THREE ESSAYS ON DYNAMIC PRODUCTION AND PRICING DECISIONS FOR NEW PRODUCTS

by

Wenjing Shen

A dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy (Business Administration) in The University of Michigan 2008

Doctoral Committee:

Professor Izak Duenyas, Co-Chair Associate Professor Roman Kapuscinski, Co-Chair Professor Mayuram S. Krishnan Assistant Professor Volodymyr Babich

Copyright © $\frac{\text{Wenjing Shen}}{\text{All Rights Reserved}}$ 2008

CONTENTS

LIST	OF FI	GURES	iv						
LIST	OF AF	PPENDICES	V						
СНАР	TER								
1	Over	view	1						
2	New	Product Diffusion Decisions Under Supply Constraints	<u> </u>						
	2.1	Introduction	Į.						
	2.2	Literature Review	Ĝ						
	2.3	Model	13						
	2.4	Infinite Capacity	17						
	2.5	Limited Capacity							
		2.5.1 Optimal demand and price trajectories under lost sales .	23						
		2.5.2 Optimal demand, and price trajectories under partial							
		backlogging	30						
	2.6	Numerical Study	37						
	2.7	Conclusions	41						
3	Optio	mal Production and Ordering Policies with Random Yield							
	Prod	uction and Queueing	43						

3.1 Introduction
3.2 Literature Review
3.3 Basic Model
3.3.1 Optimal Lot Size Decisions
3.3.2 Overtime or Subcontracting Decisions
3.4 Summary
4 Make-or-Buy Decisions When Cost Reduction Requires Mini-
mum Production Quantity 66
4.1 Introduction
4.2 Literature Review
4.3 Model
4.3.1 Centralized Problem
4.3.2 Decentralized Game
4.3.3 Effect of Parameter Changes 81
4.3.4 A Special Example
4.4 Three-period Deterministic Problem
4.5 Conclusion
5 Summary
APPENDICES
BIBLIOGRAPHY

LIST OF FIGURES

Figure

2.1	Optimal demand and price with unlimited capacity	20
2.2	Optimal policies with complete lost sale in make-to-order environment	24
2.3	Optimal policies with complete lost sale in make-to-stock environment	30
2.4	Optimal demand, inventory and backorder trajectories with make to	
	stock and partial backordering	36
2.5	Optimal price trajectories with make to stock and partial backordering	36
2.6	Value of pricing flexibility as a function of capacity K and imitation	
	factor q	38
2.7	Value of holding inventory as a function of capacity K and imitation	
	factor q	38
2.8	Value of allowing backorder	39
2.9	Optimality gap of skimming pricing strategies	39
2.10	Optimality gap of penetration pricing strategies	41
3.1	Cost difference of heuristic policies and optimal policy with different	
	material costs	58

3.2	Cost difference of heuristic policies and optimal policy with different	
	service rates	58
3.3	Cost difference of heuristic policies and optimal policy with different	
	yield distributions	59
4.1	Supplier's profits under three options	80
4.2	Manufacturer and Supplier's profits change with costs	83

LIST OF APPENDICES

Appendix

A	Main	Proofs in	Chapter 2	 	 	 	 	 	 	. !	98
В	Main	Proofs in	Chapter 3	 	 	 	 	 	 	. 13	20
\mathbf{C}	Main	Proofs in	Chapter 4							1:	36

CHAPTER 1

Overview

This dissertation focuses on dynamic production and pricing issues related to new products. Three essays in this dissertation analyze different aspects of the operations and marketing decisions. The first essay is focused on the dynamic pricing issue for new product diffusion process when capacity is limited, the second essay is focused on the production lot sizing decisions for a pharmaceutical firm that manufactures new medicines for clinical trial, and the third essay deals with outsourcing decisions when firms manufacture a new product and production costs can be reduced through "learning-by-doing" effect.

In Essay 1, "New Product Diffusion Decisions Under Supply Constraints," I study a capacity constrained firm prices products during new product introductions. Management of new product introductions is critical for nearly all firms, and one of its most important dimensions is the management of demand during the introduction. Research analyzing this area uses predominantly various versions of diffusion model (Bass 1969) to capture the demand trajectory of a new product with a fixed potential market. The classical Bass model assumes that price is fixed and capacity is unlimited. In reality, both factors critically influence firms' strategies. Price fluctuations for a new product are common and price is often a critical lever that helps

to shape the demand. Also, often firms have rather significant capacity constraints, especially during new production introduction times, which influence feasibility of their strategies.

I assume that the demand rate is influenced by price. When capacity is insufficient, I allow some customers to either be lost or to be backlogged and, therefore, for the imitation effect to be delayed. In order to understand the effect of both pricing and capacity, using control-theory framework (a generalization of the classic Bass model), I consider the integrated optimal pricing, production, and sales decisions. Most of the results are fairly robust and apply under the assumption of complete lost sales and partial backlogging, as well as make to order and make to stock environments. I show that in most of the cases, the optimal trajectory of demand is unimodal, as in the Bass model, but the optimal price trajectory can have multiple local maxima when capacity is limited. Lack of pricing flexibility might lead to intentional creation of backorders or lost sales, a phenomenon not observed when price adjustments are allowed. I also explore when pricing flexibility is most valuable using a numerical study. I find that benefits are highest when capacity is not unlimited nor very little, and when imitation effect dominates innovation effect. I also find that the capability to adjust prices is significantly more effective than the option of producing in advance and holding inventory.

In Essay 2, "Optimal Production and Ordering Policies with Random Yield Production and Queueing," I study lot sizing decisions with random yield, rigid demand and significant delay costs. The essay is motivated by lot sizing for active ingredi-

ent production for experimental drugs in the pharmaceutical industry. Clinical trials require a fairly rigid amount of products, whereas production yields for new drugs tend to be low and highly variable. Finally, delaying clinical trial is very costly, while the materials for making the active ingredients are also very expensive. Thus firms have a major tradeoff: releasing too much material would ensure on-time completion but waste materials, while releasing too little would delay this drug and potentially future ones because of compounding of delays.

I model this problem as a lot sizing decision with random yield and rigid demand in an M/G/1 queue environment. I derive structural results on the optimal policy and show that under reasonable assumptions about the yield distributions, the optimal production quantities are increasing in the number of waiting orders and the remaining quantity to be produced for the current lot. I also show that the ratio of the lot size to required amount is decreasing in required quantity.

I conduct a numerical study to compare the optimal average cost with several heuristic policies commonly used in the industry. The results indicate that in most cases the simple heuristic policies used in practice are very inefficient and substantial savings can be obtained by appropriate lot sizing policies.

In an extension, I consider a more complicated model where the firm also has the choice to outsource certain lots or use overtime and show that optimal policies follow a threshold structure.

In the third essay, "Make-or-Buy Decisions When Cost Reduction Requires Minimum Production Quantity," I study a problem of outsourcing some or all of the production (Make or Buy) in a dynamic setting, where both Manufacturer and his Supplier may improve their operations and decrease production costs. Such cost reductions, however, require costly effort and are only possible when experimenting with actual production methods. That is some minimum quantity of production enables potential improvements. Make or Buy is closely related to the question of dual sourcing, which has been studied before. While various reasons have been provided for dual sourcing, I provide a new explanation, which is driven by the fact that in-house production may facilitate learning about potential process improvements leading to eventual cost reductions.

In a two-period setting, I show that the choice between Manufacturer fully outsourcing to the Supplier, fully producing in house, or partial production and partial outsourcing depends predictably on the cost difference between Manufacturer and Supplier, but other factors matter as well. I find that dual sourcing is more likely when the learning process is more uncertain, when Manufacturer's minimum quantity required for learning and improvement is larger, or when the two plants incentives are aligned, e.g., they are part of the same company.

An extension to three periods suggests that it may be optimal to switch back and forth from full to partial outsourcing. Although Manufacturer has no incentive to subsidize Supplier's learning in two periods, in longer horizons the dynamics is more subtle. On one hand, such incentive may appear, on the other hand, the options to subsidize may hurt Manufacturer.

CHAPTER 2

New Product Diffusion Decisions Under Supply Constraints

2.1 Introduction

We are interested in the joint pricing, sales, and production decisions made by a manufacturer introducing a new product to the market and facing capacity constraints. Pricing of new products is a critical decision that firms make, and often firms face capacity constraints that affect their capability to satisfy demand during new product introductions. We are thus interested in the following groups of questions:

1. How does a firm manage pricing, production, and inventory decisions over time during new product introductions? What is the structure of optimal pricing, production and inventory policies? 2. What is the effect of the firm's capability to adjust prices? How effective is pricing as a tool to shape demand? How much does a firm lose if it has to stick to constant prices during a product's lifetime? 3. What effect does the firm's capability to make to stock prior to product introduction have on its profitability? Is pricing or inventory a more effective tool to maximize profit and satisfy demand during product introductions?

This setting is directly related to a fundamental stream of research. A classical model that characterizes demand diffusion dynamics for an innovative product is

the Bass model (Bass 1969). It assumes that (i) products can be produced in any quantities and are always available for purchase, and (ii) price is fixed or has no impact on the diffusion process. In reality, however, these assumptions are often violated.

While capacity shortages can be observed in many settings, they are quite likely to occur when new products are introduced to market and the appropriate infrastructure does not yet exist. Examples of supply shortages for new products are known in abundance. In December 2005, the Apple Store noted that the 1GB iPod shuffle is "Currently Unavailable. Sold out for holiday. Expected availability mid-January." Similarly, in 2006 Sony announced a shortage of PlayStation 3's due to insufficient manufacturing capacity of its key component, blue-laser diode. It is worth noting that capacity constraints, which are not captured in the classical Bass model, have a more significant impact on innovative products than they do on mature products. For an innovative product, not only does capacity shortage result in backlogged demand or lost sales, it also slows down the product diffusion process, because customers who have not received the product are not able to generate the positive word-of-mouth effect. Firms clearly need to consider manufacturing capacity and resulting product availability as critical elements of new product roll-outs.

The other critical factor we focus on is pricing. While prices of all products are adjusted over time, dynamic changes in prices have potentially big consequences for new products. Management of such price adjustments can be critical, as recently

¹The opposite effect not studied in this paper is the increase of desirability of product driven by perceived unavailability of products, see Stock and Balachander (2005).

observed in the introduction of the Apple iPhone. It is common, especially for manufacturers of electronic products, to adjust prices of products several times during their lifetime.

Given the practical importance of how firms' pricing, production, and inventory decisions change in the presence of capacity constraints, we are interested in both the characteristics of the optimal demand and price trajectories as well as in the choice of appropriate tools (levers) that can be used to manage this process. Clearly, to maximize profit when facing capacity shortage, a firm can use multiple tools, some of which have been analyzed in the existing literature in isolation. For example, producing in advance of product launch to build early supply has been considered by Ho et al. (2002), deliberately holding sales to slow down diffusion has been considered by Kumar and Swaminathan (2003), and changing prices to shape demand during new product introductions has been widely considered in the marketing literature. We are interested in how a manufacturer chooses any of these levers when all are available.

Despite the relevance of these questions, the effect of limited capacity and dynamic pricing, for new products remains relatively unclear in the research literature and the choice of tools sometimes even controversial. Qualitatively, capacity makes the pricing/production/selling decisions more complicated, as there is a potential benefit of influencing demand dynamically with pricing in order to match demand with available capacity. While there is a wide marketing literature on the role of pricing, it does not generally consider capacity constraints. The operations literature that deals with

product introductions does not generally consider pricing. Furthermore, predictions are sometimes contradictory. E.g., Kumar and Swaminathan (2003), considering a model where price is not a lever, proved that intentionally holding sales to slow diffusion (i.e., refusing to sell to customers even when the firm has sufficient inventory) can be optimal, while Ho et al. (2002) suggests that it is never optimal. We resolve these inconsistencies, when price cannot be adjusted.

Further, we show that analyzing production/inventory decisions without pricing as a lever can lead to results that fundamentally differ from when pricing can be used as a lever. We show that the strategy described in Kumar and Swaminathan, considering a model where price is not a lever, is no longer optimal when price can be dynamically adjusted. Instead, myopic policy, following the spirit of Ho et al., becomes optimal.

Based on the modified Bass demand dynamics that captures capacity constraint and pricing, we analyze the jointly optimal pricing, sales, and production decisions. We start with unlimited capacity, and show that both demand and price trajectories are unimodal over time. For limited capacity, we analyze a sequence of scenarios, starting from the simplest case of lost sales in make-to-order environment. Later, we examine the incremental effects of the possibility of holding inventory (make to stock) and partial backlogging. While the optimal trajectory of demand is unimodal for most cases (except when backlogging is possible and firm builds to stock), the optimal price trajectory may be more complicated and multiple local maxima may be observed. Our results can be interpreted in the framework of skimming and

penetration pricing strategies. Although multiple local optima may be optimal, our numerical experiments show that a penetration pricing strategy, widely considered in marketing literature under infinite-capacity assumptions, is often very close to optimal under finite capacity as well. Another interesting issue is how much a firm loses by not being able to change prices during a product's lifecycle or not being able to build inventory of the product or not being able to backorder the product. Our results indicate that pricing is a much more effective tool than inventory or backordering and losing that the capability to adjust prices affects profits much more severely than losing the capability to hold inventories or backorder.

The rest of the paper is organized as follows. In Section 2.2 we review the relevant literature. The Bass model is modified to incorporate capacity constraint and pricing in Section 2.3. In Section 2.4 we analyze pricing decisions when capacity is not limited, while Section 2.5 analyzes the optimal decisions for the limited capacity case. We present a numerical study to gain insights into the effectiveness of the pricing, inventory, and backordering levers in new product introduction, in Section 2.6. Finally, we provide conclusions in Section 2.7.

2.2 Literature Review

Among various demand diffusion models analyzed in marketing literature, the Bass (1969) model is one of the most recognized. The model describes adoption of new durable goods and assumes that there are two types of customers, innovative customers who make purchasing decisions independent of other customers' decisions, and imitative customers whose purchasing decisions are influenced by previous cus-

tomers through "word-of-mouth" effect. As a result, the instantaneous demand rate at time t is given by

$$d(t) = \left[p + \frac{q}{m}D(t)\right][m - D(t)],$$

where D(t) is the cumulative demand up to time t, m is the market size (the remaining potential market size is m - D(t)), p and q are labeled as "innovation coefficient" and "imitation coefficient" to represent the effects of direct communication and word of mouth, respectively.² Since the Bass diffusion model was introduced, it has been used to describe and forecast demand in various industries. In addition to the original consumer durable goods markets, these include, e.g., agriculture (Akinola 1986) and technology (Kalish and Lilien 1986).

The Bass model has been refined and extended in multiple directions, see Mahajan et al. (1990) for a thorough survey. Among those, one particularly relevant to our work is the influence of pricing on coefficients of innovation and imitation. Robinson and Lakhani (1975) is among the first papers to incorporate price into diffusion model. They assume that the diffusion rate is modified by $e^{-kp(t)}$, where p(t) is price at time t and k is a constant, and production cost decreases due to learning-curve effects. The paper examines a number of intuitive pricing strategies, including constant price, constant margin pricing, and constant-return-rate pricing. Through numerical examples, it is shown that all of these strategies can be far from optimality.

²Later literature, e.g., Mahajan et al. (1990), assumes the same functional form, but instead of labelling customers as innovators or imitators, assumes that individual customer's adoption, or probability of purchase, is influenced by two factors: external and internal. External factors might have the form of direct communications from mass media, such as advertisement. Internal factors refer to the internal communications between potential customers and customers who have already purchased.

This exponential form of price dependence is among the most common (if not the most common) form of incorporating price in diffusion models and we use it in our model. There exist, however, some variations in how price has been incorporated into diffusion models. Some of the papers assume the size of the potential market to be affected by price, m = m(p), or assume adoption rate to be affected by price, such as Robinson and Lakhani (1975), Dolan and Jeuland (1981), etc. Other alternative is to model customer individual behavior as maximizing individual utility functions (Kalish 1985). Krishnan et al. (1999) propose a Generalized Bass model, where the price fluctuation rate also affects the adoption rate.

The above studies, as well as several others, disagree on the structure of optimal pricing strategy. The pricing policies are generally classified into two categories: (i) skimming, where prices are decreased over time, and (ii) penetration, where low prices are offered early to attract a broad range of consumers (penetrate market) in order to create some diffusion-like effects. Usually, in the case of penetration, the price increases and may eventually decrease. Many of the earlier papers claim that the optimal price always follows the demand process (e.g. Kalish 1983). Later, Krishnan et al. (1999) propose a Generalized Bass Model and show that the pricing trajectory depends on the "discounted price coefficient", a product of the discount rate and the diffusion price parameter.

All of the above variants of Bass model assume sufficient availability of products, i.e., production capacity is never a constraint. We show that co-existence of pricing flexibility and capacity constraint influences the production, as well as pricing

decisions and that the resulting policies may have different structure than policies described in the existing literature.

In operations literature there have been several papers that consider the capacityconstrained diffusion problem. Jain, Mahajan, and Muller (1989) consider a diffusion model under capacity constraint. They find the demand pattern with limited capacity to be negatively skewed, and empirically verify such behavior using new telephone market data in Israel. Kumar and Swaminathan (2003) and Ho et al. (2002) independently propose models, which include capacity constraint in a diffusion model and analyze the optimal operational and marketing decisions. Although the two models differ only in minor details, some of their conclusions are contradictory. Both models adopt a modified Bass model and assume that only customers who purchased the product create the word of mouth effect as we do. Also, both models assume partial backlogging and focus on optimal sales decisions. Ho et al. shows that a myopic sales policy is always optimal, that is, always satisfy current demand if capacity and inventory allow for it, while Kumar and Swaminathan shows that not satisfying some of the demand, i.e., strategically holding sales by denying some customers even when they can be satisfied by current inventory/capacity, might be optimal in order to slow down the diffusion process. While the above operations-focused papers incorporate capacity constraint, they ignore possibility of using pricing to influence sales. Dynamic pricing is both used in practice and promising from a conceptual point of view, since unlike the strategy of holding sales, resulting in backlogging or lost-sale costs, or the strategy of delaying launching, that delays cash flow, raising price may

slow down diffusion process without incurring these costs. We re-examine the existing models and put it in a broader framework with price flexibility. Our model shows that price flexibility changes the qualitative insights and leads to myopic sales decisions in settings analyzed in these papers.

Thus, we focus on a diffusion model incorporating both realistic features, capacity constraints and pricing flexibility, and evaluate appropriate policies that should be used.

2.3 Model

Consider a diffusion process of a new product with limited capacity. We use a modification of the Bass diffusion model (Bass 1969), where price, sales, and production are decision variables that need to be determined during the whole product life cycle.

The Bass model assumes that a new durable product is introduced to a potential market of size m. Given cumulative demand D(t) at time t, demand rate d(t) is given by:

$$d(t) = \left[p + \frac{q}{m}D(t)\right][m - D(t)],$$

where p is an innovation coefficient and q an imitation coefficient.

Capacity shortage not only results in immediate unsatisfied demand, but also influences the future diffusion process of demand, because unsatisfied customers cannot spread product information without having experienced it. We define a sale as when a customer receives a product. S(t) denotes the cumulative sales at time t. As in Kumar and Swaminathan (2003) and Ho et al. (2002), only those customers who have received the product can be the source of the word-of-mouth effect. (In the case when backlogs are allowed, we assume that backlogged customers start generating word of mouth only when they have received the product.)

We make three assumptions about customer behavior:

- 1. We do not model detailed customer strategic purchasing behavior. Some customers may be acting strategically and waiting for lower prices and some customers may not. We assume that d(t) represents the overall demand we face as a function of the price we set.
- 2. If the firm is not able to fully satisfy demand at time t, a fraction ξ of the unsatisfied customers is backlogged with backlogging cost b per customer per unit time. Backlogged customers wait until they obtain the product, while the rest of the customers leave and never return. A complete lost sale case is captured by ξ = 0 and a complete backlogging case by ξ = 1. For backlogged customers, it is critical to specify which price is charged. We assume that customers pay the price when they place orders and the firm collects the payment immediately,³ but backlogged customers are not included in the total number of sales, S(t), that generates the word of mouth effect.
- 3. When the firm produces new units, backlogged customers are given priority over new customers in receiving the product. All backorders are satisfied at the end of horizon t = T at production cost.

During a finite horizon [0,T], a firm with fixed capacity K determines price $\pi(t)$,

³If customers paid the price at the time they receive the product, this could be easily gamed by the retailer. E.g., the retailer could accumulate high backlog and then force the backlogged customers to pay an extremely high price.

sales rate s(t), and instantaneous production rate x(t), in order to maximize total discounted profit with discount factor $r \geq 0$. Production incurs linear cost c and excess inventory I(t) is stored with linear holding cost h.

We assume that the price for the product at time t, $\pi(t)$, influences both the innovation coefficient p and imitation coefficient q in the same fashion, i.e., $p(\pi) = pe^{-\theta\pi}$, and $q(\pi) = qe^{-\theta\pi}$, as is typically assumed in the marketing literature (e.g., Dolan and Jeuland 1981, and Robinson and Lakhani 1975). To make the proofs easier to follow, we use $\theta = 1$.⁴ Thus, the production rate is constrained by capacity and price is the decision variable:

$$d(t) = \left[p + \frac{q}{m} S(t) \right] \left[m - D(t) \right] e^{-\pi(t)}.$$

We define $d_B(t) = \left[p + \frac{q}{m}S(t)\right][m - D(t)]$ and, therefore, $d(t) = d_B(t)e^{-\pi(t)}$. Clearly $\pi(t) = \log d_B(t) - \log d(t)$.

A discrete analog of our model would have the following interpretation. In period t, the firm's starting inventory is I(t) and backlog is W(t). The firm decides to produce quantity x(t), subject to capacity constraint K, and sets price $\pi(t)$ and observes demand d(t). It decides that it will sell a quantity s(t) that cannot exceed x(t) + I(t). Any left product becomes inventory in the next period I(t+1). If s(t) < W(t) + d(t), backlogged customers from period t are satisfied first and a fraction of the new customers $\xi[d(t) - (s(t) - W(t))^+]$ become the backlogged customers W(t+1) for period t+1.

We analyze the continuous-time model. In this model, we determine at all points

⁴Proofs for $\theta \neq 1$ are similar and do not change the structure of optimal demand and price.

in time t, the price level $\pi(t)$ and the production rate x(t). The price level determines the demand rate d(t) we observe and we also determine what sales rate s(t) to allow. We first focus on the instantaneous revenue collected at time t. Because backlogging has higher priority than new demand, there are no sales to new customers as long as backlog remains positive, W(t) > 0, i.e., each unit of production is immediately used to satisfy a backlogged customer and thus revenue is collected at a rate of $\xi \pi(t)d(t)$. (This is because we assume backlogged customers had already paid, and we only collect revenues from new customers who agree to be backlogged). If W(t) = 0, then $s(t) + \xi[d(t) - s(t)]$ customers will purchase, among whom $\xi[d(t) - s(t)]$ will be backlogged, and the collected revenue is $\pi(t)[s(t) + \xi(d(t) - s(t))]$. Once again, we assume that at time T, all backlogs are satisfied and any remaining inventory can be salvaged at c_s per unit. Thus, our objective function can be stated as follows:

$$\Pi = \max_{\pi(t), s(t), x(t)} \int_0^T \left\{ \pi(t) \left[\xi d(t) + (1 - \xi) s(t) 1_{\{W(t) = 0\}} \right] - cx(t) - hI(t) \right\}$$

$$-bW(t)\} e^{-rt}dt - cW(T)e^{-rT} + c_s I(T)e^{-rT}$$
 (2.1)

s.t.
$$d(t) = \left[p + \frac{q}{m}S(t)\right][m - D(t)]e^{-\pi(t)},$$
 (2.2)

$$\frac{dD}{dt} = d(t), (2.3)$$

$$\frac{dS}{dt} = s(t), (2.4)$$

$$\frac{dI}{dt} = x(t) - s(t), \tag{2.5}$$

$$\frac{dW}{dt} = \xi d(t) - s(t) + (1 - \xi)s(t)1_{\{W(t) = 0\}}, \tag{2.6}$$

$$I(t) \ge 0, W(t) \ge 0, 0 \le x(t) \le K, 0 \le s(t) 1_{\{W(t)=0\}} \le d(t),$$
 (2.7)

$$D(0) = S(0) = I(0) = W(0) = 0. (2.8)$$

The objective function (2.1) includes the revenue rate for new customers minus production cost, inventory holding cost, and backlogging cost. (2.2) describes the diffusion process. Equations (2.3-2.6) define cumulative demand, cumulative sales, inventory and backlogging dynamics. When backlog W(t) = 0, it (potentially) increases at a rate of $\xi[d(t) - s(t)]$, otherwise with positive backlog, fraction ξ of new demand is backlogged, while new production is used for previously backlogged demand. Constraints (2.7) require nonnegative inventory, nonnegative backlogging, production not to exceed available capacity, and sales rate smaller than demand rate in case of no backlog. Note that the constraints for s(t), when backlog occurs, are implicitly given by equations (2.5-2.7). Equation (2.8) provides the initial conditions. All models we consider are a special case of this general formulation.

