposition A.3.9. The average profit of charging a single price $p(\bar{\mu}^{-1})$ is

$$V_1 \quad = \quad \frac{p(\frac{1}{\bar{\mu}}) - \frac{hS}{2}\bar{\mu} - \frac{h}{2}\sigma(\bar{\mu}^{-1})^2\bar{\mu}^2 - a(S)}{\bar{\mu}}.$$

We have

$$\begin{aligned}
&V_N^* - V_1^* \\
&\leq V_N^* - V_1 \\
&= \frac{\sum_{n=1}^N \left[p(\frac{1}{\mu_n^*}) - p(\frac{1}{\bar{\mu}})\right] - \frac{hS}{N}\sum_{n=1}^N \left[(N-n+\frac{1}{2})\mu_n^* - \frac{N}{2}\bar{\mu})\right] - \frac{hN}{2}\left[\frac{1}{N}\sum_{n=1}^N \sigma(\frac{1}{\mu_n^*})^2\mu_n^{*2} - \sigma(\frac{1}{\bar{\mu}})^2\bar{\mu}^2\right]}{\sum_{n=1}^N \mu_n^*}.
\end{aligned}$$

The first term in the numerator, $\sum_{n=1}^N \left[p(\frac{1}{\mu_n^*}) - p(\frac{1}{\bar{\mu}})\right] \leq 0$, since $p(\frac{1}{\mu})$ is concave in μ; and the last term in the numerator $\frac{1}{N}\sum_{n=1}^N \sigma(\frac{1}{\mu_n^*})^2\mu_n^{*2} - \sigma(\frac{1}{\bar{\mu}})^2\bar{\mu}^2 \geq 0$, since $\sigma(\frac{1}{\mu})^2\mu^2$ is convex in μ. Thus,

$$\begin{aligned}
V_N^* - V_1^* \quad &\leq \quad \frac{-\frac{hS}{N}\sum_{n=1}^N \left[(N-n+\frac{1}{2})\mu_n^* - \frac{N}{2}\bar{\mu})\right]}{\sum_{n=1}^N \mu_n^*} \\
&= \quad \frac{hS}{N}\left(\frac{\sum_{n=1}^N n\mu_n^*}{\sum_{n=1}^N \mu_n^*} - \frac{N+1}{2}\right) \\
&= \quad \frac{hS}{N}\left(\frac{2\mu_1^* + 4\mu_2^* + \cdots + 2N\mu_N^* - (N+1)(\mu_1^* + \cdots + \mu_N^*)}{2\sum_{n=1}^N \mu_n^*}\right) \\
&= \quad \frac{hS}{N}\left(\frac{(N-1)(\mu_N^* - \mu_1^*) + (N-3)(\mu_{N-1}^* - \mu_2^*) + \ldots}{2N\bar{\mu}}\right)
\end{aligned}$$

where in the last line, the series ends with $\mu^*_{N/2+1} - \mu^*_{N/2}$ if N is even, and ends with $2(\mu^*_{(N+3)/2} - \mu^*_{(N-1)/2})$ if N is odd.

Applying Lemma A.3.1 and the identity:

$$(N-1)^2 + (N-3)^2 + \cdots + (N+1-2\left\lfloor\frac{N}{2}\right\rfloor)^2 = \frac{(N-1)N(N+1)}{6},$$

we have

$$\begin{aligned}
V_N^* - V_1^* &\leq \frac{hS^2}{\bar{\mu}\left(G+\frac{B}{h}\right)}\left(\frac{(N-1)^2 + (N-3)^2 + \cdots + (N+1-2\lfloor\frac{N}{2}\rfloor)^2}{2N^3}\right) \\
&= \frac{hS^2}{\bar{\mu}\left(G+\frac{B}{h}\right)}\frac{(N-1)N(N+1)}{12N^3} \\
&= \frac{hS^2\left(1-N^{-2}\right)}{12\bar{\mu}\left(G+\frac{B}{h}\right)}.
\end{aligned}$$

∎