2.4 Infinite Capacity

Before analyzing the case with both capacity constraint and dynamic pricing, we start with the problem with infinite capacity. We show that in this case, both demand and price trajectories have unimodal structures, and the demand peak (as a function of time) precedes the price peak, when the discount factor r is positive.⁵

Clearly, holding inventory is expensive and its only purpose is to hedge against future capacity shortage. Therefore, under infinite capacity, we will not hold inventory. Also, strategically backlogging some demand slows down the diffusion process and would only be useful when capacity is insufficient. Using price adjustments, we can decrease the sales (slow down the diffusion) and achieve higher profit. Thus, myopic

⁵Without discounting, the optimal demand is flat over the time horizon, see Lemma 4.

sales and production policies are optimal, i.e., $s^*(t) = x^*(t) = d^*(t)$ and the problem simplifies to:

$$\Pi = \max_{\pi(t)} \int_0^T [\pi(t) - c] d(t) e^{-rt} dt$$

Since price $\pi(t) = \log d_B(t) - \log d(t)$, the problem can be reformulated as finding the optimal demand $d^*(t)$:

$$\begin{split} \Pi &= & \max_{d(t)} \int_0^T [\log d_B(t) - \log d(t) - c] d(t) e^{-rt} dt \\ \text{s.t.} & & \frac{dD(t)}{dt} = d(t) \end{split}$$

In order to solve this control-theory problem, we use Hamiltonian H, which after incorporating factor e^{rt} into $\lambda(t)$ becomes:

$$H(D, d, \lambda, t) = (\log d_B(t) - \log d(t) - c)d(t) + \lambda(t)d(t)$$

Notice that, because S(t) = D(t), $d_B(t)$ is unimodal (negative quadratic) in D, and because D(t) is non-decreasing in t, we have that $d_B(t)$ is unimodal in t, and reaches its peak at time τ_B , such that $D(\tau_B) = (q - p) \frac{m}{2q}$. In the following lemmas, we show that the optimal demand and price are both unimodal in time, as illustrated in Figure 2.1. We further characterize conditions under which demand and price are monotonically increasing, monotonically decreasing, or increasing-decreasing. Finally, we demonstrate that, in the presence of positive discount factor, the demand peak always precedes the price peak.

Lemma 1 Assume discount factor r > 0 and capacity $K = +\infty$. Then, the optimal demand $d^*(t)$ is unimodal over time. It increases on [0,T] if the optimal cumulative

demand $D^*(T) < (q-p)\frac{m}{2q}$, decreases if $D^*(T) \ge (q-p)\frac{m}{2q}$ and $\lambda(0) \le 0$, and is increasing-decreasing otherwise.

Lemma 2 When discount factor r > 0 and capacity $K = +\infty$, the optimal price $\pi^*(t)$ is unimodal over time. It increases on [0,T] if $D^*(T) < (q-p)\frac{m}{2q}$, decreases if $D^*(T) \ge (q-p)\frac{m}{2q}$ and $r\lambda(0) - (q-p)e^{-\pi^*(0)} \ge 0$, and is increasing-decreasing otherwise.

Lemmas 1 and 2 describe the shape of demand and price trajectories when capacity is not limited and when discount factor is positive. Note first that the conditions for demand and price to be increasing over the whole horizon are the same $(D^*(T) < (q-p)\frac{m}{2q})$, which is most likely when the time horizon T is relatively short. If the selling season is long enough, usually both demand and price will eventually start decreasing.

Second, given the price and demand decrease at the end of the horizon, the lemmas describe, whether initial increase may take place or they decrease throughout the whole horizon. Clearly, demand decreases throughout the selling season if the marginal benefit of cumulative demand is already negative at the beginning of horizon $(\lambda(0) \leq 0)$, i.e., when the saturation effect already exceeds the diffusion effect. This typically takes place in a market with strong innovation effect (p) and weak imitation effect (q). The condition for price to drop throughout the time horizon typically requires a small discount factor and very weak imitation effect (q). A small discount factor ensures that it is not necessary to offer promotional price at the very beginning to boost instant revenue, and a market with weak imitation effect is quickly

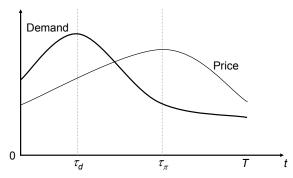


Figure 2.1. Optimal demand and price with unlimited capacity.

shrinking and thus price cut is needed later on.

Let τ_d and τ_{π} denote the time at which we reach the maximum (peak) of demand and of price, respectively. We show that they satisfy the following relationship.

Lemma 3 When discount factor r > 0 and capacity $K = +\infty$, the peaks of optimal demand, and price are ordered as follows: $\tau_d \leq \tau_{\pi}$. Furthermore, assume that the firm does not have pricing flexibility, and is forced to set the same fixed price π_F for all times $t \leq T$. Consider any constant price and let the time at which the peak demand is reached under the constant price be τ_F . Then we have that $\tau_d \leq \tau_F \leq \tau_{\pi}$.

With infinite capacity and positive discount factor, the firm prefers collecting revenue earlier, thus would like to stimulate the diffusion at a faster pace, making demand trajectory more negatively skewed.

Without discounting the profit, optimal demand and price can be derived in closed-form.

Lemma 4 If discount factor r = 0 and capacity $K = +\infty$, then for all $t \in [0, T]$, demand rate is constant, while price is correspondingly adjusted over time:

$$d(t) = \bar{d} = \frac{1}{2qT^2/m} \left\{ -(p-q)T - e^{1+c} + \sqrt{((p-q)T + e^{1+c})^2 + 4pqT^2} \right\}$$

$$\pi(t) = 1 + c + \log \frac{(m - \bar{d}t)(p + q\bar{d}t/m)}{(m - \bar{d}T)(p + q\bar{d}T/m)}$$
(2.9)

Furthermore, price is decreasing on [0,T] if and only if q < p.

Without discounting, the price structure is similar to that in Dolan and Jeuland (1981). Surprisingly, the optimal demand remains flat over the time horizon.⁶ That is, the firm should modify the price, according to (2.9), to maintain constant demand.

This may contradict the intuitive expectation that the firm might want to set a constant price over the whole time horizon when profit is not discounted. However, since the firm's myopic profit rate at time t is given by $(\pi(t) - c)d(t) = (\pi(t) - c)(m - D(t))(p + qD(t)/m)e^{-\pi(t)}$, constant price $\pi(t) = 1 + c$ maximizes only the firm's myopic profit, while ignoring the indirect effect of pricing on future product diffusion. This indirect effect takes place through cumulative demand. For example, a low price drives a faster increase in cumulative demand. Fast increase in cumulative demand is desirable at the beginning of the time horizon, as it translates into needed word-of-mouth effect, not desirable in the middle phase due to market saturation, and eventually not critical at the end of the time horizon. Therefore, if the time horizon is long enough, the firm first stimulates word-of-mouth effect through a low price, then charges a higher price, and a low price at the end of the life cycle to clear the market. Consequently, the demand should be much flatter than the case where the firm is forced to set a constant price for the whole time horizon.

It is easy to verify that $\pi(t)$ peaks at $\frac{q-p}{2aq/m}$. Therefore, price is decreasing throughout the whole life cycle (skimming) if and only if q < p, i.e., if imitation effect is dominated by innovation effect. Otherwise, we see initial penetration. This is easy to understand in the extreme case: without any imitators, q = 0, all sales are

⁶Dolan and Jeuland (1981) assume no discounting, but do not analyze the demand function.

driven by the innovators and, since the remaining market size decreases, the price correspondingly decreases. (The same logic applies to the general q < p case.)

Note that the lemma above has immediate implications for the interactions of marketing and operations in new product markets with fairly small discount rates. Obviously, the analysis of unlimited-capacity case applies to all cases where capacity is simply sufficient (not constraining the demand trajectory). Given that the "optimal" diffusion implies constant demand, a direct lesson is that the firm would build a facility which would serve a fairly constant demand over time, while all potential non-stationarities are absorbed by the pricing policy. The interesting fact here is that although it is generally recommended in the operations literature to level demand to help operations, in this case, it turns out to be optimal even when operations do not create a constraint. If discounting is significant, however, the firm is more impatient to satisfy customers and stimulates demand more aggressively in the early phase of product life cycle.

2.5 Limited Capacity

For the rest of the paper, we consider the more general situation, when capacity is constrained. We start with a case of lost sale and no inventory. In the following subsections, we incorporate inventory and backorders in the model. Recall that Kumar and Swaminathan (2003) allow for constrained capacity, but assume constant price. Their paper suggests that strategically holding sales to slow the diffusion of demand can be optimal. That is the firm may decide not to sell the product, even when holding some inventory of the product. We show that this never happens with dynamic

pricing, which emphasizes the strategic importance of pricing in new product introductions. The joint decisions of production quantity, sales level, and pricing makes the problem more complicated and we first describe the optimal demand and price trajectories. We will do this first for the case where unsatisfied demand is assumed to be completely lost and then for the most general case where partial backlogging is allowed.

2.5.1 Optimal demand and price trajectories under lost sales

We will now characterize the optimal pricing, production and sales trajectories over time. In this subsection, we assume that all unsatisfied demand is lost. We first characterize the optimal trajectories for the make-to-order case. Thus $\xi = 0$, x(t) =s(t), and W(t) = I(t) = 0.

Under make-to-order, our assumption is that if the demand rate for arriving customers exceeds our capacity rate, we experience lost sales. Clearly, lost sales cannot be optimal when we can control pricing. Note that allowing demand to be higher than available capacity results in lost sales, reduces future potential market, and does not create word-of-mouth effect. Thus, the firm is better off raising prices to reduce demand, which increases revenues and does not reduce future market potential.

Lemma 5 Assume products are made-to-order and unsatisfied demand is completely lost. The optimal production, sales, and demand are equal and never exceed capacity, i.e., $x^*(t) = s^*(t) = d^*(t) \leq K$.

As lost sales will never occur, price becomes the only decision variable, and equa-

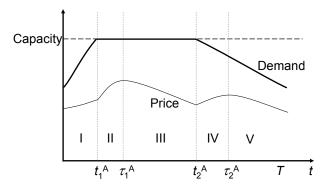


Figure 2.2. Optimal policies with complete lost sale in make-to-order environment

tions (2.1-2.8) are further simplified to

$$\Pi = \max_{d(t)} \int_0^T (\log d_B(t) - \log d(t) - c) d(t) e^{-rt} dt$$
 s.t.
$$\frac{dD}{dt} = d(t),$$

$$0 \le d(t) \le K, D(0) = 0.$$

Theorem 1 Assume that products are made to order and unsatisfied demand is completely lost. There exist time thresholds $0 \le t_1^A \le t_2^A \le T$, such that

- (a) Optimal demand $d^*(t)$ increases on $(0, t_1^A)$, $d^*(t) = K$ on (t_1^A, t_2^A) , and decreases on (t_2^A, T) .
- (b) Optimal price $\pi^*(t)$ increases on $(0, t_1^A)$, is unimodal on (t_1^A, t_2^A) , and unimodal on (t_2^A, T) .
- (c) There exists \bar{K} such that $t_1^A = t_2^A$ for all $K \geq \bar{K}$.

Theorem 1 states that, when capacity is insufficient, the optimal demand trajectory becomes "truncated unimodal" (there still exists only one demand peak but when this demand peak is equal to the firm's capacity, demand may equal capacity for an interval), while optimal price can be bimodal. Let τ_1^A and τ_2^A be the two price peaks. We can, accordingly, divide product life cycle into five phases as shown in Figure 2.2. In Phase I, $(0, t_1^A)$, the product has been just introduced with no or minimal word of mouth. Thus, demand is low and a low promotion price is offered in order to boost demand. This behavior continues until demand reaches capacity. In Phase II, (t_1^A, τ_1^A) , the firm's demand has picked up and may exceed its capacity. To avoid this, the firm raises the price in order to bring the demand rate down to its capacity rate. In Phase III, (τ_1^A, t_2^A) , a noticeable portion of the market has already been penetrated and the need for price premium to keep demand at the capacity level is first increased and then reduced. In Phase IV, (t_2^A, τ_2^A) , demand drops to a level where it is not necessary to increase prices to maintain sales at the capacity level. Since the effect of the word of mouth is less significant in this phase, the firm may in fact raise prices, as in case with infinite capacity. Finally, in Phase V, (τ_2^A, T) , the remaining market is very small, demand is decreasing and clearance prices are charged.

When capacity is sufficiently large, part (c) indicates that the constrained phase (t_1^A, t_2^A) disappears, and the price and demand trajectories are identical to those derived in the infinite capacity case.

Theorem 1 indicates that limited capacity significantly affects the structure of the pricing policy. Although the optimal demand behavior is similar to the classical Bass diffusion model, the optimal price is not unimodal anymore. The firm may charge a penetration price, and then charge a premium (possibly interpreted as "skimming") price when demand is high to bring it down to the capacity level. (Skimming would be strictly driven by limited capacity.) In later stages, when demand falls below capacity, it might be optimal to increase price again.

Having characterized the structure of optimal trajectories for the make-to-order case, we now extend our results to the make-to-stock case. We show that it is optimal for the firm to produce at the demand rate for an interval of time, and then to switch to producing at full capacity at a threshold time to build excess inventory. The firm uses this excess inventory to satisfy demand during the time when demand will exceed capacity. (Since inventory is allowed to be built, demand can exceed capacity at a certain interval of time.) The optimal price trajectory remains bimodal.

Previously, when holding inventory was not allowed, we had demand, production, and sales rates equal, $x^*(t) = s^*(t) = d^*(t) \le K$. When carrying inventory is allowed, the production rate can exceed the sales rate, $x^*(t) \ge s^*(t)$, and also the sales rate may exceed capacity when using inventory to satisfy demand, $s^*(t) \ge K$. While the model allows demand to exceed sales, we once again show that this can never occur under an optimal policy, i.e., there will be no lost sales.

Lemma 6 Assume a make-to-stock environment with complete lost sales. Then $d^*(t) = s^*(t)$, for all t.

Using the result from Lemma 6, we have that $S^*(t) = D^*(t)$, and recalling that

$$d_B(t) = [m - D(t)] \left[p + \frac{q}{m} D(t) \right], \text{ equations (1-8) become:}$$

$$\Pi = \max_{d(t), x(t)} \int_0^T [(\log d_B(t) - \log d(t)) d(t) - cx(t) - hI(t)] e^{-rt} dt + c_s I(T) e^{-rT}$$
s.t.
$$\frac{dD}{dt} = d(t),$$

$$\frac{dI}{dt} = x(t) - d(t),$$

$$I(t) \geq 0, I(0) = 0,$$

$$0 \leq x(t) \leq K,$$

$$D(0) = 0.$$

The Hamiltonian for this problem is:

$$H(D, I, d, x, \lambda_1, \lambda_2, t) = (\log d_B(t) - \log d(t))d(t) - cx(t) - hI(t)$$
$$+\lambda_1 d(t) + \lambda_2 (x(t) - d(t))$$
(2.10)

By Pontryagin's Maximum Principle, the optimal demand and production pair $(d^*(t), x^*(t))$ maximizes H subject to constraints $I(t) \geq 0$ and $0 \leq x(t) \leq K$. For any given λ_1 and λ_2 , partial derivatives of H w.r.t. d(t) and x(t) are:

$$\frac{\partial H}{\partial d} = \log d_B(t) - \log d(t) - 1 + \lambda_1(t) - \lambda_2(t)$$

$$\frac{\partial H}{\partial x} = -c + \lambda_2(t)$$

We note that $\lambda_1(t)$ can be interpreted as the discounted shadow price of the cumulative demand and $\lambda_2(t)$ can be interpreted as the discounted shadow price of inventory.

Due to the constraint on the state variable I(t), the analysis in this case is more complicated than that of make-to-order case in two aspects. First, the terminal value

 $\lambda_2(T)$ is not necessarily 0, but can take any nonnegative values, since remaining inventory at the end of the horizon can have some salvage value. Second, $\lambda_2(t)$ may be discontinuous. When the state constraint becomes binding, i.e., when I(t) transitions from positive to zero, $\lambda_2(t)$ can have a downward jump.

Lemma 7 For an open interval: (a) If $I^*(t) > 0$, then almost everywhere $x^*(t) = K$ and $\lambda_2(t) \ge c$. (b) If $I^*(t) = 0$, then almost everywhere $x^*(t) = d(t)$.

Intuitively, Lemma 7(a) states that a firm cannot produce less than its capacity and hold inventory at the same time. If a firm has inventory at the moment, and is producing less than capacity, it would have been better off by shifting some production from a previous time to the present. By doing so, demand can still be satisfied while production cost is delayed and inventory cost is decreased. For part (b), simply note that if inventory remains 0 for any open interval, production and demand must be equal.

Using similar logic, we show that it is not optimal to have positive inventory at the end of the horizon, since the excess inventory at the end of the horizon is of no use to the firm, while incurring extra holding cost and production cost.

Lemma 8 There is no inventory left at the end of the horizon, i.e., $I^*(T) = 0$.

Therefore, the optimal decisions, can be characterized as a function of inventory and the coefficients $\lambda_1(t)$ and $\lambda_2(t)$.

Lemma 9 Given I(t), $\lambda_1(t)$, and $\lambda_2(t)$, the optimal demand and production quantities are:

1. If
$$I^*(t) > 0$$
, then $\lambda_2(t) \ge c$, $x^*(t) = K$, and $\log d^*(t) = \log d_B(t) - 1 + \lambda_1(t) - \lambda_2(t)$.

- 2. If $I^*(t) = 0$ and $\lambda_2(t) \ge c$, then $x^*(t) = K$, $\log d^*(t) = \min\{\log K, \log d_B(t) 1 + \lambda_1(t) \lambda_2(t)\}$.
- 3. If $I^*(t) = 0$ and $\lambda_2(t) < c$, then $x^*(t) = d^*(t)$, $\log d^*(t) = \min\{\log K, \log d_B(t) 1 + \lambda_1(t) c\}$.

The production strategy is determined by the discounted shadow price of inventory λ_2 . When λ_2 is greater than the production cost, the firm produces as much as possible, which is full capacity. If λ_2 is smaller than the production cost, it is not beneficial to fully utilize capacity. In that case, the firm's production perfectly matches its current demand.

We are now ready to describe the whole trajectory of price, demand, and production.

Theorem 2 In a make-to-stock environment (with holding cost $h < +\infty$), there exist time thresholds $0 \le t_1^B \le t_2^B \le t_3^B \le T$, such that

- (a) Optimal demand $d^*(t)$ increases on $(0, t_1^B)$, is unimodal on (t_1^B, t_2^B) , $d^*(t) = K$ on (t_2^B, t_3^B) , and decreases on (t_3^B, T) .
- (b) Optimal price $\pi^*(t)$ increases on $(0, t_1^B)$, is unimodal on (t_1^B, t_3^B) , and is unimodal on (t_3^B, T) .
- (c) Optimal production $x^*(t) = d^*(t)$ on $(0, t_1^B) \cup (t_3^B, T)$, and $x^*(t) = K$ on (t_1^B, t_3^B) . Inventory $I^*(t) > 0$ and is unimodal on (t_1^B, t_2^B) and 0 otherwise.

Theorem 2, illustrated in Figure 2.3, states that with the ability to hold stock, the optimal demand and price trajectories are similar to those of make-to-order product, except for an intermediate phase (t_1^B, t_2^B) . At t_1^B , demand is still below capacity, but the firm starts producing at full capacity rate and accumulates inventory. Demand

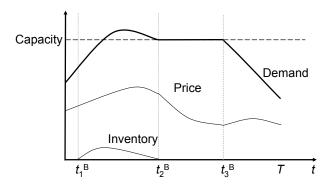


Figure 2.3. Optimal policies with complete lost sale in make-to-stock environment

at some time exceeds the firm's capacity and inventory is naturally depleted. At t_2^B , inventory is sold out and the firm continues selling at capacity rate in time interval (t_2^B, t_3^B) . Finally, demand drops below capacity after t_3^B . Essentially, the ability to hold inventory allows the firm to start producing larger amounts than demand early on so that it can sell more than what it is producing at a later time. Thus, pricing does not negate the need for inventory build-up. However, in this case, with pricing, we do not have to resort to withholding sales whereby we turn away customers even though we have inventory, which implies qualitatively different results from those of Kumar and Swaminathan (2003).

2.5.2 Optimal demand, and price trajectories under partial backlogging

When unsatisfied demand is completely lost, we showed that price is used to shape demand so that demand always equals sales and is never lost. We were also able to characterize the demand, price, and inventory trajectories. Consider now the most general case with partial backlogging and inventory - as defined in Section 2.3. In this general setting, the optimal demand can indeed exceed sales, resulting in some backlogged customers. We show that before any backlog appears, the firm builds some inventory and uses it when demand exceeds capacity (as in Section 2.5.1).

After inventory is depleted, selling more than capacity may continue, and the excess demand will be then backlogged. After demand drops below capacity level, the firm will continue to produce at full capacity level until the backlogged demand is cleared. We will first characterize the structure of the production and sales policies and compare them to those derived in Kumar and Swaminathan (2003) and Ho et al. (2002). What we will show is that, with pricing as an available option, a myopic sales policy that sells all the available inventory up to demand at any time is optimal. This is the policy derived in Ho without pricing as an available option, but we will in fact derive a counterexample to show that Ho's conclusion is invalid without pricing as an available control and that, as Kumar and Swaminathan predicted, without pricing, it is potentially optimal for the firm to withhold sales even if it is capable of satisfying the demand, something that would never happen with dynamic pricing.

We consider here the non-trivial case of partial or full backlog, $0 < \xi \le 1$ (lost sale $\xi = 0$ has already been analyzed in the previous sections). The problem formulation is exactly the same as in equations (2.1-2.8) and the Hamiltonian depends on whether the backlog is positive:

1. If W(t) = 0:

$$H(D, S, I, W, d, x, s, \lambda_1, \lambda_2, \lambda_3, \lambda_4, t) = (\log d_B(t) - \log d(t))(\xi d(t) + (1 - \xi)s(t))$$
$$-cx(t) - hI(t) - bW(t) + \lambda_1(t)d(t) + \lambda_2(t)s(t) + \lambda_3(t)(x(t) - s(t))$$
$$+\lambda_4(t)\xi(d(t) - s(t))$$

2. If W(t) > 0:

$$H(D, S, I, W, d, x, s, \lambda_1, \lambda_2, \lambda_3, \lambda_4, t) = (\log d_B(t) - \log d(t))\xi d(t) - cx(t)$$
$$-hI(t) - bW(t) + \lambda_1(t)d(t) + \lambda_2(t)s(t) + \lambda_3(t)(x(t) - s(t))$$
$$+\lambda_4(t)(\xi d(t) - s(t))$$

It is possible to interpret λ_1 , λ_2 , λ_3 , and λ_4 as, respectively, the discounted shadow prices of cumulative demand, cumulative sales, inventory, and backorders. We first show that it is not possible to have, at the same time, backlogged customer and positive inventory (as it is beneficial to use inventory to immediately satisfy backlog).

To avoid trivialities, we assume that b > rc, which implies that the backlogging cost is higher than what the firm can save by postponing production a unit of time. If this does not hold, the firm may have an incentive to intentionally delay all production till the end of the horizon.

Lemma 10 (a) If $I^*(t) > 0$ on an open interval, then $W^*(t) = 0$, $x^*(t) = K$ and $s^*(t) = d^*(t)$ almost everywhere.

(b) If b > rc and $W^*(t) > 0$ on an open interval, then $I^*(t) = 0$, $x^*(t) = s^*(t) = K$ almost everywhere.

(c) If $I^*(t) = W^*(t) = 0$ on an open interval, then $x^*(t) = s^*(t) = d^*(t)$ almost everywhere.

The lemma above states that although backlogging is possible, deliberately backlogging demand when inventories are available or capacity is sufficient, as shown optimal in Kumar and Swaminathan (2003), is no longer optimal when price is dynamically adjusted. Inventory, if available, is used to immediately clear any backo-

rders and, by the same token, if no inventory is available, then the firm uses current production to satisfy the backorders rather than build any inventory. Lemma 10 confirms that the production strategy used in Lemma 7 continues to hold when backlogging is allowed, that is, the firm produces at full capacity if it is optimal to keep positive inventory on hand.

Recall that in two previous papers, Kumar and Swaminathan and Ho et al. obtained conflicting results about the optimality of delaying sales in models where price is fixed. Kumar and Swaminathan showed that delaying sales can be optimal while in Proposition 1 of their paper, Ho et al., claimed that a myopic policy that never delays sales is always optimal as long as the firm has positive profit margins, and positive holding costs. Thus, Ho et al. claimed that it is never optimal for a firm to carry inventory and backlog (or turn away) customers at the same time while Kumar and Swaminathan showed the opposite. We provide a counter example that contradicts the conclusion of Proposition 1 in Ho et al.

To explain the counterexample, we first need to explain two variables that Ho et al. use in their model. First, they define a launch time t_l , where product is launched and demand diffusion can begin. Inventory, however, can be built before product launch. In our model, we can always achieve that by setting prices arbitrarily high, so that demand equals zero for some initial period of time. Thus, we do not need this additional variable. Ho et al. also allow backlogged customers to wait for an average of 1/l units of time before departing the system. Our counter example to Ho et al.'s result uses the following parameters: p = 0.03, q = 0.4, m = 3000, K=100, discount

factor r=0.01, holding cost h=0.001, profit margin a(t)=1.3 ($a(t)=\pi(t)-c$ in our setting) for all t, complete lost sales, so 1/l=0 (where 1/l is the average time that a backlogged customer will wait before departing the system) launch time $t_l=0$. These parameter values satisfy the condition of Proposition 1 in Ho et al. Solving the continuous time equation that these parameters result in (equations (11-16) in Ho et al.), we obtain the profit under the myopic sales policy which never delays sales to be 2270.07. However, we consider an alternative policy under which we intentionally delay sales by $\tau=5$ time units. That is, for $t\in[0,\tau]=[0,5]$, customers come to purchase, but the firm does not sell to them even if inventory is available. For $t\geq\tau$ a myopic sales policy is used. We obtain a profit of 2311.92. (Please see the Appendix for the derivation of profits.) Therefore, in this case a myopic sales policy is dominated by a policy that delays sales even when the firm has inventory. This is exactly what Kumar and Swaminathan showed.

This counterexample suggests that the myopic sales policy suggested in Ho et al. is not an optimal policy when price cannot be dynamically adjusted. In our counterexample, the firm has a small discount factor so delaying sales does not hurt profit very significantly. When capacity is limited, delaying sales slows down the diffusion process, thus reducing further sale losses from a sudden onslaught of demand that exceeds capacity. Our setting differs in that our model allows dynamic price changes, and we find that myopic sales policy is optimal when pricing flexibility is available. Clearly, when the firm has pricing flexibility, it can adjust price to control the diffusion process, without incurring the cost of losing sales. Therefore, pricing flexibility

acts as another lever to hedge against capacity shortage. An interesting question is how powerful pricing is as a tool compared to delaying sales or building inventory in advance by delaying the launching of the process as Kumar and Swaminathan and Ho et al. allow. We explore these questions in the next section.

We conclude this section by characterizing the optimal demand and price trajectories for the most general setting with partial backorders. We show that there is at most one backlogging interval, and the optimal price trajectory has at most three modes.

Theorem 3 In a make-to-stock environment with partial backlogging, if b > rc, and if demand is continuous, then there is at most one backordering interval, and the optimal price has at most three local maxima. There exist time thresholds $0 \le t_1^C \le t_2^C \le t_3^C \le t_4^C \le T$, such that

- (a) Optimal demand $d^*(t) \leq K$ and increases on $(0, t_1^C)$, $d^*(t)$ is unimodal on (t_1^C, t_2^C) , $d^*(t) = K$ on (t_2^C, t_3^C) , $d^*(t)$ is unimodal on (t_3^C, t_4^C) , and $d^*(t) \leq K$ and decreases on (t_4^C, T) .
- (b) Optimal price $\pi^*(t)$ is unimodal on $(0, t_3^C)$, (t_3^C, t_4^C) , and (t_4^C, T) .
- (c) Optimal production $x^*(t) = d^*(t)$ on $(0, t_1^C) \cup (t_4^C, T)$, and $x^*(t) = K$ on (t_1^C, t_4^C) . Inventory I(t) > 0 and is unimodal on (t_1^C, t_2^C) .
- (d) Optimal sales $s^*(t) = d^*(t)$ on $(0, t_3^C) \cup (t_4^C, T)$, and $s^*(t) = K$ on (t_3^C, t_4^C) . Backlog W(t) > 0 on (t_3^C, t_4^C) .

In the special case where the product is made to order (holding cost $h = +\infty$), then $t_1^C = t_2^C$ and the optimal demand trajectory is unimodal over time on [0,T], and the

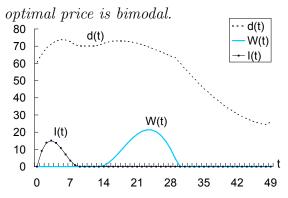


Figure 2.4. Optimal demand, inventory and backorder trajectories with make to stock and partial backordering

Figure 2.5. Optimal price trajectories with make to stock and partial backordering

The option of backorders, either partial or full, complicates the firm's pricing strategies. A typical situation is shown in Figures 2.4 and 2.5, where the selling season includes a positive inventory period and a positive backorder period. Parameters used are $m = 3000, p = 0.03, q = 0.4, c = 70, \alpha = 0.05, \beta = 0.9, h = 0.01, w = 0.01, T = 50$, where β is the discrete discount factor corresponding to r in continuous time. During the positive inventory period, the firm first produces at full capacity to build up some inventory while capacity is sufficient, and then sell more than capacity using the inventory prepared. After inventory is depleted, the firm may sell at capacity level for a while. And then, when the firm anticipates that the capacity-constrained phase will end in the near future, it may decide to sell more than it can offer (from inventory or from production) and hold some of the excess demand as backorders.

Theorem 3 indicates that a tri-modal price may be a possibility while Figure 2.5 shows an example with a bimodal price. (We have actually not succeeded in creating a numerical example with a tri-modal pricing structure). However, in the marketing literature and in fact in practice, it is not that easy to see products that

had a trimodal or even a bimodal pricing structure during their life cycle. As we stated in the introduction, the most commonly encountered pricing structures are skimming pricing where the prices only decrease over time and penetration pricing which essentially is unimodal. In fact, our numerical results in the next section indicate that penetration pricing is very close to optimal when used in conjunction with optimal inventory and production policies,

2.6 Numerical Study

In the last section, we have characterized the optimal demand, price, inventory and backorder trajectories under different situations. Although the firm can use price, preparing inventory, or allowing backorders to hedge against capacity shortage, it is not clear how effective these tools are. We conduct numerical study to compare the benefit of these tools.

We have shown that the optimal price trajectory has a complicated bimodal (or potentially tri-modal) structure. The interesting question is how much benefit the firm has gained from using this complicated structure as opposed to simpler structures, which we also investigate in this section.

Similar to Kumar and Swaminathan (2003), we solve the discrete-time version of the continuous-time model. For all results presented in this section, unless otherwise stated, we use the following parameters: m = 3000, p = 0.03, q = 0.4, c = 0.5, r = 0.1, h = 0.01, b = 0.005, $\xi = 1$, and T = 50 periods.

• Value of pricing flexibility.

Pricing flexibility acts as a lever to hedge against capacity shortage. We define

the value of pricing flexibility as

 $\frac{\text{Optimal profit with dynamic pricing} - \text{Profit with the optimal fixed price}}{\text{Profit with the optimal fixed price}} \times 100\%$

Figure 2.6 shows that, as capacity increases, the value of pricing flexibility first increases and then decreases and, thus, it is most beneficial when the firm's capacity is in the intermediate range. Intuitively, if the firm has sufficient capacity, demand is seldom constrained and thus price, while actively used, has moderate variation, driven solely by revenue maximization. When the firm has low capacity, demand is always constrained, so use of price is very important, but it is used solely to shape demand to fit available capacity and, therefore, price does not vary dramatically. In-between, both of these reasons exist and the seller more actively accelerates and decelerates demand to maximize revenue and take advantage of slack capacity in the beginning and the end of product life cycle, while constraining the sales in the middle of product life cycle.

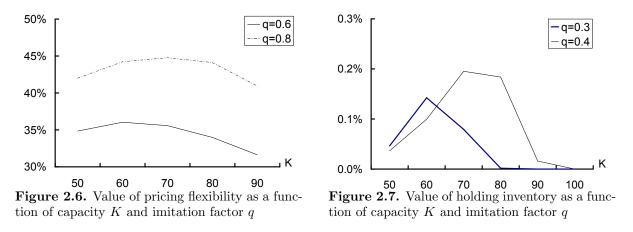


Figure 2.6 also shows that the value of pricing flexibility increases in the imitation factor q. As the customers are more easily influenced by other people's purchases, a low promotion price can create a significant word-of-mouth effect

and thus is more helpful.

• Value of holding inventory. Figure 2.7 shows the value of holding inventory, which is defined similarly, as the percentage profit increase from make-to-order environment to make-to-stock environment. Similar to the value of pricing flexibility, the value of holding inventory is also highest for intermediate range of capacity. However, compared to value of pricing flexibility (20%-50%), value of holding inventory is very small (0-0.2%), which implies that carrying inventory does not help much when price adjustments are a viable tool, even for a firm with potential capacity constraints.

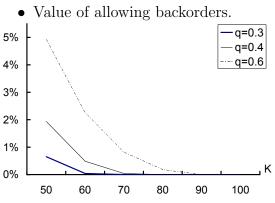


Figure 2.8. Value of allowing backorder

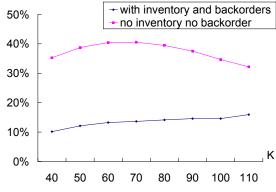


Figure 2.9. Optimality gap of skimming pricing strategies

Figure 2.8 illustrates the value of allowing backorders, defined as the percentage profit increase when backlogging is allowed as compared to the complete lost sales case. The value of allowing backorders is typically larger than that of holding inventory, and it decreases in capacity. Clearly, when capacity is large, backorders/lostsale do not occur very often. When capacity is small, it may, however, be significant, as shown in Figure 2.8. Similarly, it is higher for bigger imitation factors, q.

• Comparison of the optimal pricing strategy with simple pricing strategies

Although the marketing literature has suggested that the optimal price should
have either a skimming or a penetration structure, they have assumed that the
firm has infinite capacity. We have shown that when the firm has capacity
constraints, the optimal pricing strategies may have a complicated structure.

However, implementing such a complicated dynamic pricing structure may be
difficult and explaining it to customers may be even harder. Therefore, we compare the optimal pricing strategy with the simple pricing strategies suggested
in the marketing literature (when those simple policies are used in conjunction
with optimal production/inventory policies) and examine the profit implications.

We define a skimming pricing strategy as the price decreasing over time, and a penetration pricing strategy as the price increasing and then decreasing over time, thus the penetration pricing trajectory is unimodal. We define the optimality gap for skimming price as:

Profit with the optimal price — Profit with the optimal skimming price

Profit with the optimal skimming price

Similarly, we define the optimality gap for penetration price as:

Profit with the optimal price – Profit with the optimal penetration price

Profit with the optimal penetration price

Figures 2.9 and 2.10 show the optimality gaps for skimming and penetration price policies, respectively. The parameters used are $m = 3000, p = 0.03, q = 0.3, c = 0.05, r = 0.1, h = 0, b = 0.01, \xi = 1, T = 50$. We can see that for

some parameters, skimming pricing strategy still does not perform very well, and the profit difference can be 10% to 30%, although the profit difference is smaller when the firm also has other levers such as holding inventory and allowing backorders. On the other hand, the penetration strategy performs very close to the optimal strategy. This is very encouraging since we have shown that even though the optimal price trajectory may have multiple modes, the simple unimodal strategy, suggested in marketing literature without capacity constraints, actually works well in the situation of capacity constraints. Therefore, managers can avoid the frustration of using a complicated pricing policy as long as they focus on coordinating optimal production/inventory policies with the right penetration pricing.

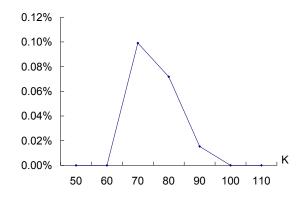


Figure 2.10. Optimality gap of penetration pricing strategies

2.7 Conclusions

In this work, we analyzed a capacity-constrained firm introducing a new product and making decisions about how to price and manufacture the product for the most profitable diffusion of the product into the marketplace. Our model generalizes Kumar and Swaminathan (2003) and Ho et al. (2002) by incorporating price decisions. As

multiple decisions about production, sales, and pricing interact with each other, we examined these interactions in several situations, starting from the simplest ones, to eventually include complete lost sales and partial backlogging, and make-to-order and make-to-stock environments.

While marketing literature had considered pricing decisions, it had typically ignored capacity constraints. Our first contribution was to derive the structure of integrated pricing, production and inventory policies under capacity constraints. We demonstrated that the presence of supply constraints may result in fairly complex optimal pricing strategies as well as production/inventory policies. Thus, the penetration and skimming pricing policies suggested in marketing are no longer optimal.

While the operations literature had considered capacity, it had ignored pricing. Compared to Kumar and Swaminathan (2003), our results suggest that pricing flexibility is a very effective lever to use when facing capacity constraints. Moreover, when dynamic pricing is used, intentionally holding sales is not necessarily optimal any longer, as pricing flexibility allows the firm to delay diffusion process without losing any profits.

We examined the benefit of pricing flexibility and found that it is most effective when the firm has intermediate range of capacity, and when imitation effects are high. Finally, we showed that a penetration pricing policy, while no longer optimal, does perform very well when used in conjunction with optimal production/inventory policies.

CHAPTER 3

Optimal Production and Ordering Policies with Random Yield Production and Queueing

3.1 Introduction

In this paper, we address a problem commonly encountered in the pharmaceutical industry. Pharmaceutical firms manufacture a batch of drugs for testing purposes in job-shop like environments called *kilo labs* or *pilot plants*. Since most compounds are found to be toxic or ineffective in initial trials, they are manufactured only once. Furthermore, in this initial manufacturing process, yields tend to be fairly low and variable. The chemicals required to make the active ingredients in the drug are often extremely expensive (it is common for them to cost many times the cost of gold).

Since yields are random, the firm procures more chemicals and starts a larger batch than the amount that will be needed in the trials. The major tradeoff that is faced is as follows: If the firm starts a much larger quantity than needed, it will have wasted a lot of expensive chemicals; on the other hand if the yield turns out to be too low to yield a quantity sufficient for the animal and/or clinical trials, then the trial has to be postponed until the firm can produce a second (or third etc.) batch until the required amount is obtained. Delaying animal and clinical trials is very

expensive¹. Thus, the firm faces a problem of determining the optimal batch size for the particular phase of clinical or animal trials.

The arrival of a batch of a new compound request to the kilo lab is a random process as it depends on when the researchers (and the firm management) feel that the compound is promising enough to merit a test. Thus, the kilo lab may face a situation where it receives a large number of requests in a given month. Given this variability in the arrival process and the resulting delays, it is clear that the batch size should depend on the number of batches waiting to be manufactured. Taking a risk, which results in having to produce a second batch, is much more expensive if there are many batches (and therefore many trials) being delayed. Our experience with the industry however is that it is common for pharmaceutical firms to use fairly simplistic policies for determining batch sizes. Common policies inflate the amount needed for the clinical trial by dividing the amount needed for the clinical/animal trial by the average or minimum percentage yield the researchers observed in their labs. Apart from the fact that the yields obtained in the researchers' labs may not be perfectly representative of the yields obtained in the pilot plants, the main problem with the simplistic models is that they do not take into account the costs of delays or materials or the number of compounds waiting to be manufactured (i.e., how loaded the plant is). We therefore model this problem to obtain an understanding of optimal batch sizes and their dependence on these factors.

¹Some firms do not postpone clinical trials but schedule them far enough that even with very long production lead times the trial can still start on time. Of course, this kind of delay also has significant costs as delaying clinical trials in this manner still delays the time when the product can be introduced to the market, potentially costing millions of dollars

Versions of the problem that we address have been recognized in the literature and "Multiple Lot sizing in Production to Order system with rigid demand" (MLPO). Most of the studies assume a known probability distribution of the yield, given the input lot size, and a setup cost for each production run. For every production run, lot size is determined to minimize the total material and setup costs. Most of the pharmaceutical production processes yield random outputs and have rigid demand and are MLPO processes. The demand is often rigid because failing to provide sufficient volume of drugs will delay the trials. However, the pharmaceutical production process is characterized by some properties not studied before. First, the effect of limited capacity combined with rigid demand, which often results in queueing, has not been studied in the literature. In practice, capacity is often scarce since the facilities to make these batches are expensive and even the largest pharmaceutical companies such as Pfizer or Merck have only a few kilo labs where they can make such batches. Clearly more capacity is needed if smaller batch sizes are used as more production runs are needed. When there are a lot of batches waiting to be manufactured, smaller batch sizes can result in queueing and long delays. The existing literature only provides solutions to the MLPO problem for a single customer, while ignoring the externality effect of serving one customer on the other waiting orders. In fact, the lack of kilo lab capacity in the large pharmaceutical companies has led to the existence of contract manufacturers that manufacture batches for all pharmaceutical firms on demand. A common decision that the firms have to make is when to outsource a particular batch to the contract manufacturer because the kilo lab is

already too overloaded. We also model this decision which has not been addressed in these problems.

The rest of this paper is organized as follows. In Section 2, we provide a literature review. Section 3 first provides a basic model with only production decisions and assuming infinite raw materials, thus focusing on the tradeoff between material expense and delay expense. We show the existence and the structure of the optimal policy and also provide managerial insights on the factors that affect the batch size. We conduct a numerical study to compare the optimal policy with heuristic policies commonly used in the pharmaceutical industry. We then extend the model to the case where the firm also has the option to outsource a batch to a contract manufacturer or run overtime productions. Finally, we summarize our work in Section 4.

3.2 Literature Review

Our research is related to two streams of literature: queueing control models and MLPO random yield models, and is the first one that belongs to both streams, while having differences with previous work in both fields.

The first stream of relevant research is traditional queueing control. A typical problem in this literature is that a machine can be operated at different rates with different costs, and waiting orders or materials incur inventory holding costs. In some cases a switching cost is also charged when the production rate changes. The objective is to find an optimal policy to control the production rate, based on the length of the queue. For detailed review of related work see Grabill, Gross and Magazine (1977). Our model essentially controls the production rate by changing lot sizes, but

it differs from previous work as the production rate is not completely controllable. Even though lot size is used to control the rate, the number of production runs to complete the service remains uncertain in our setting.

The second stream of research, random yield problems, has received considerable amount of attention since the 1980s. An extensive review of literature on random yield production was provided by Yano and Lee (1995). They summarize various models of production policies with the objective of minimizing costs, which often include setup costs, salvage costs, inventory holding costs, and backlogging or lost sale penalty costs, etc. Some papers study rigid demand problems, while others assume that production takes place only once or a fixed number of times for each customer. For example, Shih (1980) generalizes the EOQ model and Henig and Gerchak (1990) generalize the newsvendor model to include yield uncertainty, which was further generalized to a variation of DP models for multi-period problem by Moinzadeh (1987). The optimal production policies with random yield have a critical order point below which it is optimal to order, but the order-up-to level is not constant. A review of the subgroup of random yield models, the MLPO models, is provided by Grosfeld-Nir and Gerchak (2004). In the basic model they assume that multiple production runs have to be completed in order to fulfill the entire order of a single customer. Each production attempt is costly for the firm, due to equipment setup costs and labor cost, etc. While it may be natural to increase production output, reduce the number of repeated attempts, and order a larger amount of materials. However, this may result in unnecessary production. The key question in the related research is the tradeoff between batch size (and the potential of overproduction) and the number of production runs.

One of the earliest papers on MLPO is Beja (1997), where a single stage MLPO with binomial yield is analyzed. He proves intuitive properties of the optimal production policy, such as the monotonicity of lot size in demand. Later research has focused mostly on generalizing this model to more general yield distributions, including both discrete and continuous ones. Typical discrete yield distributions are Binomial (Beja (1997), Interrupted Geometric (Porteus (1990) and (1986)), Uniform (Anily (1993)), and All-or-Nothing (Grosfeld-Nir and Gerchak (2004)). For continuous yield distribution, on the other hand, the optimal lot size cannot be computed recursively and thus has received less attention. To the best of our knowledge, the only MLPO model with continuous distribution is by Grosfeld-Nir and Gerchak (1990). They assume a power distribution and provide closed-form solutions for special cases with power parameters k = 1 and 2. Our model assumes a more general yield distribution.

Our model also generalizes the previous models to include continuous yield distributions and, at the same time, incorporate queueing effects. These effects have not been studied before in the MLPO literature, and the optimal solution to the tradeoff between the material cost and delay cost remains unclear in systems with random yield, rigid demand, and order queueing.

Almost all of the MLPO papers focus on material cost savings, but ignore the effect of multiple production runs on customer delays. Most papers assume single-demand problems, in which only one customer requires to be served. Orders' waiting

time, however, is often an important performance measure and is associated with costs in a make-to-order environment. Reduction in the number of production runs not only shortens the waiting time for the customer being served, but creates an external effect of reducing all other orders' waiting time.

There exist some random yield models with multiple orders, but with the restrictive assumption of only limited number of production runs. Most models include assumptions of one production run per period, linear holding and shortage costs, complete backlogging and, with one exception, zero lead times (Yano and Lee (1995)). None of them studies rigid demand in make-to-order system.

We next present our basic model that looks at the tradeoff between material costs and order delay costs.

3.3 Basic Model

We study a make-to-order production system with a single server. Orders arrive according to a Poisson process with rate λ , each of whom demands \bar{D} . Products are unique to orders, therefore the product is made to order. Orders are served FCFS, and only one customer can be served at one time. We assume that orders wait until getting the entire demand. Any excessive production has to be discarded and cannot be used in the future.

The production process yields random output. While most of the continuous yield literature is restricted to stochastically proportional yield, meaning the portion of production output to input follows a common distribution, we allow a more general distribution. We also allow yield distribution to be non-stationary. The yield

distribution is often improved from one trial to another. This is because the firm's chemists learn about the root cause of yield loss, and use this information in consecutive runs. Given lot size Q and the production run number i of a particular order, we define Y to be the amount of product obtained, with cdf F(i,Q,y) and pdf $f(i,Q,y) = \frac{\partial F(i,Q,y)}{\partial y}$. Grosfeld-Nir and Gerchak (1990) assume the yield proportion has a support [0,1], while we assume the support of F is [a,b], $0 \le a \le b \le 1$. Although this does not change the analytical results, we feel it is more realistic for pharmaceutical production processes where the yield rate has a minimum lower bound. If the products obtained from one production run is not sufficient, another production run is restarted, and the process continues until sufficient products are obtained. We assume F and f are twice continuous differentiable.

The time for each production run is a random variable T, and is assumed i.i.d with cdf G(.), pdf g(.) and mean $\frac{1}{\mu}$. This is consistent with the pharmaceutical production processes, where the production time is often determined by the chemical reaction time in the reactor. G, g and μ are the independent of lot size Q. We assume $\lambda < \mu$ for system stability.

The costs we consider include linear material cost c_M per unit material used, linear customer waiting cost c_W per unit time per order, and resource cost c_R per unit time for the resources (e.g., machine, labor, etc.) used to make the products. Setup cost for each production run is ignored, but it can be easily included without changing the results significantly.

A batch size Q needs to be determined before starting each production run. If Q

is too large, one or two production runs may be able to provide sufficient products, and this accelerates service speed at the cost of large material waste. On the other hand, if Q is too small, the desired amount can be obtained with less material waste, but this may cause a long queue of waiting orders. So the objective is to find the optimal production lot size Q^* that minimizes the expected long-run average cost.

Intuitively, lot size decisions are influenced by demand and how heavily the system is loaded. So we define the state variables as:

(i, D, N) = (production run number of the first order, remaining demand of the first order, number of orders in the system). $(i, D, N) \in N \times [0, \bar{D}] \times Z^+$.

We further restrict the decision space. Since any $Q \geq \frac{D}{a}$ will produce enough quantity to meet demand, without loss of generality we can restrict attention to $Q \leq \frac{D}{a}$. Furthermore, we assume $Q \geq \frac{D}{b}$ so that there is positive probability to obtain more than demand with one production run. When b=1 this is consistent with the assumption $Q \geq D$ in Grosfeld-Nir and Gerchak (1990). One can create possible instances where Q > D may not be optimal when the delay costs are trivially small, but in practice the delay costs are significant enough that one would never produce a quantity less than demand. In fact we believe that this assumption is even more reasonable in our setting compared to settings where delay costs were not considered. So the decision space is $\Omega(D) = \left[\frac{D}{b}, \frac{D}{a}\right]$, if a=0, then $\Omega(D) = \left[\frac{D}{b}, +\infty\right)$.

We determine lot sizes for each production run in order to minimize total material and waiting costs. Let $\Gamma(T)$ be the number of production runs started until time T, and $1_{N(t)}$ be an indicator function, then the objective is to minimize the expected average cost:

$$g = \min_{Q_j \in \Omega(D_j)} \lim_{T \to \infty} \frac{1}{T} E \left[\sum_{j=1}^{\Gamma(T)} c_M Q_j + \int_0^T c_W N_t + c_R 1_{N(t)} dt \right]$$

The problem can be formulated as a Markov Decision Process. Decisions are made after each production run or when a customer arrives in an empty system. If there are N orders in the system, the first order requires D more units to complete the order, and it is the ith production run for the first order, then after one production run, if the yield Y is greater than D, the new state is $(1, \bar{D}, N + N' - 1)$, where N' is the number of orders that arrived during the production process. If the yield Y is smaller than D, then the new state is (i + 1, D - Y, N + N'). The distribution of N' is $Prob\ (N' = n) = E_T\ Prob\ (N' = n|T) = \int_0^\infty \frac{(\lambda s)^n e^{-\lambda s}}{n!} dG(s)$.

For each production run, the associated costs are material cost, waiting cost of the orders in the queue, and waiting cost of the orders who just arrived during this production run. Material cost is straightforward and defined to be $c_M Q$; the expected waiting cost of orders in the queue during the production run is $W = c_W/\mu$ per order; and the expected waiting cost of the new arrivals W_{new} can be computed in the following way: Given production time T, the expected number of arrivals is $\lambda * T$. Since the arrivals follow a Poisson process, these arrivals are i.i.d uniformly distributed on [0, T], thus the expected waiting time for each order is T/2. Therefore the expected total waiting time of new arrivals is $W_{new} = c_W \lambda * E(T^2/2)$. Note that both W and W_{new} are independent of i, D, N and Q.

We can write the optimality equations as the following:

$$V(i, D, 0) + g * 1/\lambda = V(1, \bar{D}, 1);$$
 (3.1)

$$V(i, D, N) + g * 1/\mu = \min_{Q \in \Omega(D)} \{N * W + W_{new} + c_R/\mu + c_M Q + (3.2)\}$$

$$\sum_{N'=0}^{\infty} \text{Prob}(N') \left[\bar{F}(i, Q, D) V(1, \bar{D}, N + N' - 1) \right]$$
 (3.3)

$$+ \int_{aQ}^{D} V(i+1, D-y, N+N') dF(i, Q, y) \bigg] \}, \qquad (3.4)$$

for all
$$i \in N, D \in (0, \overline{D}], N \ge 1$$
.

We also define $V(i, 0, N) = \lim_{D \to 0^+} V(i, D, N)$.

The three terms in (3.2) include the waiting costs for orders already in queue, the waiting costs of new arrivals, and the material costs. (3.3) is the cost-to-go when yield is large enough to satisfy the current order after this production run. Finally, (3.4) is the cost-to-go when yield from this production run is not enough to satisfy the current order.

We show that a unique optimal production policy exists and also provide the structure. We define a cost difference function $H(i,D,N) = V(i,D,N) - V(1,\bar{D},N-1)$ for all $i \in N, D \in [0,\bar{D}], N \geq 1$. For $N \geq 1$, we can rearrange the terms and use only H(i,D,N) functions with $N \geq 1$ in the optimality equations.

$$V(i, D, N) - V(1, \bar{D}, N - 1) + g/\mu = \min_{Q \in \Omega(D)} \{N * W + W_{new} + c_R/\mu + c_M Q + \sum_{N'=0}^{\infty} \text{Prob}(N')\bar{F}(i, Q, D)V(1, \bar{D}, N + N' - 1) + \sum_{N'=0}^{\infty} \text{Prob}(N') \int_{aQ}^{D} V(i + 1, D - y, N + N') dF(i, Q, y) - V(1, \bar{D}, N - 1)\} (3.5)$$

$$= \min_{Q \in \Omega(D)} \{ N * W + W_{new} + c_R/\mu + c_M Q$$

$$+ \sum_{N'=0}^{\infty} \operatorname{Prob}(N') (1 - \int_{aQ}^{D} dF(i, Q, y)) V(1, \bar{D}, N + N' - 1)$$

$$+ \sum_{N'=0}^{\infty} \operatorname{Prob}(N') \int_{aQ}^{D} V(i + 1, D - y, N + N') dF(i, Q, y)$$

$$- \sum_{N'=0}^{\infty} \operatorname{Prob}(N') V(1, \bar{D}, N - 1) \}$$
(3.6)

$$= \min_{Q \in \Omega(D)} \{ N * W + W_{new} + c_R/\mu + c_M Q + \sum_{N'=0}^{\infty} \text{Prob}(N') \left[V(1, \bar{D}, N + N' - 1) - V(1, \bar{D}, N - 1)) \right] + \sum_{N'=0}^{\infty} \text{Prob}(N') \int_{aQ}^{D} \left(V(i+1, D-y, N + N') - V(1, \bar{D}, N + N' - 1) \right) dF(i, Q, y) \}$$

$$= \min_{Q \in \Omega(D)} \{ N * W + W_{new} + c_R/\mu + c_M Q$$

$$+ \sum_{N'=0}^{\infty} \operatorname{Prob}(N')((V(1, \bar{D}, N + N' - 1) - V(1, \bar{D}, N + N' - 2)$$

$$+ (V(1, \bar{D}, N + N' - 2) - V(1, \bar{D}, N + N' - 3))$$

$$+ \dots + (V(1, \bar{D}, N) - V(\bar{D}, N - 1)))$$

$$+ \sum_{N'=0}^{\infty} \operatorname{Prob}(N') \int_{aQ}^{D} (V(i + 1, D - y, N + N')$$

$$- V(1, \bar{D}, N + N' - 1)) dF(i, Q, y) \}.$$

$$(3.7)$$

Replace $V(i, D, N) - V(1, \overline{D}, N - 1)$ by H(i, D, N), then

$$H(i, D, N) + g/\mu = N * W + W_{new} + c_R/\mu + \sum_{N'=0}^{\infty} \text{Prob}(N') \sum_{j=N}^{N+N'-1} H(1, \bar{D}, j)$$

$$+ \min_{Q \in \Omega(D)} \{ c_M Q + \sum_{N'=0}^{\infty} \text{Prob}(N')$$

$$\int_{aQ}^{D} H(i+1, D-y, N+N') dF(i, Q, y) \}.$$
(3.8)

(3.9)

Notice that the first three terms in (3.8) do not depend on Q while the last two terms do depend on Q.

3.3.1 Optimal Lot Size Decisions

Intuitively, we expect a larger lot size if more people are waiting or if the quantity needed to satisfy the first order increases. Theorem 4 provides conditions under which a unique optimal policy exists and increases in both demand and queue length.

Theorem 4 There exists a unique optimal policy $Q^*(i, D, N)$ increasing with D and N for each i, if for each i (a) F(i, Q, y) is decreasing and convex in Q, and (b) F(i, Q, y) is submodular in (Q, y).

Proof: All proofs are presented in appendices.

It is easy to find yield distributions that satisfy conditions in Theorem 4. For example, consider stochastically proportional yield distributions, where the yield rate is independent of lot size, i.e., Y/Q follows the same distribution for any Q. In mathematical terms, if there exists some probability distribution P(x), such that F(i, Q, y) can be written as $P_i\left(\frac{y}{Q}\right)$ for any i, Q, then F is a stochastically proportional yield distribution.

For a stochastically proportional yield distribution, $F(i, Q, y) = P_i\left(\frac{y}{Q}\right)$, then condition (a) and (b) in Theorem 4 become:

$$F_Q'(i,Q,y) = P_i'\left(\frac{y}{Q}\right)\frac{-y}{Q^2} \le 0; \tag{3.10}$$

$$F_Q''(i,Q,y) = P_i'\left(\frac{y}{Q}\right)\frac{2y}{Q^3} + P_i''\left(\frac{y}{Q}\right)\left(\frac{-y}{Q^2}\right)^2 = \frac{y}{Q^3}\left(2P_i'\left(\frac{y}{Q}\right) + P_i''\left(\frac{y}{Q}\right)\frac{y}{Q}\right) \ge 0;$$
(3.11)

$$F_{Qy}''(i,Q,y) = P_i'\left(\frac{y}{Q}\right)\frac{-1}{Q^2} + P_i''\left(\frac{y}{Q}\right)\frac{-y}{Q^2}\frac{1}{Q} = \frac{-1}{Q^2}\left(P_i'\left(\frac{y}{Q}\right) + P_i''\left(\frac{y}{Q}\right)\frac{y}{Q}\right) \le 0.$$

$$(3.12)$$

Therefore for a stochastically proportional yield distribution, if $P'_i(x) + xP''_i(x) \ge 0$ for any $x \in [a, b]$ and i, i.e., if $xP'_i(x)$ is increasing in x, then condition (a) and (b) in Theorem 4 are satisfied.

The condition that $xP_i'(x)$ is increasing is satisfied by any stochastically proportional yield distributions with increasing densities, such as power distributions with pdf $P(x) = x^r, x \in [0,1], r \geq 1$, scaled power distributions with pdf $P(x) = \left(\frac{x-a}{1-a}\right)^r, x \in [a,1], r \geq 1$, and exponential pdf $P(x) = \frac{e^x-1}{e-1}, x \in [0,1].$ $xP_i'(x)$ is also increasing for some stochastically proportional yield distributions with non-monotone densities, such as the incomplete triangular distribution

$$P_i'(x) = \begin{cases} 3x, & x \in \left[0, \frac{1}{2}\right]; \\ 2 - x, & x \in \left[\frac{1}{2}, 1\right], \end{cases} \text{ as well as some with decreasing densities, such as}$$

$$P_i'(x) = \frac{5}{4} - \frac{1}{2}x, \, x \in [0, 1].$$

If yield rate Y/Q is dependent on Q, we can also find distributions that satisfy condition (a) and (b) in Theorem 4. An example is

$$F(i,Q,y) = \begin{cases} \frac{y}{Q}, & \text{for } Q \le 1, 0 \le y \le Q; \\ \left(\frac{y}{Q}\right)^{Q}, & \text{for } Q > 1, 0 \le y \le Q. \end{cases}$$

We are interested in the relative inflation rate, which is the ratio of optimal lot size over demand, Q^*/D . If yield follows stochastically proportional power distribution, we show that although the optimal lot size increases in demand, the inflation rate decreases in D. With smaller demand, a relatively large lot can be used to accelerate service, since the cost of wasting material is small. When demand grows larger, it may be better to try several production runs with lot sizes close to demand, to reduce the risk of poor yield. The result is summarized in Theorem 5.

Theorem 5 If the yield distribution is a stochastically proportional power distribution, i.e., if for any i, Q, $F(i, Q, y) = (\frac{y}{Q})^{r_i}$, for some $r_i \geq 1$, then the optimal inflation rate $Q^*(i, D, N)/D$ decreases in D.

We also show that the optimal lot size monotonically increases in the unit waiting cost, the arrival rate, and the production time.

Theorem 6 When a = 0, the optimal production quantity Q^* increases 1) if material cost c_M decreases, 2) if waiting cost c_w increases, 3) if λ increases, 4) if the production time T stochastically increases, or 5) if yield Y stochastically decreases and $F'_Q(i, Q, y)$ also decreases.

Theorem 6 shows that the behavior of the optimal lot sizes is intuitive. Higher material costs result in lower production quantities (i.e., potentially more runs), be-

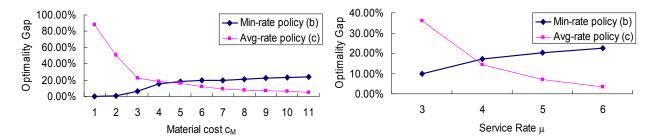


Figure 3.1. Cost difference of heuristic policies and optimal policy with different material costs.

Figure 3.2. Cost difference of heuristic policies and optimal policy with different service rates.

cause the firm does not want to waste valuable material. Higher delay costs have the opposite effect. More demand or slower production completion also result in larger batch sizes.

We have proved the policy structures for the optimal dynamic lot sizing decisions. In reality, firms often use some simple strategies to control lot sizes. A common strategy is to use the same inflation rate for any demand or queue length. We conduct a numerical study to compare the profits under these strategies with the optimal policy. We consider the following three policies: (a) The optimal policy, (b) fixed inflation by minimum yield, i.e., Q = D/a, and (c) fixed inflation by average yield, i.e., Q = D/EX.

For all the figures, we use $\lambda = 1$, a = 1, $c_W = 150$, $c_R = 0$, $\bar{D} = 50$. We use $\mu = 4$, r = 1, $c_M = 4$ if these parameters are not changed in the graph. We set a maximum queue length as 20, and use a step size 0.1. The program converges if $\max_{D,N} V^{k+1}((D,N)-V^k(D,N)) - \min_{D,N} (V^{k+1}((D,N)-V^k(D,N)))$ is smaller than 0.1.

We compare the costs resulting from these three policies. Figure 3.1, 3.2, and 3.3 show the cost difference of policy (b) and (c) as a percentage of the optimal average

cost with different parameters.

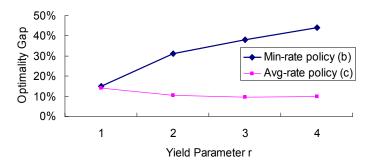


Figure 3.3. Cost difference of heuristic policies and optimal policy with different yield distributions.

Figure 3.1 shows the relative costs of the heuristic policies as the material cost changes. Clearly, policy (b) performs better with smaller material cost. This is intuitive since if material is cheap, a larger lot size can be used and thus Q = D/a is closer to optimal. policy (c) performs in the opposite way. When material is more expensive, policy (c) performs better. The other two graphs show a similar pattern when the production rate changes or when yield distribution changes.

In summary, policy (b) is the most conservative policy, and it works the best when material is cheap, when production is slow, or when yield is poor. Policy (c) is an intermediate solution and it works better when material is more expensive, when production is slower, or when yield is better. Both policies can perform significantly badly compared to the optimal policy, with 25 - 50% higher costs easily observable in many cases. This indicates that taking the externality of delays caused by the lot sizing decisions into account has a large impact on costs.

3.3.2 Overtime or Subcontracting Decisions

In our setting, the firm has to rerun the production when any demand remains to be satisfied. In reality, firms sometimes have an option to schedule excess capacity (in the form of overtime or subcontracting for capacity) to finish the current order without delaying other orders. Overtime production often incurs higher resource costs,
therefore the firm needs to make a decision as to when to run overtime production.
Similarly, the use of subcontracting plants to outsource some of the active ingredient
production is fairly common in the pharmaceutical industry, but the subcontracting
plants may charge a higher price than using internal production. If the firm has the
option to subcontract, it also needs to make the optimal subcontracting decisions
depending on the remaining demand and system utilization.

As these two decisions share some common elements, we consider the two decisions in the same section. In the overtime situation, we assume that after a production run for a particular order with lot size Q, if the output is not sufficient to satisfy the order, the firm may either decide to rerun the same product, or to start producing the next order and schedule overtime for the current order without delaying other orders. (This is fairly common in the companies we have worked with). We make an assumption that the firm has infinite overtime capacity. The material cost for overtime production is the same, while the resource cost is higher, $c_O > c_R$ per unit of time. During the overtime production, the firm also needs to make optimal lot sizing decisions. Therefore, the subproblem during overtime production is equivalent

to Grosfeld-Nir and Gerchak (1990).

$$V(i, D, N) + g/\mu = \min_{Q \in \Omega(D)} \begin{cases} N * W + W_{new} + c_R/\mu + c_M Q + \sum_{N'=0}^{\infty} \text{Prob}(N') \\ [\bar{F}(i, Q, D)V(1, \bar{D}, N - 1 + N') \\ + \int_{aQ}^{D} \min \left[V(i+1, D-y, N+N'), \\ C_O(i+1, D-y) + V(1, \bar{D}, N+N'-1) \right] dF(i, Q, y) \end{bmatrix}$$

$$C_O(i, D) = \min_{Q \in \Omega(D)} \begin{cases} W + c_O/\mu + c_M Q + \int_{aQ}^{D} C_O(i+1, D-y) dF(i, Q, y) \end{cases}$$

In the equations above, during the regular production, if the order is not satisfied after a production run, the firm can choose to continue production for this order, or to work on the next order and do overtime production for the current order. $C_O(i, D)$ is the total production cost if overtime production is used for the *ith* trial of a particular order and if the unsatisfied demand for that order is D. The overtime production cost is similar to the regular production cost except that delay costs for other orders are not incurred.

In the case of subcontracting, we assume that when the firm is producing for a particular order, the firm may decide a lot size Q and start producing for this order, or the firm may decide to switch to the next order, and subcontract this order to a subcontractor. The subcontractor charges a price $C_S(D)$ which is assumed to be

increasing in D. The optimality equations for $N \geq 1$ then become:

$$V(i, D, N) + g/\mu = \min_{Q \in \Omega(D)} \begin{cases} N * W + W_{new} + c_R/\mu + c_M Q + \sum_{N'=0}^{\infty} \text{Prob}(N') \\ [\bar{F}(i, Q, D)V(1, \bar{D}, N - 1 + N') \\ + \int_{aQ}^{D} \min \left[V(i+1, D-y, N+N'), \\ C_S(D-y) + V(1, \bar{D}, N+N'-1) \right] dF(i, Q, y) \end{bmatrix} \}$$

For both situations, we can apply similar techniques and define $H(i, D, N) = V(i, D, N) - V(1, \bar{D}, N - 1)$. Then

$$H(i, D, N) + g/\mu = \min_{Q \in \Omega(D)} \begin{cases} N * W + W_{new} + c_M Q + \sum_{N'=0}^{\infty} \text{Prob}(N') \\ \sum_{n=N}^{N+N'-1} H(1, \bar{D}, n) + \int_{aQ}^{D} \min[H(i+1, D-y, N+N'), C(D-y)] dF(i, Q, y)] \end{cases}$$

where $C(D-y)=C_O(i+1,D-y)$ in the situation of overtime production, and $C(D-y)=C_S(D-y)$ in the situation of subcontracting.

For both situations, we can show that there exists a threshold in queue length for each remaining demand level, such that the firm starts to work on the next order if and only if the current queue length exceeds that threshold. The idea is intuitive: if the number of waiting orders exceeds a limit, which means the system is heavily utilized and cost of delay is significant, the firm should consider overtime production or subcontracting for the current order.

Theorem 7 If the conditions in Theorem 4 are satisfied, then there exists a unique optimal lot size $Q^*(D, N)$ increasing in D and N. After the production run if \tilde{D} unit of demand still needs to be satisfied for the current order, and if there are \tilde{N} orders

in total, then there exists an $\bar{N}(\tilde{D})$, such that the firm runs overtime production (or subcontracts production) if and only if the queue length $\tilde{N} \geq \bar{N}(\tilde{D})$; otherwise the firm starts a new regular production run for the current order.

Next we consider the threshold structure for the remaining demand. In the situation of subcontracting, if the subcontracting cost consists of a fixed cost A_S and a linear variable cost c_S , and the yield follows stochastically proportional power distribution, we can show that the optimal policy has a threshold structure for the remaining demand level depending on the fixed cost. If the fixed cost is relatively small, then there is one threshold for the remaining demand below which the firm would subcontract the production. If the remaining demand of the customer currently in service is below the threshold, then it is not worthwhile to restart the production run and make other orders wait, thus it is optimal to subcontract. Moreover, the threshold level (the remaining quantity below which the firm subcontracts production) is increasing in the queue length. However, if the fixed cost is relatively high, then there two thresholds for the remaining demand level. If the remaining demand is too low, the savings on subcontracting cannot offset the high fixed cost, thus it is optimal to continue the production in house. If the remaining demand is at an intermediate level, it is worthwhile to pay the fixed cost and subcontract the production, in order to accelerate the production and save the delay cost on other orders. The results are summarized in the following theorem.

Theorem 8 If the subcontracting cost is linear in the remaining demand, i.e., $C_S(D) = A_S + c_S D$, where A_S is a constant, and if the yield is stochastically proportional power

distribution, then for each N, there exists a constant \bar{A} , such that

- 1. If the fixed cost $A_S \leq \bar{A}$, then there exists a threshold $\bar{D}_i(N)$, such that the firm subcontracts production if and only if the remaining demand for the current order $D \leq \bar{D}_i^*(N)$, otherwise the firm starts a new regular production run for the current order. Furthermore, the threshold $D_i^*(N)$ is increasing in N.
- 2. If the fixed cost A_S > Ā, then there exist two thresholds <u>D</u>_i(N) ≤ D̄_i(N), such that the firm subcontracts production if and only if the remaining demand for the current order <u>D</u>_i(N) ≤ D ≤ D̄_i(N). Furthermore, the threshold <u>D</u>_i(N) decreases in N and D̄_i(N) increases in N.

In the situation of overtime production, the firm still makes optimal lot sizing decisions during the overtime productions. Depending on the overtime cost, we can show that the overtime decisions also have a threshold structure.

Theorem 9 Assume that the conditions in Theorem 4 are satisfied. For each queue length N, there exist two thresholds $\underline{c}_i(N) \leq \overline{c}_i(N)$, such that

- 1. If the overtime resource cost $c_O \leq \underline{c}_i(N)$, then the firm runs overtime production for the current order for any remaining demand on the current lot.
- 2. If the overtime resource cost $c_O > \bar{c}_i(N)$, then the firm continues production for the current order for any remaining demand on the current lot.
- 3. If the overtime resource cost $\underline{c}_i(N) < c_O \leq \overline{c}_i(N)$, then there exists a threshold \overline{D} such that the firm runs overtime production if and only if the remaining demand on the current lot $D < \overline{D}$.

The theorem above describes the firm's optimal overtime production policy. If the

overtime resource cost is too high or too low, continue production or overtime production dominates for any demand. If the overtime resource cost is in the intermediate range, then the firm only does overtime production if the remaining demand is small enough.

3.4 Summary

In this paper we study a lot-sizing problem in pharmaceutical industry with considerable yield losses, high material costs, high delay costs and limited capacity. We build a dynamic programming model to show how the optimal lot sizes should be adjusted based on the actual demand and queue length.

Our model builds a bridge between the classic queueing control models and MLPO random yield models, and provides solutions under joint consideration. We also adopt a more general yield distribution than the existing random yield literature. Our optimality results indicate that the simple inflation rate policies currently used in many pharmaceutical companies are very far from optimal. In certain situations the cost can be significantly decreased if the lot sizes can be dynamically adjusted as in the optimal policy.

Clearly, our model captures the two critical elements that should be taken into account for lot-sizing decisions: material cost and delay cost. Other factors may be interesting for further exploration. Future directions may include analysis of a process with multiple production stages, heterogeneous demand with priority service, learning effects that improve yield rate over time.

CHAPTER 4

Make-or-Buy Decisions When Cost Reduction Requires Minimum Production Quantity

4.1 Introduction

In recent years, increasingly many US companies seek lower production costs and outsource all or a portion of their production to developing countries, such as Mexico and Southeast Asia, mostly due to lower labor and material costs in these regions. Thus, in a variety of industries where low cost plays an important role in gaining competitive advantage, the potential for outsourcing has made the Make-or-Buy decision extremely visible and strategically important.

When making outsourcing decisions, firms might, and often do, limit themselves to the comparison of the current outsourcing and internal production costs, or in case of more sophisticated firms, they may take into consideration likely trends in the currency exchange rates and in labor rates per hour. In reality, however, costs do not remain the same over time even if these global conditions are fairly stable. Clearly, cost reductions are widely observed in various industries and sectors, such as high-tech or sectors with new products and technologies. Myopic comparisons often ignore the fact that various firms have different potentials to reduce costs, and that manufacturers have no direct control over Supplier's production cost or cost-reducing

activities. In this paper, we model two elements related to cost reduction activities, the first one being a minimum production quantity necessary for a firm to engage in an active learning, and the second one being the costly effort necessary in order to learn and reduce cost. In practice, these two elements are essential for allowing cost reductions and, sometimes, they are referred to as "learning-by-doing".

To model such a "learning-by-doing" environment, we consider a Manufacturer selling a single product with fixed demand and fixed retail price in a two-period model. Manufacturer can produce the product himself as well as outsource any portion of production to a Supplier. Each of them can invest effort in cost-reducing activities, but cost reduction can take effect only if production quantity exceeds a certain minimum level. Based on wholesale price offered by supplier, manufacturer decides whether to produce or to outsource. Clearly, the strategies of the supplier and manufacturer as well as the optimal production and outsourcing quantities are influenced by both starting production costs and learning abilities.

Outsourcing in practice means giving up some of the decision rights. Thus, a natural question is about the relative benefits of centralized versus decentralized solution, In the considered "learning-by-doing" environment, owning a facility (a centralized system) means that manufacturer can capture all benefits from cost reductions, as opposed to two independent firms acting in their own interest. We compare the decisions in these two ownership environments.

Our work is closely related to Gray et al (2007), who also study cost reduction through learning-by-doing, in a two-period model. Our model differs from theirs in a few aspects, such as modeling research questions, and the resulting insights. First, our model focuses on an settings with low price sensitivity and includes realistic features of "earning by doing" environment. Price and demand are fixed in our model as opposed to price being a linear function of demand. We require that production has to exceed a certain minimum quantity and that costly effort needs to be invested, which are not considered in Greys et al. Instead of function of production, the learning in our model is a function of effort (usually learning and cost reductions do not take place automatically in practice). We also incorporate a realistic feature of uncertainty in the cost-reducing effect. Some of the observations in Gray et al (2007) and in our model are the same. For example, Manufacturer may choose to produce internally, even if he is at a cost disadvantage, partial outsourcing is possible, and Supplier may charge below-cost wholesale prices. It is easy, however, to see that without price sensitivity, the many of the results described Greys et al would not anymore hold.

In terms of research questions, Gray et al (2007) focuses on the influence of power structure between Manufacture and Supplier, while we concentrate on identifying when and why outsourcing in general, and partial outsourcing specifically, is more likely. In settings where Supplier is Stackelberg leader, we show that the results depend on the cost difference of Manufacturer and Supplier and also on the learning ability and minimum production quantity. Also, partial outsourcing is more likely when the initial cost difference is relatively small, when Manufacturer's learning is more uncertain, and when the minimum production quantity required for Manufacturer to be able to learn is larger.

Finally, some of our additional insights follow from a three-period model which shows that Manufacturer can benefit from Supplier's learning behavior and, consequently, Manufacturer may consider investing some effort to subsidize the Supplier's cost of effort. However, sometimes this benefit disappears. If Supplier knows that Manufacturer would be willing to help, Supplier will take advantage of it and charge a higher wholesale price, forcing Manufacturer to indeed subsidize his cost. In this situation Manufacturer may end up worse off compared to not having this option.

The paper is organized as follows: Section 2 reviews the previous work on subcontracting problems; Section 3 describes the basic model and analyze the problem; Section 4 extends the model to three periods and considers the option of effort sharing. Section 5 summarizes the work.

4.2 Literature Review

The early research of outsourcing decisions considers only one decision maker and outsourcing is used as a more expensive outside production option to be used when internal production is not sufficient. Lee and Zipkin (1989) study the multi-period inventory problem assuming reliable outsourcing capacity and a constant cost. They characterize the structure of inventory policy and provide a DP algorithm to compute the optimal solution. Atamturk and Hochbaum (2001) consider a problem of acquiring a fixed capacity to meet non-stationary deterministic demand over a multi-period horizon. Production and outsourcing decisions are made after capacity is acquired. They construct a Linear Programming model and provide an algorithm to find the optimal decisions.

Later work takes supplier's reaction into account and focuses on the contract design and supply chain coordination between manufacturer and supplier. Kamien and Li (1990) study a multi-period production planning problem and define feasible, transaction-feasible, efficient, and dominant contracts. They show that coordination contracts, which achieve social optimum, are always feasible and always dominate other contracts. An example with quadratic production cost is used to demonstrate that subcontracting is a useful tool for production smoothing. Van Mieghem (1999) adds capacity investment cost to cost factors and studies four types of outsourcing contracts. He shows that an incomplete contract, where production quantities are jointly decided in order to maximize total profit, can coordinate production decisions, but only state-dependent contract, where price or quantity is agreed up-front as a function of capacities and demands, can coordinate both capacity and production decisions.

In both of the above papers on contract design, costs and demands are exogenously given. Plambeck and Taylor (2001) propose a model in which firms can use effort to stimulate demands and also the firms can outsource their production to a joint supplier. They show that when benefits of capacity pooling are significant and profits are retained (or mostly retained), manufacturers may be better off by subcontracting their production. Lewis and Sappington (1991) propose a model close to ours, where costs can be reduced by technological improvements. Both technology and costly effort can reduce production cost, while technology can also increase the efficiency of effort and reduce the cost of the effort. They show that technology improvement

leads to a more frequent use of internal production. However, both manufacturer and society may be worse with more frequent technology changes. Their model is a one-period game focusing primarily on the effect of asymmetric information. In our paper, we have assumed that demand is exogenous and that production cost can be reduced over time due to learning by doing. We are interested in understanding how outsourcing decisions are made and how production and learning decisions are related to the outsourcing option.

Another related question is often labeled as "make or buy" decision and a subset of papers in operations and economics literature asks when it is in firms' interest to both make and buy. Kamien, Li, and Samet (1989) study the effect of adding subcontracting option to Bertrand game. Because of strictly convex production costs, subcontracting is beneficial, as it allows for reallocating production and thus a decrease in the marginal cost. Spiegel (1993) studies subcontracting within Cournot framework. Costs include a convex upstream cost (manufacturing) and a linear downstream cost (marketing and sale). Two types of subcontracting agreements are allowed: ex ante agreements signed before the firms compete in the market, and ex post agreements signed after competition takes place. He shows that subcontracting can allocate production more efficiently under asymmetric costs. In both papers, concurrent sourcing is due to the diseconomy of scale resulting from convex production costs. None of the models above considers the interaction between production process and cost reduction. In our model firms can reduce costs with effort, but only if a certain production requirement is satisfied. We show that in absence of the diseconomy of scale dual

production can still appear.

We described the work by Greys et al in the previous section. Our results are complementary in the sense that they apply in price insensitive environment, and include the realistic features of cost of effort, minima over which learning may take place, and uncertainty in the effects of effort. We also explain how the multi-period behavior may differ from a stylized two-period setting. At the end of the paper, we also consider the possibility for Manufacturer to share learning effort of Supplier. A related work is Krishnan et al (2004), which considers the impact of effort sharing between a manufacturer and a retailer, and compares different contract mechanisms.

Other research related to our work appears in corporate strategy framework. Parmigiani (2002) claims that firms may choose concurrent sourcing if (a) no clear price advantage exists for either internal or external suppliers; (b) quality monitoring is difficult; (c) disputes are not likely; (d) both autonomous and coordinative incentives exist. Our results confirm not only that (a) is a possible condition for dual sourcing, but provide a model that explains potential reasons for this outcome.

4.3 Model

We consider a Manufacturer selling a single product in two periods (i = 1, 2). In each period, prices are exogenously given as P_1 and P_2 , and demand in Period i is a random variable Q_i with expectation \bar{Q}_i . Demand has to be satisfied in full.

Manufacturer is considering to outsource some production to a Supplier. Both Manufacturer and Supplier's production costs are linear. Without other considerations, Manufacturer would outsource all production if Supplier quotes a wholesale price lower than Manufacturer's cost. However, both firms have the ability of learning and consequently reducing their production costs if they have the chance to conduct sufficient amount of production. These cost-reducing activities also require some costly effort, and the cost reduction takes place at the beginning of the next period.

Specifically, if firm I's unit production cost in Period 1 is c_1^I , then if firm I produces q_1^I less than a certain minimum quantity q_{min}^I , his cost stays the same in the next period. If he produces more than q_{min}^I , then he can invest effort ϵ^I , and his cost will be reduced by $g^I(\epsilon^I, \delta^I)$ in the next period, where δ^I is a random disturbance. Therefore,

$$c_2^I = \begin{cases} c_1^I & \text{if } q_1^I < q_{min}^I, \\ c_1^I - g^I(\epsilon^I, \delta^I) & \text{else.} \end{cases}$$

Let $\bar{g}^I(\epsilon)$ be the expectation of $g^I(\epsilon, \delta)$.

In each period i = 1, 2, events happen in the following sequence:

- 1. Supplier quotes linear wholesale price ω_i .
- 2. Demand realizes, Manufacturer decides in-house production quantity q_i^M and outsourcing quantity q_i^S , so that $q_i^M + q_i^S = Q_i$.
- 3. Both Manufacturer and Supplier determine effort level ϵ_i^M and ϵ_i^S .
- 4. Production takes place, and costs are reduced at the beginning of the next period if production quantities exceed minimum amount.

We make the following assumptions:

- 1. $P_i \ge c_1^M$ and $P_i \ge c_1^S$, i = 1, 2. We assume that production makes nonnegative profit at both firms in every period.
- 2. The two random disturbance δ^M and δ^S are independent.

- 3. $\bar{g}^M(\epsilon)$ is concave in ϵ , so as Manufacturer invests more effort, the learning effect diminishes.
- 4. Without loss of generality, we assume the unit monetary cost of effort is 1.
- 5. In each period, the realized demand is large enough for two firms to possibly learn, i.e., $Q_i \geq q_{min}^M + q_{min}^S$ for any realized Q_i , i = 1, 2.
- 6. If Manufacturer is indifferent between producing q_1 and q_2 in house, he always chooses to produce the smaller quantity.

The last assumption restricts the game outcome on one path. It is reasonable since Supplier is the Stackleburg leader, he can always reduce wholesale price by a tiny amount, then Manufacturer will choose to produce the smaller quantity.

Both firms maximize expected total profit for two periods,

$$\pi^{M} = E_{Q_{1},Q_{2},\delta^{M},\delta^{S}} \sum_{i=1}^{2} (P_{i} - c_{i}^{M}) q_{i}^{M} + (P_{i} - \omega_{i}) q_{i}^{S} - \epsilon_{1}^{M}$$

$$\pi^{S} = E_{Q_{1},Q_{2},\delta^{M},\delta^{S}} \sum_{i=1}^{2} (P_{i} - c_{i}^{S}) q_{i}^{S} - \epsilon_{1}^{S}$$

4.3.1 Centralized Problem

As a benchmark we first consider the case where a single firm owns both facilities M and S, and considers the problem of allocating demand in each period in order to maximize expected total profit. Then similarly we have the two-threshold properties of quantity allocation:

Lemma 11 There exist two thresholds $x_1 \le 0 \le x_2$ such that the production quantity at M is:

$$q_1^M = Q_1 \text{ if } c_1^M - c_1^S \le x_1, \ q_1^M = Q_1 - q_{min}^S \text{ if } x_1 \le c_1^M - c_1^S \le 0, \ q_1^M = q_{min}^M \text{ if } 0 \le c_1^M - c_1^S \le x_2, \ and \ q_1^M = 0 \text{ if } x_2 \le c_1^M - c_1^S.$$

If both learning functions are deterministic, then one of the two thresholds x_1 and x_2 must be zero, since we know exactly which facility can have lower cost in the second period and thus will not invest effort in the other facility.

4.3.2 Decentralized Game

Next we analyze the game equilibrium when Manufacturer and Supplier do not belong to the same company. We analyze this game in a backward fashion. First we examine the pricing and production strategies in the second period, given costs c_2^M and c_2^S . Since no learning effects need to be considered in the second period, the results are straightforward and are summarized in the following lemma.

Lemma 12 If the costs at the beginning of the second period are c_2^M and c_2^S , then

- (a) If $c_2^M \geq c_2^S$, then Supplier offers wholesale price $\omega_2 = c_2^M$ and produces all $q_2^S = Q_2$.
- (b) If $c_2^M < c_2^S$, then Manufacturer produces all $q_2^M = Q_2$.
- (c) Manufacturer makes expected profit $R_2^M = (P_2 c_2^M)\bar{Q}_2$, and Supplier makes expected profit $R_2^S = (c_2^M c_2^S)^+\bar{Q}_2$, where x^+ represents the positive part of x.

We observe that Manufacturer makes the same profit no matter where the production takes place, and also no matter how cheaply Supplier can produce. This indicates that in the first period Manufacturer does not have an incentive to let Supplier improve his cost, on the other hand, he has a strong incentive to reduce his own cost.

Supplier only makes positive profit if his cost is lower than Manufacturer's. Furthermore, the difference between Manufacturer's cost and his own cost represents the profit margin he is able to obtain. This implies that in the first period Supplier also

has an incentive to reduce his own cost, but at the same time he would try to prevent Manufacturer from improving.

We examine the effort investment after production and outsourcing decisions are made in the first period. Apparently no effort is invested if the production quantity is smaller than the minimum requirement. If it exceeds the minimum quantity, then Manufacturer invests a certain level of effort ϵ^{M^e} regardless of initial costs and Supplier's improvement, while Supplier's effort level is a function of initial costs c_1^M and c_1^S , and also depends on whether Manufacturer improves or not. The results are summarized in the following lemma.

Lemma 13 Given wholesale price ω_1 and quantity allocation q_1^M , q_1^S ,

- (a) Manufacturer invests a unique effort ϵ^{M^e} if $q_1^M \geq q_{min}^M$, where $\epsilon^{M^e} = \arg\max(-\epsilon + \bar{g}^M(\epsilon)\bar{Q}_2)$;
- (b) Supplier invests effort ϵ^{S^e} if $q_1^S \geq q_{min}^S$, where

$$\epsilon^{S^e} = \begin{cases} \arg \max\{-\epsilon^S + E_{\delta^S}(c_1^M - c_1^S + g^S(\epsilon^S, \delta^S))^+ \bar{Q}_2\} & \text{if } q_1^M < q_{min}^M, \\ \arg \max\{-\epsilon^S + E_{\delta^M, \delta^S}(c_1^M - c_1^S + g^S(\epsilon^S, \delta^S)) - g^M(\epsilon^{M^e}, \delta^M))^+ \bar{Q}_2\} & \text{else.} \end{cases}$$

As an intermediate result, we find that no matter whether outsourcing or producing internally, Manufacturer always invests in the same level of effort. On the other hand, Supplier will not invest in any learning effort if he expects that he is not going to obtain any production in the second period. If Supplier will produce in the second period, he will invest in some learning effort in the first period, the amount of which depends on the quantity he will produce. The larger the outsourcing quantity in the second period, the more effort Supplier invests in.

Since Manufacturer does not benefit from Supplier's learning, we find that Manufacturer does not have an incentive to help Supplier learn. As we will notice later, this conclusion is only suitable for short planning horizon (two periods in this case). When the problem is extended into three periods, Manufacturer may be willing to help Supplier learn from production.

We redefine some notations to simplify analysis below. Let $G^M(\bar{Q}_2) = -\epsilon + \bar{g}^M(\epsilon)\bar{Q}_2$, then G^M is the extra profit Manufacturer gains from learning. Manufacturer benefits more from learning if there is a larger market in the second period. Also let

$$\begin{split} R^{S^e} &= \max_{\epsilon^S} \{ -\epsilon^S + E_{\delta^M, \delta^S} (c_1^M - c_1^S + g^S(\epsilon^S, \delta^S))^+ \bar{Q}_2 \} \\ R^{MS^e} &= \max_{\epsilon^S} \{ -\epsilon^S + E_{\delta^M, \delta^S} (c_1^M - c_1^S + g^S(\epsilon^S, \delta^S) - g^M(\epsilon^{M^e}, \delta^M))^+ \bar{Q}_2 \} \\ R^{M^e} &= E_{\delta^M} (c_1^M - c_1^S - g^M(\epsilon^{M^e}, \delta^M))^+ \bar{Q}_2 \end{split}$$

Then R^{S^e} , R^{MS^e} , R^{M^e} represent Supplier's maximum second-period profit subtracted by effort cost when (a) only Supplier can learn, (b) both firms can learn, and (c) only Manufacturer can learn, respectively.

Lemma 14 Manufacturer produces more in house as wholesale price ω_1 increases. More specifically,

$$q_1^M = \begin{cases} 0 & if \, \omega_1 \leq \bar{\omega} = c_1^M - \frac{G^M(\bar{Q}_2)}{q_{min}^M}, \\ q_{min}^M & if \, \bar{\omega} < \omega_1 \leq c_1^M, \\ Q_1 & else. \end{cases}$$

Notice that other than full internal production $(q_1^M = Q_1)$ and full outsourcing $(q_1^M = 0)$, there is a situation when partial outsourcing $(q_1^M = q_{min}^M)$ takes place. The lemma indicates that Manufacturer allocates production quantity according to the

wholesale price Supplier offers. Using his own current production cost as a benchmark, if the wholesale price is higher than his cost, Manufacturer produces all the demand internally. If the wholesale price is much lower than his own cost, Manufacturer outsources all the production. However, if the wholesale price is lower than Manufacturer's cost, but not significantly lower, Manufacturer chooses to outsource most of the production, but still keep a minimum level of internal production, in order to learn and reduce cost in the next period. This may happen no matter outsourcing takes place or not in the next period, because Manufacturer's profit only depends on his own cost, so he always has an incentive to learn and reduce cost.

Knowing Manufacturer's response to any wholesale price, Supplier chooses one to maximize his own expected profit. We observe that $\bar{\omega}$ is independent of demand in the first period Q_1 , therefore Manufacturer's quantity allocation is independent of demand realization. Since any wholesale price $\omega_1 \in [0, \bar{\omega}]$ results in the same quantity allocation and effort investment, and similarly for any $\omega_1 \in (\bar{\omega}, c_1^M]$, Supplier quotes the highest wholesale price by which he can obtain a certain amount of production. So Supplier considers three options and chooses the one with the highest expected profit: (a) Offer deep discount $\omega_1 = \bar{\omega}$ and produce all demand $q_1^S = Q_1$ for any realized demand Q_1 ; (b) Match Manufacturer's cost $\omega_1 = c_1^M$ and produce partial demand $q_1^S = Q_1 - q_{min}^M$; and (c) Offer $\omega_1 > c_1^M$ and produce nothing $q_1^S = 0$.

When considering these three options, Supplier makes tradeoff between price and quantity, and also between current period profit and future gains. Myopically, a larger quantity can only be obtained through a deep discount; For future considerations,

Supplier benefits from large quantity in two ways: having enough production provides

Supplier the opportunity of learning, and having extremely large production prevents

Manufacturer from learning. The following table lists the pros and cons for each
option.

	$(a)q_1^S = Q_1$	(b) $q_1^S = Q_1 - q_{min}^M$	(c) $q_1^S = 0$
Wholesale price in Period 1	Small	Medium	Large
Quantity in Period 1	Large	Medium	Zero
Wholesale price in Period 2	Large	Small	Small
Cost reduction in Period 2	Large	Large	Small

We show in the following theorem that the choice of Supplier is monotone in initial cost difference. So if Supplier's initial cost c_1^S is significantly lower than Manufacturer's, he will produce all, if Supplier's initial cost is significantly higher than Manufacturer's, he will produce nothing and if their costs are close, he will produce partial demand.

Theorem 10 There exists two thresholds $x_1 \leq x_2$ such that

$$\begin{cases} q_1^M = Q_1, & \omega_1 > c_1^M, & \text{if } c_1^M - c_1^S < x_1 \\ q_1^M = q_{min}^M, & \omega_1 = c_1^M, & \text{if } x_1 \le c_1^M - c_1^S < x_2 \\ q_1^M = 0, & \omega_1 = \bar{\omega}, & \text{else} \end{cases}$$

Furthermore, $x_1 < 0$, $x_2 \ge 0$ if the following condition holds:

$$\bar{Q}_1 \frac{G^M(\bar{Q}_2)}{q_{min}^M} \ge \max_{\epsilon^S} (-\epsilon^S + \bar{g}^S(\epsilon^S)\bar{Q}_2) - \max_{\epsilon^S} (-\epsilon^S + E_{\delta^M,\delta^S}(g^S(\epsilon^S,\delta^S) - g^M(\epsilon^{M^e},\delta^M))^+ \bar{Q}_2)$$

 $x_1 < 0$ indicates that as long as Supplier has a lower cost, he will outsource at least part of the production. This is intuitive since Manufacturer can still learn by doing so, and also will not be hurt by Supplier's learning.

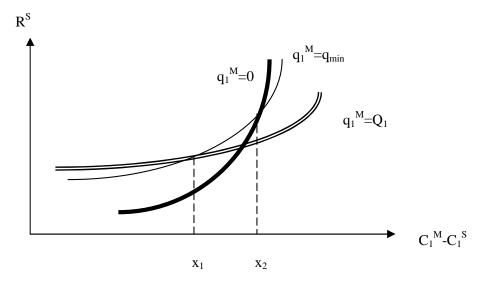


Figure 4.1. Supplier's profits under three options.

 x_2 can be positive or negative, which implies that even if Supplier has a lower cost than Manufacturer, sometimes he cannot produce full demand, because Manufacturer always wants to reserve some production in house for the purpose of learning. This happens when the condition in Theorem 10 holds.

Whether the condition holds depends on many factors. Specifically speaking, it holds when (a) average demand in Period 1 is large enough, or (b) the minimum requirement for Manufacturer q_{min}^M is small enough, or (c) the extra profit Manufacturer gets from learning $G^M(\bar{Q}_2)$ is large enough, or (d) the optimal cost reduction for Manufacturer $g^M(\epsilon^{M^e}, \delta^M)$ is small enough.

We make a couple of observations related to this outcome. First, if both firms are symmetric, i.e., they have the same initial cost $c_1^M = c_1^S$, and the same learning functions $g^M = g^S$, then Manufacturer always produces minimum quantity q_{min}^M . Secondly, if one firm produces everything in the first period, that firm will produce everything in the second period, as well. The reason is that the other firm did not

have the chance to learn, so the cost disadvantage remains there or becomes even larger in the second period.

By comparing the centralized and decentralized problems, we found that in some situations Manufacturer always produce more in the decentralized case. If we label x_2^d and x_2^c as the thresholds for decentralized and centralized cases respectively, then we have the following lemma.

Lemma 15 If
$$g^M > g^S$$
, then $x_2^d > x_2^c$.

Thus, if Manufacturer has a stronger learning ability, then Manufacturer is more reluctant to outsource (even partially) in the decentralized case then in the centralized case.

4.3.3 Effect of Parameter Changes

Effect of Variance: We also examine the effect of variance of learning on the outcome. We find that both firms' strategies only depend on the expectation of decrease in cost, independent of variance.

For a given wholesale price, variance of Manufacturer's cost-reducing function $g^M(\epsilon, \delta)$ also does not influence Manufacturer's quantity allocation and both firms' effort decisions. Since Manufacturer's second-period profit only depends on his own cost, for a fixed wholesale price, only the expectation of future improvement matters, thus the outsourcing and effort decisions remain the same.

However, the variance does have an impact on Supplier's optimal wholesale price decision, and thus on the final outcome. Supplier compares three options. For producing 0 or Q_1 , the variance of g^M does not influence Supplier's expected profit.

However, for producing $Q_1 - q_{min}^M$, Supplier's profit is larger if Manufacturer's learning is more uncertain. This is because Supplier only makes positive profit in the second period if his cost is lower than Manufacturer's. With a larger variance of g^M , there is a higher chance for this to happen. Therefore with a larger variance of g^M , x_1 decreases, x_2 increases, and Manufacturer is more likely to produce minimum quantity q_{min}^M .

Variance of Supplier's cost-reducing function $g^s(\epsilon, \delta)$ has a similar but slightly different effect. A larger variance also increases the probability that Supplier will have a lower cost than Manufacturer in the second period, but this makes both options of producing partially and producing everything at Supplier's plant more promising than before. In other words, **As cost-reduction of Supplier becomes** more variable, x_1 decreases, and outsourcing (at least partially) tends to be more likely. Once the wholesale price is announced, it does not influence any further decisions.

Profits change with costs: As shown in Figure 4.2, for a given c_1^M , Supplier's profit is a decreasing convex function of c_1^S . However, Manufacturer's profit is not monotone in c_1^M . When Manufacturer produces Q_1 or q_{min}^M , his profit is linearly decreasing in his cost c_1^M , regardless of Supplier's cost. Similarly, when Manufacturer outsources all demand, his profit is again linearly decreasing in his cost and independent of Supplier's cost. However, if a tiny increase in Manufacturer's cost or a tiny decrease in Supplier's cost happens, so that Manufacturer starts to outsource all rather than produce some amount in house, then Manufacturer's profit dramatically increases.

The amount of increase in profit is $(\frac{\bar{Q}_1}{q_{min}^M} - 1)G^M(\bar{Q}_2)$.

Supplier's profit is influenced by Manufacturer's initial cost. As c_1^M increases, Supplier's profit increases.

Manufacturer's profit is only slightly influenced by Supplier's initial cost. By "slightly" we mean that as c_1^S increases, Manufacturer's profit does not change unless Manufacturer switches to outsource some production from producing everything internally, in which case Manufacturer's profit drops.

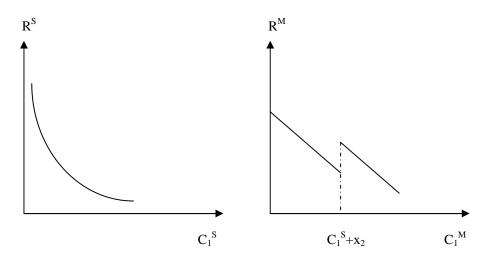


Figure 4.2. Manufacturer and Supplier's profits change with costs.

Effect of demand in the first period Q_1 : We examine the effect of Q_1 on x_1 and x_2 separately. x_1 is the cost difference of Manufacturer and Supplier at which Supplier is indifferent between producing 0 and $Q_1 - q_{min}^M$. With a larger demand in the first period, Supplier still makes zero profit if he produces nothing, For the option of producing partially, Supplier offers the same wholesale price $\omega_1 = c_1^M$, invests the same level of effort, and makes the same profit in the second period, but produces more in the first period. As we have known that $x_1 < 0$, so at $c_1^M - c_1^S = x_1$ Supplier loses money by producing in period one. He loses more if he produces more in the first

period. Therefore as the demand in the first period increases, x_1 increases, and it is more likely to produce everything in Manufacturer's facility than producing at both places.

 x_2 is the cost difference between Manufacturer and Supplier, at which Supplier is indifferent between producing partially versus producing everything. Again, for a fixed wholesale price, Q_1 has no impact on the subsequent decisions. Q_1 only changes the amount of production in the first period. For Supplier, the profit margin of producing partially is greater than that of producing everything, so at $c_1^M - c_1^S = x_2$, as Q_1 increases, Supplier's profit increases more (or decreases less) if he produces partial production. In other words, as Q_1 increases, x_2 increases, and it is more likely to outsource partial production than to outsource all.

In summary, Manufacturer tends to produce more when demand in the first period is larger.

Effect of demand in the second period Q_2 : The effect of Q_2 on the outcome of the game is not immediately clear. In general x_1 and x_2 can change in both directions.

For x_1 , we compare the option of outsourcing partially and not outsourcing at all. Supplier always makes zero profit for the latter option, while for the former option Supplier is more eager to learn and hopefully can make more profit in the second period if demand is larger in the second period. However Manufacturer thinks in the same way and also invests more to reduce cost. If this demand increases influences Manufacturer's learning more than Supplier's, Supplier is worse off and thus is more likely to give up production, while if it influences Supplier's learning more, Supplier is better off and thus is more likely to produce some of the production.

Similarly, for x_2 , we compare the option of outsourcing partially and outsourcing everything. As discussed above we are not clear whether a larger demand in the second period makes the former option more or less promising. For the latter option, demand change also affects Supplier's profit in two ways. With a larger demand, on one hand Supplier is more eager to learn and has a larger demand in the second period if he produces, on the other hand, Manufacturer is also eager to learn, so Supplier has to offer a much deeper discount in order for Manufacturer to give up production. It is not immediately clear which of these two impacts dominates the other.

Effect of minimum quantity of Manufacturer q_{min}^{M} :

At $c_1^M - c_1^S = x_1 < 0$, a larger q_{min}^M increases Supplier's profit if he produces partial demand in the first period. This is because Supplier loses profit by making $Q_1 - q_{min}^M$ in the first period, and a larger q_{min}^M reduces this profit loss while keeping all the other decisions the same. Therefore as q_{min}^M increases, x_1 decreases, and it is more likely to produce at both facilities than to produce everything in house.

Similarly, at $c_1^M - c_1^S = x_2$, if Supplier produces partial demand, a larger q_{min}^M increases Supplier's profit if $x_2 < 0$, and decreases Supplier's profit if $x_2 > 0$. If Supplier produces everything, since Supplier has to offer a discount $\omega_1 = c_1^M - G^{M*}(Q_2)/q_{min}^M$, a larger q_{min}^M allows Supplier to offer a higher wholesale price and thus increases his profit. Therefore if $x_2 > 0$, as q_{min}^M increases, x_2 decreases, and it is more likely to outsource everything than to produce at both facilities. If $x_2 < 0$ it is not clear how x_2 changes with q_{min}^M

Effect of minimum quantity of Supplier q_{min}^S : Since Manufacturer does not benefit from Supplier's learning, q_{min}^S does not influence Manufacturer's outsourcing decisions, and thus has no effect on the game outcome.

4.3.4 A Special Example

In this section we look at an example with special cost-reducing functions: $g^M(\epsilon, \delta) = a\sqrt{\epsilon}$, and $g^S(\epsilon, \delta) = b\sqrt{\epsilon}$. In this case learning activities at both firms are deterministic and take a special square-root form. Coefficients a and b represent the ability of learning, larger a and b imply stronger learning abilities.

The equilibrium outcome can be characterized in the following lemma: Note that sometimes the outcome $x_1 = x_2$ is possible. In this case it is never optimal to both make and buy. In the deterministic case, the necessary and sufficient condition for $x_1 = x_2$ is:

$$\frac{Q_1}{q_{min}^M}G^{M^*}(Q_2) - g^M(\epsilon_1^{M^e})Q_2(Q_1 + Q_2) < (\frac{Q_1}{q_{min}^M}G^{M^*}(Q_2) - G^{S^*}(Q_2))q_{min}^M < 0$$

Lemma 16 If the cost reducing functions take special forms $g^M(\epsilon, \delta) = a\sqrt{\epsilon}$, and $g^S(\epsilon, \delta) = b\sqrt{\epsilon}$, let $A = \frac{1}{4}(b^2 - 2a^2)Q_2^2$ and $B = \frac{1}{4}(b^2 - \frac{Q_1}{q_{min}^M}a^2)Q_2^2$, then there are four cases of possible equilibrium outcome:

Case 1: Manufacturer produces (Q_1, Q_2) if $c_1^M - c_1^S < \frac{-B}{Q_1 + Q_2}$, and (0, 0) otherwise.

Case 2: Manufacturer produces
$$(Q_1, Q_2)$$
 if $c_1^M - c_1^S < \frac{C_1 + C_2}{Q_1 + Q_2 - q_{min}^M}$, $(q_{min}^M, 0)$ if $\frac{-A}{Q_1 + Q_2 - q_{min}^M}$ $\leq c_1^M - c_1^S < \frac{B - A}{q_{min}^M}$, and $(0, 0)$ otherwise.

Case 3: Manufacturer produces (Q_1, Q_2) if $c_1^M - c_1^S < 0$, (q_{min}^M, Q_2) if $0 \le c_1^M - c_1^S < \frac{-A}{Q_2}$, and (0,0) otherwise.

Case 4: Manufacturer produces (Q_1, Q_2) if $c_1^M - c_1^S < 0$, (q_{min}^M, Q_2) if $0 \le c_1^M - c_1^S < \frac{-A}{Q_2}$, $(q_{min}^M, 0)$ if $\frac{-A}{Q_2} \le c_1^M - c_1^S < \frac{B - A}{q_{min}^M}$, and (0, 0) otherwise.

Conditions for the four cases are as follows:

$$\begin{split} &Case \ 1: \ B \geq 0, \frac{A}{Q_1 + Q_2 - q_{min}^M} \leq \frac{B}{Q_1 + Q_2}. \\ &Case \ 2: \ A \geq 0, \frac{A}{Q_1 + Q_2 - q_{min}^M} \geq \frac{B}{Q_1 + Q_2}. \\ &Case \ 3: \ A \leq 0, B \leq 0, \frac{A}{Q_2} \leq \frac{B}{Q_2 + q_{min}^M}. \\ &Case \ 4: \ A \leq 0, B \leq 0, \frac{A}{Q_2} \geq \frac{B}{Q_2 + q_{min}^M}. \end{split}$$

An example of situation which satisfies this condition could be a case when Manufacturer can reduce his cost very significantly but the profit does not significantly in the absolute terms, while Supplier gains significant extra profit due to learning, and q_{min}^{M} is close to Q_{1} . In this case dual sourcing is not promising and only fully internal production or full outsourcing can be optimal.

For the deterministic cost-reducing functions we can easily provide examples for the ambiguous effect of parameters, which was discussed in the last section. For example, $x_2 = \frac{B-A}{q_{min}^M} = \frac{1}{4}a^2Q_2^2(2-\frac{Q_1}{q_{min}^M})\frac{1}{q_{min}^M}$, so if B > A then x_2 increases in Q_2 , otherwise x_2 decreases in Q_2 .

Similarly, note that x_2 increases in q_{min}^M in this case. However, if it is in case 1, then $x_2 = \frac{-B}{Q_1 + Q_2} = \frac{1}{4} \frac{1}{Q_1 + Q_2} (a^2 \frac{Q_1}{q_{min}^M} - b^2)$. Since B > 0 in this case, x_2 decreases in q_{min}^M .

4.4 Three-period Deterministic Problem

We extend our two-period model into three periods, but for analytical trackability, we only consider a deterministic problem.

Manufacturer's profit-to-go function is no longer monotone in cost when time horizon is extended to three periods. Manufacturer may benefit from higher cost, and consequently be reluctant to improve. Anticipating this response Supplier may charge higher wholesale price and nonintuitive quantity allocations may appear in the equilibrium.

Lemma 17 In the first period, (a) Manufacturer never produces $(Q_1 - q_{min}^S)_+$, and (b) Manufacturer produces $Q_1 - q_{min}^S$ only when he produces (0,0) in the next two periods.

In this section, we analyze the case where the equilibrium quantity in the last two periods could only be (Q_2, Q_3) or (0,0). Analysis of the other scenarios is similar.

We say Supplier "fully participates" when $(q_2^M, q_3^M) = (0, 0)$. Let $A = \frac{M}{Q_1 + Q_2}$. Based on our analysis of two-period problem, Supplier fully participates only if $c_2^M - c_2^S \geq A$. Supplier's profit is given by:

$$R_2^S = \begin{cases} 0 & \text{if } c_2^M - c_2^S < A \\ (c_2^M - c_2^S)(Q_2 + Q_3) + N & \text{otherwise} \end{cases}$$

First we study a benchmark case where R^M takes the same form as $c_2^M < c_2^S + A$ for all c_2^M , i.e., $R^M = (P_2 - c_2^M)Q_2 + (P_3 - c_3^M)Q_3 + G^{M*}(Q_3)$. Then Manufacturer's profit is independent of c_2^S . Similar to two-period problem, Supplier invests $\varepsilon^{S*}(Q_2 + Q_3)$ if he can produce full quantities in the next two periods. Manufacturer always invests

 $\varepsilon^{M*}(Q_2+Q_3)$. When making production decision, Manufacturer produces $q_1^M=Q_1$ if $\omega_1>c_1^M$, $q_1^M=q_{min}^M$ if $c_1^M\geq\omega_1>\bar{\omega}=c_1^M-\frac{G^{M*}(Q_2+Q_3)}{q_{min}^M}$, and $q_1^M=0$ otherwise.

The equilibrium is given by comparing Supplier's profits of the six options:

$$R^{S}(Q_{1}, Q_{2}, Q_{3}) = 0$$

$$R^{S}(Q_{1}, 0, 0) = 0 + (c_{2}^{M} - \Delta c^{M}(Q_{2} + Q_{3}) - c_{2}^{S})(Q_{2} + Q_{3}) + G^{S*}(Q_{3})$$

$$R^{S}(q_{min}^{M}, Q_{2}, Q_{3}) = (c_{1}^{M} - c_{1}^{S})(Q_{1} - q_{min}^{M})$$

$$R^{S}(q_{min}^{M}, 0, 0) = (c_{1}^{M} - c_{1}^{S})(Q_{1} - q_{min}^{M}) + (c_{2}^{M} - \Delta c^{M}(Q_{2} + Q_{3}) - c_{2}^{S})(Q_{2} + Q_{3})$$

$$+G^{S*}(Q_{3}) + G^{S*}(Q_{2} + Q_{3})$$

$$R^{S}(0, Q_{1}, Q_{2}) = (\omega_{1} - c_{1}^{S})Q_{1}$$

$$R^{S}(0) = (\omega_{1} - c_{1}^{S})Q_{1} + (c_{2}^{M} - c_{2}^{S})(Q_{2} + Q_{3}) + G^{S*}(Q_{3}) + G^{S*}(Q_{2} + Q_{3})$$

Note that if R^M takes the same form as $c_2^M > c_2^S + A$ for all c_2^M , i.e., $R^M = (P_2 - c_2^M)Q_2 + (P_3 - c_3^M)Q_3 + \frac{Q_2}{q_{min}^M}G^{M*}(Q_3)$, the equilibrium is the same as above since the marginal benefit of effort is unchanged.

Lemma 18 Supplier always invests in $\varepsilon^{S^*}(Q_2+Q_3)$ if $c_2^M-c_1^S \geq A-\frac{G^{S^*}(Q_2+Q_3)+N}{Q_2+Q_3} = A-D$ and otherwise invests no effort.

Lemma 19 If only Manufacturer can learn, then there exist $E = \Delta c^M(Q_2 + Q_3)$ and $B \ge 0$ satisfying $(\frac{Q_2}{q_{min}^M} - 1)G^{M*}(Q_3) + f^{M-1}(B) + B(Q_2 + Q_3) = G^{M*}(Q_2 + Q_3)$, such that Manufacturer invests $\varepsilon^{M*}(Q_2 + Q_3)$ if $c_1^M - c_2^S < A + B$ or $c_1^M - c_2^S > A + E$, where f^{M-1} is the inverse function of f^M and c_2^S is Supplier's second-period cost given by c_1^M , q_{min}^M and Lemma 18. Otherwise he invests $f^{M-1}(c_1^M - c_1^S - A)$ and after learning $c_2^M - c_2^S = A$ (See Figure ??).

Theorem 11 (a) for $c_1^M - c_1^S < A - D$ or $c_1^M - c_1^S \ge A + E$, the equilibrium is identical to the benchmark case. (b) For $A - D \le c_1^M - c_1^S < A + E$, $(Q_1 - q_{min}^S, 0, 0)$ may take place in the equilibrium. (c) For $A - D \le c_1^M - c_1^S < A + E$, if the equilibrium quantity $q_1^M = 0$, then the wholesale price ω_1 is higher than the one in the benchmark case.

Since Manufacturer greatly benefits from Supplier's full participation, he may prefer keeping own cost at a higher level, providing Supplier more bargaining power and thus increasing wholesale price.

Because the profit-to-go functions under optimal effort decisions are very irregular, we cannot prove that the equilibrium quantity allocation is monotonic in initial costs for three-period model. We can, however, show that only a subset of the allocations are possible to appear in the equilibrium, and we offer the interpretation for each possible allocation later.

Theorem 12 In three-period model, the only possible quantities in the equilibrium are: (Q_1, Q_2, Q_3) , (q_{min}^M, Q_2, Q_3) , $(q_{min}^M, q_{min}^M, Q_3)$, $(q_{min}^M, q_{min}^M, 0)$, $(q_{min}^M, 0)$, $(q_{min}^M, 0)$, $(Q_1 - q_{min}^S, 0, 0)$, $(Q_1, 0, 0)$, $(Q_1, q_{min}^M, 0)$, and $(Q_1, q_{min}^M, 0)$. Furthermore, if demands in each period are stationary, i.e., $Q_1 = Q_2 = Q_3$, then $(Q_1, 0, 0)$, $(Q_1, q_{min}^M, 0)$, and $(Q_1, q_{min}^M, 0)$, and $(Q_1, q_{min}^M, 0)$, and $(Q_1, q_{min}^M, 0)$, are not possible in the equilibrium.

With non-stationary demand, it is easy to provide examples where the quantities listed in Theorem 12 are possible:

• $(0, q_{min}^M, 0)$: Suppose $Q_1 \ll Q_2$, $\Delta c^M(Q_2 + Q_3)$ is large and $G^{M*}(Q_2 + Q_3) \approx 0$. Then $\omega_1 \approx c_1^M$, since Q_1 is small, the profit loss of offering discount is small, and $q_1^M = 0$ can be optimal. In the second period, Q_2 is large, so the profit loss of offering discount is large as long as $G^{M^*}(Q_3) > 0$, and $q_2^M = q_{min}^M$ may be optimal.

Such quantities are also possible for stationary demand. Let $f^S=0$, $f^M=1+\frac{\varepsilon^M}{2Q}$ for $0\leq \varepsilon^M\leq 8Q$ and $f^M=5$ otherwise. Assume $q^M_{min}=Q/4$, then it can be verified that $G^{M^*}(Q)=Q$, $G^{M^*}(2Q)=2Q$, M=4Q, N=Q, it satisfies the condition of scenario 3 in Theorem 1 with $A_1=0$, $A_2=1$, $A_3=12$. For $4.5< c^M_1-c^S_1<5$, compare the profit of Manufacturer for $(0,q^M_{min},0)$, (Q,Q,Q) and (q^M_{min},Q,Q) , we can verify that the equilibrium quantity is $(0,q^M_{min},0)$. With higher internal, most of the production takes place at Supplier. But in the second period Manufacturer produces the minimum quantity to reduce cost. With only one period to go, Manufacturer can only reduce cost by a small amount of effort, not hurting Supplier's future profit too much. ($\Delta c^{M^*}(Q)=1$, $G^{M^*}(Q)=Q$.) In the first period, however, Manufacturer wants to improve a lot given more demand in the future. But the profit increase from cost reduction is relatively small due to the high effort cost ($\Delta c^{M^*}(2Q)=8$, $G^{M^*}(2Q)=2Q$). So Manufacturer is willing to accept a small discount to give up improvement.

• $(Q_1, 0, 0)$: Suppose $Q_1 >> Q_2$, $f^S = 0$, and $c_1^M < c_1^S$. Then $\omega_1 \le c_1^M < c_1^S$, and Supplier gets the order and incurs large profit loss due to the large quantity. In period 2 with smaller demand, Supplier can afford to offer $\omega_2 < c_2^S = c_1^S$ in order to benefit in Period 3. This example shows that when demands are non-stationary and Supplier has higher cost, he has to strategically choose the

time to step in the game, i.e., in the period with little demand so that he does not need to lose too much.

The explanation for $(Q_1, q_{min}^M, 0)$ is similar.

• Example for $(0, q_{min}^M, Q_3)$ when demands are not stationary:

Suppose $f^M(\epsilon^M) = 1$ for all ϵ^M , $f^S(\epsilon^S) = \epsilon^S$ if $\epsilon^S \le 1$ and $f^S(\epsilon^S) = 1$ otherwise. Let $Q_2 = 8$, $Q_3 = 0.5$, $q_{min}^M = 1$, then we can check that $(0, q_{min}^M, Q_3)$ is the equilibrium quantity for $-0.75 < c_1^M - c_1^S < 0$.

In this example $Q_2 >> Q_3$. Although Manufacturer has bigger incentive to learn in the first period than in the second one, Supplier prefers blocking Manufacturer from learning by offering a significant discount. In Period 2, however, Supplier does not improve any more with small demand in Period 3, only produces part of current-period demand, and gives up production in Period 3.

We can interpret the other possible quantity allocations as follows:

- 1. (Q_1, Q_2, Q_3) : Manufacturer has very low cost or very strong learning ability.
- 2. $(q_{min}^M, q_{min}^M, Q_3)$ and (q_{min}^M, Q_2, Q_3) : Manufacturer has high initial cost but strong learning ability. He can exploit the Supplier's low cost at the beginning but finally match Supplier's cost.
- 3. $(q_{min}^M, 0, 0)$ and $(q_{min}^M, q_{min}^M, 0)$: Manufacturer wants to exploit Supplier's strong learning ability, but he reduces his own cost since it is the upper bound of Supplier's wholesale price.
- 4. (0,0,0): Supplier has very low cost or very strong learning ability.
- 5. $(Q_1 q_{min}^S, 0, 0)$: Supplier fully participates only if he is allowed to learn and

Manufacturer's next period cost is not too low. Thus, Manufacturer accepts a wholesale price higher than his own cost, $\omega_1 > c_1^M$, and invests less effort in order to induce Supplier to fully participate.

In the two-period game, Manufacturer is always better off with lower Supplier cost. One interesting question is whether Manufacturer is willing to help Supplier improve. Consider the following revised game: Supplier offers wholesale price ω_1 , and Manufacturer decides quantity. After that Manufacturer announces his own effort level ε^M , and Manufacturer will offer effort ε^{MS} to help Supplier improve. Finally Supplier decides his effort level ε^S . Supplier's improvement is $f^S(\varepsilon^{MS} + \varepsilon^S)$, but he only incurs effort cost ε^S , while Manufacturer incurs effort cost $\varepsilon^M + \varepsilon^{MS}$ with cost reduction $f^M(\varepsilon^M)$.

Observation 1 In a three-period game with deterministic additive cost-reducing functions and the option of offering effort cost sharing, Manufacturer can be worse off if (a) The conditions of scenario 4 in 1 are satisfied; (b) $\varepsilon^{S^*}(Q_2 - q_{min}^M) = 0$; (c) A - D > 0 where A is the smallest cost difference $c_2^M - c_2^S$ for (0,0), and A - D is the minimal cost difference $c_2^M - c_1^S$ at which Supplier can earn nonnegative profit in the last two periods by learning; (d) $(\frac{Q_2}{q_{min}^M} - 1)G^{M^*}(Q_3) > G^{M^*}(Q_2 + Q_3)$; and (e) There exists $c_1^M - c_1^S < A - D$ such that (0,0,0) is the optimal quantity.

Despite that Manufacturer always has the choice of sharing nothing, he may still be hurt just by having the option of offering effort sharing. Manufacturer's internal learning is not as important as before since he has the option of helping Supplier improve. Knowing this Supplier does not need to offer a wholesale price as low as before to gain the total production, hurting Manufacturer's profit in the long run.

4.5 Conclusion

In this paper, we study a "make-or-buy" problem when both Manufacturer and Supplier can invest in a costly effort to reduce production cost. This learning requires the plant to produce a minimum amount of production to provide sufficient experience for post-savings activities to take place. Although dual sourcing has been studied before and various reasons have been provided for this decision, we provide a new explanation for this phenomenon, which relates partial internal production to learning objective that enable cost savings.

In the two-period problem, we have shown that depending on the cost difference between Manufacturer's and Supplier's, Manufacturer can either fully outsource to the Supplier, fully produce in house, or proceed with partial production and partial outsourcing. We find that this dual sourcing is more likely when the learning process is more uncertain, when Manufacturer's minimum quantity required to learn is larger, or when the two plants belong to the same company.

We extend the problem to three periods and find that the production quantity can switch back and forth from full to partial outsourcing. Although Manufacturer has no incentive to help Supplier learn in two periods, he does have such an incentive in longer horizons. However, this is not always the case because sometimes simply having such an option allows Supplier to worsen his offer and exploit Manufacturer.

CHAPTER 5

Summary

This dissertation is focused on the dynamic production and pricing decisions for new products. The three essays contained in the dissertation provide theoretical results and numerical studies for three different aspects of new product decision making.

The first essay studies how a capacity constrained firm should price products during new product introductions. using control-theory framework (a generalization of the classic Bass model), I consider the integrated optimal pricing, production, and sales decisions. I show that in most of the cases, the optimal trajectory of demand is unimodal, as in the Bass model, but the optimal price trajectory can have multiple local maxima when capacity is limited. Lack of pricing flexibility might lead to intentional creation of backorders or lost sales, a phenomenon not observed when price adjustments are allowed. I also explore when pricing flexibility is most valuable using a numerical study. I find that benefits are highest when capacity is not unlimited nor very little, and when imitation effect dominates innovation effect. I also find that the capability to adjust prices is significantly more effective than the option of producing in advance and holding inventory.

The second essay is focused on the production lot sizing decisions for a pharmaceutical firm that manufactures new medicines for clinical trial. I derive structural results on the optimal policy and show that under reasonable assumptions about the yield distributions, the optimal production quantities are increasing in the number of waiting orders and the remaining quantity to be produced for the current lot. I also conduct a numerical study to compare the optimal average cost with several heuristic policies commonly used in the industry. The results indicate that in most cases the simple heuristic policies used in practice are very inefficient and substantial savings can be obtained by appropriate lot sizing policies.

The third essay deals with outsourcing decisions when firms manufacture a new product and production costs can be reduced through "learning-by-doing" effect. I provide a new explanation for dual sourcing, which is driven by the fact that in-house production may facilitate learning about potential process improvements leading to eventual cost reductions.

These three essays deal with different issues that are important for new product production and marketing. Various extensions are likely based upon this work.

APPENDICES

APPENDIX A

Main Proofs in Chapter 2

A.1 Proof of Lemma 1

Proof: Applying Pontryagin's Maximum Principle (Sethi and Thompson, 2000, p. 33), a necessary condition for optimality is that there exist optimal $d^*(t)$, $D^*(t)$, $d_B^*(t)$ and a continuous function $\lambda(t)$ such that $\frac{\partial H}{\partial d} = 0$, $\frac{d\lambda}{dt} = r\lambda(t) - \frac{\partial H}{\partial D}$, and $\lambda(T) = 0$. Since,

$$\frac{\partial H}{\partial d} = \log d_B^*(t) - \log d^*(t) - c - 1 + \lambda(t)$$

we can obtain the optimal demand as a function of λ :

$$\log d^*(t) = \log d_B^*(t) - c - 1 + \lambda(t) \tag{A.1}$$

$$\frac{d\lambda}{dt} = r\lambda(t) - \frac{\partial H}{\partial D} = r\lambda(t) - d^*(t) \frac{\partial \log d_B^*(t)}{\partial D} = r\lambda(t) - \frac{\partial D^*(t)}{\partial t} \frac{\partial \log d_B^*(t)}{\partial D}
= r\lambda(t) - \frac{d\log d_B^*(t)}{dt}$$
(A.2)

Taking the first derivative of t on both sides of (A.1) and substituting (A.2) for $\frac{d\lambda}{dt}$, we have

$$\frac{d\log d^*(t)}{dt} = \frac{d\log d_B^*(t)}{dt} + \frac{d\lambda}{dt} = \frac{d\log d_B^*(t)}{dt} + r\lambda(t) - \frac{d\log d_B^*(t)}{dt} = r\lambda(t) \quad (A.3)$$

Therefore, if r > 0, the trajectory of the optimal demand depends on λ : $d^*(t)$ increases if $\lambda(t) > 0$, and decreases if $\lambda(t) < 0$.

 $\lambda(t)$ can be solved from differential equation (A.2) and $\lambda(T) = 0$:

$$\lambda(t) = \int_{t}^{T} e^{r(t-u)} d\log d_{B}^{*}(u)$$
(A.4)

To examine the shape of $\lambda(t)$, we further examine equation (A.2). Since $\lambda(T)=0$, $\frac{d\lambda}{dt}|_{t=T}=-\frac{d\log d_B(t)}{dt}|_{t=T}$. Note first that $\frac{d\log d_B(t)}{dt}|_{t=T}>0$ is equivalent to $\frac{\partial d_B}{\partial D}|_{t=T}>0$, i.e. , $D(T)<(q-p)\frac{m}{2q}$. Consider the following cases:

Case 1: $D(T) < (q-p)\frac{m}{2q}$ Since $d_B(t)$ is unimodal in t, $\frac{d\log d_B(t)}{dt}|_{t=T} > 0$ implies $\frac{d\log d_B(t)}{dt} > 0$ for all $t \le T$. By (A.4) $\lambda(t) \ge 0$ for all $t \in [0,T]$. Thus the optimal demand $d^*(t)$ is monotonically increasing in [0,T].

Case 2: $D(T) \geq (q-p)\frac{m}{2q}$ In this case $\lambda(T) = 0$ and $\frac{d\lambda}{dt}|_{t=T} \geq 0$. So $\lambda(T-\delta) \leq 0$ for small enough $\delta > 0$. We show first that there is at most one additional zero-point of $\lambda(t)$ for $t \in [0,T)$. Assume that there exist $\tau_1 < \tau_2 \in [0,T]$ such that $\lambda(\tau_1) = 0$, $\lambda(\tau_2) = 0$, $\frac{d\lambda}{dt}|_{t=\tau_1} > 0$ and $\frac{d\lambda}{dt}|_{t=\tau_2} < 0$. From (A.2), we have $\frac{d\log d_B(t)}{dt}|_{t=\tau_1} < 0$ and $\frac{d\log d_B(t)}{dt}|_{t=\tau_2} > 0$, which is impossible given $\tau_1 < \tau_2$ and d_B is unimodal. Depending on the value of $\lambda(0)$ we have two possibilities

- Case 2.1: If $\lambda(0) \leq 0$, then due to no other zero points, we must have $\lambda(t) \leq 0$ for all t, and the optimal demand $d^*(t)$ monotonically decreases over time.
- Case 2.2: If $\lambda(0) > 0$, then there must exist $\tau_d \in [0,T]$ such that $\lambda(t) > 0$ for $t < \tau_d$ and $\lambda(t) \leq 0$ otherwise. So the optimal demand $d^*(t)$ first increases and then decreases over time.

A.2 Proof of Lemma 2

Proof: From equation (A.1) and $\pi(t) = \log d_B(t) - \log d(t)$, we have $\pi^*(t) = 1 + c - \lambda(t)$. Therefore $\pi^*(t)$ increases if and only if $\lambda(t)$ decreases. We will show that $\lambda(t)$ has a unique local maximum, by showing $\frac{d^2\lambda}{dt^2} \geq 0$ whenever $\frac{d\lambda}{dt} = 0$. Since $\frac{d\pi^*(t)}{dt} = -\frac{d\lambda(t)}{dt}$, when $\frac{d\lambda(t)}{dt} = 0$, $\frac{d\pi^*(t)}{dt} = 0$. From (A.2), at local minimum,

$$\frac{d^{2}\lambda(t)}{dt^{2}} = r\frac{d\lambda(t)}{dt} - \frac{d^{2}}{dt^{2}}\log d_{B}(t) = -\frac{d^{2}}{dt^{2}}\log d_{B}(t)$$

$$\frac{d}{dt}\log d_{B}(t) = \frac{d}{dt}(\log(m-D(t)) + \log(p+qD(t)/m)$$

$$= \frac{-d(t)}{m-D(t)} + \frac{qd(t)/m}{p+qD(t)/m}$$

$$= -\frac{(m-D(t))(p+qD(t)/m)e^{-\pi(t)}}{m-D(t)} + \frac{q/m(m-D(t))(p+qD(t)/m)e^{-\pi(t)}}{p+q/mD(t)}$$

$$= -(p+qD(t)/m)e^{-\pi(t)} + q/m(m-D(t))e^{-\pi(t)}$$

$$= (-(p+qD(t)/m) + q/m(m-D(t)))e^{-\pi(t)}$$

$$= (q-p-2qD(t)/m)e^{-\pi(t)}$$

$$= (4^{2}dt^{2}\log d_{B}(t)) = -2qd(t)e^{-\pi(t)}/m - (q-p-2qD(t)/m)e^{-\pi(t)}\frac{d\pi}{dt}$$
(A.5)

Since $\pi'(t) = -\lambda'(t) = 0$, $\frac{d^2\lambda(t)}{dt^2} = -\frac{d^2}{dt^2}\log d_B(t) = 2q/md(t)e^{-\pi(t)} \ge 0$. Therefore

 $\lambda(t)$ has a unique local minimum and $\pi^*(t)$ has a unique local maximum.

Case 1: If $D(T) < (q-p)\frac{m}{2q}$, we have $\frac{d\log d_B(t)}{dt}|_{t=T} > 0$. In this case λ is monotonically decreasing, therefore $\pi^*(t)$ monotonically increases.

Case 2: If $D(T) \ge (q-p)\frac{m}{2q}$, we have $\frac{d\log d_B(t)}{dt}|_{t=T} \le 0$.

Case 2.1: If
$$\frac{d\lambda}{dt}|_{t=0} > 0$$
, i.e., $r\lambda(0) - \frac{d}{dt}\log d_B(t)|_{t=0} = r\lambda(0) - (\frac{-d(t)}{m - D(t)} + \frac{q/md(t)}{p + q/mD(t)})|_{t=0} = r\lambda(0) - mpe^{\pi^*(0)}(q/p - 1)/m = r\lambda(0) - (q - p)e^{\pi^*(0)} > 0$

0, then $\lambda(t)$ is monotonically increasing, therefore $\pi^*(t)$ monotonically decreases.

Case 2.2: If $\frac{d\lambda}{dt}|_{t=0} \leq 0$, *i.e.* , $r\lambda(0) - \frac{d}{dt}\log d_B(t)|_{t=0} \leq 0$, then there exists $\tau_{\pi} \in [0,T]$, such that $\pi^*(t)$ increases for $t \in [0,\tau_{\pi}]$, and decreases afterwards.

A.3 Proof of Lemma 3

Proof: We showed that d, d_B and π are all unimodal. Furthermore, if we are forced to set a fixed price, then the peak time τ_F is the same as the peak time of d_B . Thus, we consider first-order conditions. From (A.3), $\lambda(\tau_d) = 0$, and from proof of Lemma 1 $\frac{d\lambda}{dt}|_{\tau_d} \leq 0$. By (A.2) $\frac{d\log d_B}{dt}|_{\tau_d} = r\lambda(\tau_d) - \frac{d\lambda}{dt}|_{\tau_d} \geq 0$, implying $\tau_d \leq \tau_F$.

Since price π is maximized at the minimum of λ , we have $\frac{d\lambda}{dt}|_{\tau_{\pi}} = 0$. Combined with $\lambda(T) = 0$ it implies $\lambda(\tau_{\pi}) \leq 0$. By (A.2), $\frac{d \log d_B}{dt}|_{\tau_{\pi}} = r\lambda(\tau_{\pi}) - \frac{d\lambda}{dt}|_{\tau_{\pi}} \leq 0$. Thus, $\tau_B \leq \tau_{\pi}$.

A.4 Proof of Lemma 4

Proof: When the discount factor r = 0, equation (A.4) simplifies to

$$\lambda(t) = \log d_B^*(T) - \log d_B^*(t) \tag{A.6}$$

Substitute (A.6) into (A.1),

$$\log d^*(t) = \log d_B^*(t) - c - 1 + (\log d_B^*(T) - \log d_B^*(t)) = -1 - c + \log d_B^*(T)$$

Therefore $d^*(t)$ is a constant, say $d^*(t) = \bar{d}$ for all t, $D(T) = \bar{d}T$, and \bar{d} solves the following equation:

$$\log \bar{d} = -1 - c + \log[(m - \bar{d}T)(p + \frac{q}{m}\bar{d}T)]$$

or,

$$(m - \bar{d}T)(p + \frac{q}{m}\bar{d}T) = e^{1+c}\bar{d}$$

This quadratic equation has one positive solution:

$$\bar{d} = \frac{1}{2q/mT^2} \{ -(p-q)T - e^{1+c} + \sqrt{((p-q)T + e^{1+c})^2 + 4pqT^2} \}$$

The optimal price is given by $\pi(t) = \log d_B(t) - \log d(t)$:

$$\pi(t) = 1 + c - \lambda(t) = 1 + c - (\log d_B(T) - \log d_B(t)) = 1 + c + \log \frac{(m - \bar{d}t)(p + q/m\bar{d}t)}{(m - \bar{d}T)(p + q/m\bar{d}T)} \quad \blacksquare$$

A.5 Proof of Lemma 5

Proof: The problem in Section 2.3 can be simplified as follows:

$$\Pi = \max_{\pi(t), s(t)} \int_0^T (\pi(t) - c) s(t) e^{-rt} dt$$
s.t.
$$d(t) = (p + \frac{q}{m} S(t)) (m - D(t)) e^{-\pi(t)},$$

$$\frac{dD}{dt} = d(t),$$

$$\frac{dS}{dt} = s(t),$$

$$0 \le s(t) \le K, 0 \le s(t) \le d(t),$$

$$D(0) = S(0) = 0.$$

Again it is equivalent to finding the optimal d(t) and s(t). Denote unsatisfied demand as $\tilde{s}(t) = d(t) - s(t) \ge 0$. As $\pi(t) = \log d_B(t) - \log d(t)$, the problem can be

expressed as:

$$\Pi = \max_{s(t),\tilde{s}(t)} \int_0^T (\log d_B(t) - \log(s(t) + \tilde{s}(t)) - c) s(t) e^{-rt} dt$$
s.t.
$$\frac{dD(t)}{dt} = s(t) + \tilde{s}(t),$$

$$\frac{dS(t)}{dt} = s(t),$$

$$\tilde{s}(t) \geq 0,$$

$$0 \leq s(t) \leq K.$$
(A.8)

Write Hamiltonian:

$$H(D, S, s(t), \tilde{s}(t), \lambda_1, \lambda_2, t) = (\log d_B(t) - \log(s(t) + \tilde{s}(t)) - c)s(t)$$
$$+\lambda_1(t)(s(t) + \tilde{s}(t)) + \lambda_2(t)s(t)$$

By Pontryagin Maximum Principle, $\lambda_1(T) = 0$ and

$$\frac{d\lambda_1}{dt} = r\lambda_1 - \frac{\partial H}{\partial D} = r\lambda_1 + \frac{s^*(t)}{m - D^*(t)}$$
Thus $\lambda_1(t) = -\int_t^T e^{r(t-\mu)} \frac{s^*(\mu)}{m - D^*(\mu)} d\mu$. Since $\frac{s^*(\mu)}{m - D^*(\mu)} \ge 0$, $\lambda_1(t) \le 0$ for all $t \in [0, T]$.

Since $s^*(t)$ and $\tilde{s}^*(t)$ maximizes H subject to constraints (A.7) and (A.8), and

$$\frac{\partial H}{\partial \tilde{s}} = -\frac{s(t)}{s(t) + \tilde{s}(t)} + \lambda_1(t) \le 0,$$

we have $\tilde{s}^*(t) = 0$ for all t, and $d(t) = s(t) \le K$.

A.6 Proof of Theorem 1

Proof: Define Hamiltonian H as:

$$H(D, d, \lambda, t) = (\log d_B(t) - \log d(t) - c)d(t) + \lambda(t)d(t)$$

By Maximum Principle, $d^*(t)$ maximizes H subject to the constraint $0 \le d^*(t) \le K$.

$$\frac{\partial H}{\partial d} = \log d_B(t) - 1 - c + \lambda(t) - \log d(t)$$

Denote $M(t) = \log d_B(t) - 1 - c + \lambda(t)$. The same steps as in the proof of Lemma 1 imply $\frac{dM}{dt} = r\lambda(t)$, and $\lambda(t)$ is positive, or M(t) increases, for $t \leq \tau_d$ and negative, M(t) decreases, for $\tau_d < t < T$.

Let $\bar{K} = e^{\max_{t \in [0,T]} M(t)}$. If $K \geq \bar{K}$, then $d^*(t) = e^{M(t)}$ satisfies the first order condition $\frac{\partial H}{\partial d} = 0$ and satisfies the capacity constraint. In that case it is clear that $d^*(t)$ is first increasing and then decreasing.

If $K < \bar{K}$, then let $t_1^A < t_2^A$ be the times where $M(t_1^A) = M(t_2^A) = \log K$. Then, for $t \in [0, t_1^A]$, $d^*(t) = e^{M(t)}$, which is increasing, and $d^*(t) = e^{M(t)}$ on $[t_2^A, T]$, which is decreasing, while $d^*(t) = K$ on $[t_1^A, t_2^A]$,

Consider now the trajectory of price. On $[0, t_1^A]$, demand $\log d^*(t) = \log d_B(t) - 1 + \lambda(t) - c$, so $\pi^*(t) = \log d_B(t) - \log d(t) = 1 + c - \lambda(t)$. Since $\lambda(t)$ is decreasing on $[0, t_1^A]$, $\pi^*(t)$ is increasing on $[0, t_1^A]$.

On $[t_1^A, t_2^A]$, optimal demand $d^*(t) = K$, so $\pi^*(t) = \log d_B(t) - \log d(t) = \log d_B(t) - \log K$. $\pi^*(t)$ is unimodal on $[t_1^A, t_2^A]$ since $d_B(t)$ is unimodal.

On $[t_2^A, T]$, $\log d^*(t) = \log d_B(t) - 1 + \lambda(t) - c$, so $\pi^*(t) = \log d_B(t) - \log d(t) = 1 + c - \lambda(t)$. As in Lemma 2, $\lambda(t)$ has a unique local minimum on $[t_2^A, T]$ Thus, $\pi^*(t)$ has a unique local maximum on $[t_2^A, T]$. and the theorem is proved.

A.7 Proof of Lemma 6

Proof: The proof is similar to that of Lemma 5. Defining $\tilde{s}(t) = d(t) - s(t) \ge 0$, the problem can be reformulated as:

$$\Pi = \max_{s(t), \tilde{s}(t), x(t)} \int_0^T [(\log d_B(t) - \log(s(t) + \tilde{s}(t)))s(t) - cx(t) - hI(t)]e^{-rt}dt + c_sI(T)$$
s.t.
$$\frac{dD}{dt} = s(t) + \tilde{s}(t),$$

$$\frac{dS}{dt} = s(t),$$

$$\frac{dI}{dt} = x(t) - s(t),$$

$$I(t) \ge 0, I(0) = 0,$$

$$0 \le x(t) \le K,$$

$$\tilde{s}(t) \ge 0.$$

The Hamiltonian becomes:

$$H(D, S, s(t), \tilde{s}(t), \lambda_1, \lambda_2, \lambda_3, t) = (\log d_B(t) - \log(s(t) + \tilde{s}(t)))s(t) - cx(t) - hI(t)$$
$$+\lambda_1(t)(s(t) + \tilde{s}(t)) + \lambda_2(t)s(t) + \lambda_3(t)(x(t) - s(t))$$

By Pontryagin Maximum Principle, $\lambda_1(T) = 0$, and

$$\frac{d\lambda_1}{dt} = r\lambda_1 + \frac{s^*(t)}{m - D^*(t)}$$

Thus $\lambda_1(t) = -\int_t^T e^{r(t-\mu)} \frac{s^*(\mu)}{m - D^*(\mu)} d\mu$. Since $\frac{s^*(\mu)}{m - D^*(\mu)} \ge 0$, $\lambda_1(t) \le 0$ for all $t \in [0,T]$. As \tilde{s}^* maximizes H, and

$$\frac{\partial H}{\partial \tilde{s}} = -\frac{s(t)}{s(t) + \tilde{s}(t)} + \lambda_1(t) \le 0$$

Thus $\tilde{s}^*(t) = 0$ for all t, and d(t) = s(t).

A.8 Proof of Lemma 7

Proof: (a) Since the Hamiltonian H in (10) is linear in x, if $I^*(t) > 0$, then $x^*(t) = 0$ or $x^*(t) = K$.

Assume $I^*(t) > 0$ and $x^*(t) = 0$ for an open interval $[t_1, t_2]$. Since $I^*(t_1) > 0$, there exist an open interval (s_1, s_2) before (t_1, t_2) such that I(s) > 0 and $x^*(s) = K$ for $s \in (s_1, s_2)$. Consider an alternative strategy where $\tilde{x}(s) = K - \delta$ for $s \in (s_1, s_2)$, and $\tilde{x}(t) = \delta * (s_2 - s_1)/(t_2 - t_1)$ for $t \in (t_1, t_2)$, where δ is a small positive number. All other decisions remain the same. This strategy is still feasible. As demand is not changed, the firm collects the same revenue, but incurs lower inventory holding cost and production cost, which contradicts optimality of the original policy.

By equation (2.11), H is linear in x. When $I^*(t) > 0$, we have $x^*(t) = K$, implying that $\frac{\partial H}{\partial x} \ge 0$, or equivalently, $\lambda_2(t) \ge c$.

(b) Let $I^*(t) = 0$ for a small open interval. Since $I^{*'}(t) = 0$, we have $x^*(t) = d^*(t)$. From Lemma 6, $d^*(t) = s^*(t)$.

A.9 Proof of Lemma 8

¹We can always alter the production policy on a set of measure 0, and this is why we consider the interval.

costs decrease. Additionally, the alternative strategy saves production cost. Therefore any production strategy with I(T) > 0 is dominated, contradicting our original assumption.

A.10 Proof of Lemma 9

Proof: The Lagrangian for this problem is

$$L(D, I, d, \lambda_1, \lambda_2, \mu, \eta, t) = (\log d_B(t) - \log d(t))d(t) - cx(t) - hI(t) + \lambda_1(t)d(t) + \lambda_2(t)(x(t) - d(t)) + \mu_1(t)x(t) + \mu_2(t)(K - x(t)) + \eta(t)(x(t) - d(t))$$

By maximum principle,

$$\frac{d\lambda_1}{dt} = r\lambda_1 - \frac{\partial L}{\partial D} = r\lambda_1 - \frac{d\log d_B^*(t)}{dt}$$
(A.9)

$$\frac{d\lambda_2}{dt} = r\lambda_2 - \frac{\partial L}{\partial I} = r\lambda_2 + h \tag{A.10}$$

$$\lambda_1(T) = 0$$

$$\lambda_2(T) \geq 0$$

$$\frac{\partial L}{\partial d} = \log d_B^*(t) - \log d^*(t) - 1 + \lambda_1(t) - \lambda_2(t) - \eta(t) = 0$$
 (A.11)

$$\frac{\partial L}{\partial x} = -c + \lambda_2(t) + \mu_1(t) - \mu_2(t) + \eta(t) = 0$$

$$\mu_1(t) \geq 0, \mu_2(t) \geq 0, \mu_1(t)x^*(t) = 0, \mu_2(t)(K - x^*(t)) = 0$$

$$\eta(t) \ge 0, \frac{d\eta}{dt} \le r\eta(t), \eta(t)I^*(t) = 0$$
(A.12)

If $I^*(t) > 0$, then by Lemma 7, $\lambda_2(t) \geq c$, $x^*(t) = K$. From Equation (A.12),

 $\eta(t) = 0$. From Equation (A.11), $\log d^*(t) = \log d_B(t) - 1 + \lambda_1(t) - \lambda_2(t)$.

If
$$I^*(t) = 0$$
, since $\frac{\partial H}{\partial x} = -c + \lambda_2(t)$, if $\lambda_2(t) \ge c$, then $x^*(t) = K$. As $x^*(t) = K$,

we have
$$\frac{dd}{dx} = 0$$
 and, thus, $\frac{dH}{dd} = \frac{\partial H}{\partial d} = \log d_B(t) - \log d(t) - 1 + \lambda_1(t) - \lambda_2(t)$,

subject to the constraint $d^*(t) \leq K$. Therefore, $\log d^*(t) = \min\{\log K, \log d_B(t) - 1 + \lambda_1(t) - \lambda_2(t)\}$.

If $I^*(t) = 0$ and $\lambda_2(t) < c$, then H is decreasing in x, so $x^*(t) = d^*(t)$. Substituting $d^*(t)$ for x^* in H, we have $\frac{dH}{dd} = \log d_B(t) - \log d(t) - 1 + \lambda_1(t) - c$, subject to the constraint $d^*(t) \le K$. Therefore, $\log d^*(t) = \min\{\log K, \log d_B(t) - 1 + \lambda_1(t) - c\}$.

A.11 Proof of Theorem 2

Proof: We first prove the structure of the optimal demand and production trajectories (part (a) and (c)):

- 1. By Pontryagin's Maximum Principle $\lambda_2(T) \geq 0$, and $\lambda_2(t)$ may have a negative jump whenever inventory I(t) changes from positive to zero. We first show that $\lambda_2(t)$ is piece-wise increasing.
 - Solving equation (A.10), for any maximal interval where $\lambda_2(t)$ is continuous, we have $\lambda_2(t) = Ce^{rt} \frac{h}{r}$, where C is a constant. We show that C is always positive, which implies that $\lambda_2(t)$ is increasing on that interval. Note that at the end of the interval $t = \bar{t}$, either (i) $\lambda_2(t)$ has a negative jump, or (ii) \bar{t} is the end of the horizon. (i) implies that inventory I changes from positive to zero at \bar{t} , and by Lemma 9, part 1, we have $\lambda_2(\bar{t}^-) \geq c \geq 0$. If $\bar{t} = T$, then $\lambda_2(\bar{t}) = \lambda_2(T) \geq 0$. Therefore, in both cases, $C \geq 0$.
- 2. Let $\log d_1(t) := \log d_B(t) 1 + \lambda_1(t) c$. Using the same logic as in the proof of Lemma 1, we have $\log d_1(t)$ is unimodal for all $t \in (0, T)$.
- 3. Let $\log d_2(t) := \log d_B(t) 1 + \lambda_1(t) \lambda_2(t)$. We show that $\log d_2(t)$ is unimodal for any interval where $\lambda_2(t)$ is continuous by justifying that any local optimum

is a local maximum. Consider the first derivative of $\log d_2(t)$. Based on equation

(A.9) and
$$\lambda_2(t) = Ce^{rt} - \frac{h}{r}$$
, $C \ge 0$, we have

$$\frac{d\log d_2(t)}{dt} = \frac{d\log d_B(t)}{dt} + \frac{d\lambda_1(t)}{dt} - \frac{d\lambda_2(t)}{dt}$$

$$= \frac{d\log d_B(t)}{dt} + r\lambda_1(t) - \frac{d\log d_B(t)}{dt} - rCe^{rt} = r\lambda_1(t) - rCe^{rt}$$

At the local optimum of $\log d_2(t)$, $\frac{d \log d_2(t)}{dt} = 0$, thus $\lambda_1(t) = Ce^{rt} \geq 0$. In the proof of Lemma 1, we have shown that $\lambda_1(t)$ is decreasing if $\lambda_1(t) \geq 0$, thus $\frac{d\lambda_1(t)}{dt} \leq 0$ at the local maximum of $\log d_2(t)$.

We evaluate the second derivative of $\log d_2(t)$ at $\frac{d \log d_2(t)}{dt} = 0$ and show it is negative.

$$\frac{d^2 \log d_2(t)}{dt^2} = r \frac{d\lambda_1(t)}{dt} - r^2 C e^{rt} \le 0.$$

Thus, $\log d_2(t)$ is unimodal.

- 4. We show that it is not possible for the three conditions to take place simultaneously on an open interval: $d_2(t) < K$, $d_2(t)$ is decreasing, and $\lambda_2(t) \ge c$.

 If these conditions take place on interval (t',t''), then they hold on (t',T). $\lambda_2(t) \ge c$ implies that x(t) = K (Lemma 9, parts 1 and 2). Combined with $d_2(t) < K$, we have $\frac{dI(t)}{dt} > 0$ on (t',t''). Since d_2 is unimodal, we have $d_2(t) < K$ on (t',T). For x(t) < K, we need $\lambda_2(t) < c$, for which we need a jump down, which is only possible if I(t) = 0. The initial accumulation of inventory combined with d(t) < K = x(t) makes it, however, impossible. Thus, I(T) > 0, which contradicts Lemma 8.
- 5. Based on the observations above, we consider the following cases depending on the values of $\lambda_2(0)$ and $d_2(0)$:

Case 1: If $\lambda_2(0) \ge c$ and $d_2(0) < K$:

By part 1 and 2 of Lemma 9, $x^*(0) = K$ and $d^*(0) = d_2(0) < K$, thus $\frac{dI}{dt} > 0$. Since I(T) = 0, we define $t_2^B := \min\{t \in (0,T] | I(t) = 0\}$. Lemma 9 implies $\underline{x^*(t) = K}$, $\underline{d^*(t) = d_2(t)}$, and $\underline{I(t) > 0}$ on $(0, t_2^B)$. Since $\frac{dI(t)}{dt} \leq 0$ at $t_2^{B^-}$, we have $d_2(t_2^{B^-}) \geq K$. As we have $\frac{dI}{dt} = K - d_2(t)$ on $(0, t_2^B)$, $\frac{dI}{dt}$ must be positivenegative. Thus $\underline{I^*(t)}$ is unimodal for $t \in [0, t_2^B]$.

 $\lambda_2(t)$ may have a negative jump at t_2^B . Recall that $I(t_2^B) = 0$ and either case 2 or 3 of Lemma 9 applies, depending on the value of the jump.

Case 1.1: If $\lambda_2(t_2^{B^+}) \geq c$, then $x^*(t) = K$, $d^*(t) = \min(K, d_2(t)) = K$, and $\frac{dI}{dt} = 0$ for $t \in (t_2^B, t_2^B + \delta)$, where δ is a small positive number. Since, $\lambda_2(t_2^{B^+}) \leq \lambda_2(t_2^{B^-})$ and $d_2(t_2^{B^-}) \geq K$, from Lemma 9, part 1, we have $d_2(t_2^{B^+}) \geq d_2(t_2^{B^-}) \geq K$. Thus we have, $\lambda_2(t_2^{B^+}) \geq c$ and $d_2(t_2^{B^+}) \geq d_2(t_2^{B^-}) \geq K$.

Note that if $d_2(t)$ drops below K at any $t \geq t_2^B$, then it is decreasing at that

point. However, since $\lambda_2(t)$ is increasing, we have $\lambda_2(t) \geq c$, which cannot happen simultaneously. Therefore in this case $\underline{x^*(t)} = d^*(t) = K$, and $I^*(t) = 0$ for $t \in (t_2^B, T)$. Case 1.2: If $\lambda_2(t_2^{B^+}) < c$, then $x(t) = d(t) = \min(K, d_1(t))$, and $\frac{dI}{dt} = 0$ for $t \in (t_2^B, t_2^B + \delta)$, where δ is a small positive number. Since $\lambda_2(t_2^{B^-}) \geq c$, we have $d_1(t_2^B) \geq d_2(t_2^{B^-}) \geq K$. Therefore, we define maximum $t_3^B \geq t_2^B$, such that $d_1(t_3^B) \geq K$, which implies x(t) = d(t) = K for $t \in (t_2^B, t_3^B)$ and $d_1(t_3^B) = K$. Since $d_1(t)$ is unimodal, d_1 is decreasing for $t \geq t_3^B$. Since $\frac{dd_2(t)}{dt} = \frac{dd_1(t)}{dt} - \frac{d\lambda_2(t)}{dt}$, $d_2(t)$ is also decreasing for $t \geq t_3^B$, and we must have $\lambda_2(t) < c$. Thus $\underline{x^*(t)} = d^*(t) = K$ on (t_2^B, t_3^B) , and $\underline{x^*(t)} = d^*(t) = d_1(t)$ on (t_3^B, T) .

Case 2: If $\lambda_2(0) \geq c$ and $d_2(0) \geq K$, the implication of Case 1.1. for $t_2^B = 0$ applies and $x^*(t) = d^*(t) = K$, and I(t) = 0 for $t \in (0, T)$.

Case 3: If $\lambda_2(0) < c$:

Then $\underline{x^*(t)} = d^*(t) = \min(K, d_1(t))$ for $t \in (0, t_1^B)$, where t_1^B is the first time such that $\lambda_2(t_1^B) = c$. At t_1^B , the situation becomes identical to case 1 or 2, depending whether $d_1(t_1^B) < K$ or $d_1(t_1^B) \ge K$.

Case 3.1: If $d_1(t_1^B) < K$, since $\lambda_2(t_1^B) = c$, $I(t_1^B) = 0$, $d_2(t_1^B) = d_1(t_1^B)$, the trajectory after t_1^B is as in Case 1.

Case 3.2: If $d_1(t_1^B) \ge K$, since $\lambda_2(t_1^B) = c$, $I(t_1^B) = 0$, $d_2(t_1^B) = d_1(t_1^B)$, the trajectory after t_1^B is as in Case 2.

Next we prove the structure of the optimal price trajectory (part (b)):

- 1. Let $t' \in (0, t_1^B)$ be the minimal t' such that $d_1(t') \geq K$. It t' exists, we have Case 3.2 above and for all $t \in [t', T]$, we have d(t) = K and thus $\pi(t) = \log d_B(t) \log d(t) = \log d_B(t) \log d(t)$, which clearly means that $\underline{\pi(t)}$ is unimodal. Let t'' = t' if it exists and $t'' = t_1^B$ otherwise. For $t \leq t''$: $\log d(t) = \log d_1(t) = \log d_B(t) 1 + \lambda_1(t) c$, thus $\pi(t) = \log d_B(t) \log d(t) = 1 + c \lambda_1(t)$. As shown in the proof of Lemma 2, $\underline{\pi(t)}$ is increasing since $\log d(t)$ is increasing on (0, t''). When t'' = t' the price structure is already shown, thus consider $t'' = t_1^B$.
- 2. On $t \in (t_1^B, t_2^B)$, $\log d(t) = \log d_2(t) = \log d_B(t) 1 + \lambda_1(t) \lambda_2(t)$, we will show that $\pi(t) = \log d_B(t) \log d(t) = 1 \lambda_1(t) + \lambda_2(t)$ is unimodal. Let $\lambda(t) = \lambda_2(t) \lambda_1(t)$. We will show that $\frac{d^2\pi(t)}{dt^2} = \frac{d^2\lambda(t)}{dt^2} \le 0$ whenever $\frac{d\pi(t)}{dt} = \frac{d\lambda(t)}{dt} = 0$, which implies that $\underline{\pi(t)}$ is unimodal.

Note that if
$$\frac{d\pi(t)}{dt} = \frac{d\lambda(t)}{dt} = 0$$
, by equation (A.9) and (A.10), we have $\frac{d\lambda(t)}{dt} = r\lambda(t) + \frac{d\log d_B(t)}{dt} + h$ and $\frac{d^2\lambda(t)}{dt^2} \Big|_{\frac{d\pi}{dt}} = 0 = \frac{d^2\log d_B(t)}{dt^2}$. From equation (A.5),
$$\frac{d^2}{dt^2}\log d_B(t) = -2qd(t)e^{-\pi(t)}/m - (q-p-2qD(t)/m)e^{-\pi(t)}\frac{d\pi}{dt}$$
we have $\frac{d^2\lambda(t)}{dt^2}\Big|_{\frac{d\pi}{dt}} = 0 = \frac{d^2}{dt^2}\log d_B(t) = -2q/md(t)e^{-\pi(t)} \le 0$.

3. On $t \in (t_2^B, t_3^B)$, d(t) = K, thus $\pi(t) = \log d_B(t) - \log K$, which is unimodal. We show that if $\pi(t)$ is (already) decreasing at $t_2^{B^-}$, then it is also decreasing at $t_2^{B^+}$, so $\underline{\pi(t)}$ is unimodal on (t_1^B, t_3^B) .

If $\pi(t)$ is decreasing at $t_2^{B^-}$, we have $\frac{d\log d_B(t)}{dt} - \frac{d\log d_2(t^-)}{dt} \leq 0$, which is equivalent to $\frac{d\log d_B(t)}{dt} - \frac{d\log d_1(t)}{dt} + \frac{d\log d_1(t)}{dt} - \frac{d\log d_2(t^-)}{dt} = \frac{d\lambda_1(t)}{dt} + \frac{d\lambda_2(t^-)}{dt} \leq 0$, thus $\frac{d\lambda_1(t)}{dt} \leq -\frac{d\lambda_2(t^-)}{dt} \leq 0$. From the proof of Lemma 2, $\lambda_1(t)$ is decreasing only if $\log d_B(t)$ is decreasing. Thus $\pi(t_2^{B^+}) = \log d_B(t_2^{B}) - \log K$ is also decreasing at $t_2^{B^+}$.

4. On $t \in (t_3^B, T)$, $\log d(t) = \log d_1(t) = \log d_B(t) - 1 + \lambda_1(t) - c$, thus $\pi(t) = 1 + c - \lambda_1(t)$. As shown in the proof of Lemma 2, $\pi(t)$ is unimodal.

A.12 Proof of Lemma 10

Proof: 1. If I(t) > 0 on an open interval, applying similar argument as in the proof of Lemma 7, we have $x^*(t) = K$ almost surely.

- 2. If I(t)=W(t)=0 on an open interval, then $\frac{dI(t)}{dt}=x(t)-s(t)=0$ and $\frac{dW(t)}{dt}=\xi(d(t)-s(t))=0$. Therefore, $s^*(t)=x^*(t)=d^*(t)$ and part (c) holds.
- 3. If W(t) > 0 on an open interval and b > rc, then by equation (2.11), the Hamiltonian is linear in s(t). If (i) I(t) > 0, then $s^*(t) = 0$ or $s^*(t) = +\infty$. If (ii)

I(t) = 0, then s(t) = 0 or s(t) = x(t) = K.

- (i) Assume I(t) > 0. $s^*(t) = +\infty$ is not sustainable on a positive time interval. Thus, $s^*(t) = 0$ is the only possibility. Applying the same logic as in the proof of Lemma 7, we have x(t) = c. An alternative strategy is to sell $s^*(t) = \varepsilon$ and $x^*(t) = K$ on $(t_1, t_1 + \delta)$, where ε is a small positive number, while $x^*(t) = x^*(t) + \varepsilon$ for $(t_2, t_2 + \delta)$ where t_2 is the first $t_2 \geq t_1 + \delta$ such that $x(t_2) < c$. In this case, the firm saves backordering cost, inventory holding cost, and benefits from increased cumulative sales. Therefore $s^*(t) = 0$ cannot be optimal. Thus I(t) = 0 whenever W(t) > 0 on a positive interval (which completes part (a)).
- (ii) Let I(t) = 0. Note that I(t) = 0 and s(t) = 0 on $(t_1, t_1 + \delta)$, where δ is a small positive number, cannot be jointly optimal.

When I(t)=0, applying similar logic as in the proof of Lemma 7, we have x(t)=s(t). Assume x(t)< K. W,D,S,I can be approximated by $W(t+\delta)\approx W(t)+\xi\delta d(t)$, $D(t+\delta)\approx D(t)+\delta d(t)$, $S(t+\delta)=S(t)$ and $I(t+\delta)=I(t)$. Consider an alternative strategy to sell $x^*(t)=s^*(t)=x(t)+\varepsilon$ on $(t_1,t_1+\delta)$ where ε is a small positive number. In this case, W,D,S can be approximated by $W(t+\delta)\approx W(t)+\xi\delta d(t)-\varepsilon\delta$, $S(t+\delta)=S(t)+\varepsilon\delta$, $D(t+\delta)\approx D(t)+\delta d(t)$ and $I(t+\delta)=I(t)$. Since b>rc, the alternative strategy saves on backlogging cost, and increases cumulative sales $S(t+\delta)$. By equation (2.11), $\frac{d\lambda_2(t)}{dt}=r\lambda_2(t)-\frac{\partial H}{\partial S}=r\lambda_2(t)-\frac{q/m}{p+qS(t)/m}\xi d(t)$. Note that $\lambda_2(t)$ is continuous and $\lambda_2(T)=0$. If $\lambda_2(t)=0$ for some $t\in[0,T]$, then $\frac{d\lambda_2(t)}{dt}<0$, implying that there would be no other 0-points of $\lambda_2(t)$ and $\lambda_2(t)\geq 0$ for all $t\in[0,T]$. Thus the firm would also benefit from the increased cumulative sales, which implies

A.13 Compare myopic sales policy with delay sales policy

We first study the myopic sales policy. Under myopic sales policy, at time t=0, the cumulative demand is given by $D(t)=pm(e^{(p+q)t}-1)/(q+pe^{(p+q)t})$. We solve for the time at which inventory is depleted: $D(\tau_1)=c\tau_1$, and obtain $\tau_1=0.57$. The profit during the first phase $[0,\tau_1]$ is given by $\Pi_1^{myopic}=\int_0^{\tau_1}(ad(t)-h(ct-D(t)))e^{-rt}dt=73.918$. In the second phase, demand is constrained until $d(\tau_2)=c$. The cumulative demand $D(t)=m-(m-D_1)e^{(p+q/m(D_1-c\tau_1))}\tau_1+\frac{q}{2m}c\tau_1^2-(p+q/m(D_1-c\tau_1))t-\frac{q}{2m}ct^2)$ where $D_1=D(\tau_1)$. solving for τ_2 , we have $\tau_2=14.89$. The profit in the second phase $[\tau_1,\tau_2]$ is given by $\Pi_2^{myopic}=\int_{\tau_1}^{\tau_2}ace^{-rt}dt=1724.59$. Finally, demand drops below capacity, $D(t)=m/q(q-(p+q/m(S_2-D_2))e^{-(t+Constant3)(p+q/m(S_2-D_2)+q)})/(1+e^{-(t+Constant3)}(p+q/m(S_2-D_2)+q)))$, where $S_2=D_1+c(\tau_2-\tau_1)$, $D_2=D(\tau_2)$, and $Constant3=1/(p+q/m(S_2-D_2)+q)\log((p+q/m(S_2-D_2)+D_2q/m)/(q-D_2q/m))-\tau_2$. The profit during the third phase $[\tau_2,\infty)$ is given by $\Pi_3^{myopic}=\int_{\tau_2}^{\infty}ad(t)e^{-rt}dt=471.55$. The total profit is $\Pi^{myopic}=\Pi_1^{myopic}+\Pi_3^{myopic}=2270.07$.

Then we examine the profit if we intentionally delay sales. That is, for $t \in [0, \tau] = [0, 5]$, customers come to purchase but the firm does not sell to them even if inventory is available. The profit in the first phase [0, 5] is 0. In the second phase, the cumulative demand is given by $D(t) = m/q(q-(p+q/m(S_0-D_0))e^{-(t+Constant1)(p+q/m(S_0-D_0)+q)})/(1+e^{-(t+Constant1)(p+q/m(S_0-D_0)+q)})$, where $D_0 = m(1-e^{-p\tau})$, $S_0 = 0$, and $Constant1 = 1/(p+q/m(S_0-D_0)+q)\log((p+q/m(S_0-D_0)+D_0q/m)/(q-D_0q/m))-\tau$. The time that inventory is depleted $\tau_1 = 11.356$. The profit during the second phase $[\tau, \tau_1]$ is

 $\Pi_1^{delay}=1348.87$. Then demand is constrained and the time demand drops below capacity $\tau_2=16.84$. Profit during the third phase $[\tau_1,\tau_2]$ is $\Pi_2^{delay}=619.25$. Finally demand drops below capacity and the profit during this phase $[\tau_2,\infty]$ is $\Pi_3^{delay}=343.79$. Total profit is $\Pi^{delay}=\Pi_1^{delay}+\Pi_2^{delay}+\Pi_3^{delay}=2311.92\geq\Pi^{myopic}$. Thus myopic sales policy is dominated. \blacksquare

A.14 Lemma for the Proof of Theorem 3

Denote $\log d_1(t) = \log d_B(t) - 1 + \frac{1}{\xi} \lambda_1(t) + \lambda_4(t)$, $\log d_2(t) = \log d_B(t) + \lambda_1(t) + \lambda_2(t) - \lambda_3(t)$, $\log d_3(t) = \log d_B(t) + \lambda_1(t) + \lambda_2(t) - c$, and $\pi_i(t) = \log d_B(t) - \log d_i(t)$ for i = 1, 2, 3.

Lemma 20 If b > rc, then the optimal demand is given by

(a) If
$$W(t) > 0$$
 and $I(t) = 0$, then $\log d^*(t) = \log d_B(t) - 1 + \frac{1}{\xi} \lambda_1(t) + \lambda_4(t)$.

(b) If
$$W(t) = 0$$
 and $I(t) > 0$, then $\log d^*(t) = \log d_B(t) - 1 + \lambda_1(t) + \lambda_2(t) - \lambda_3(t)$.

(c) If
$$W(t) = 0$$
 and $I(t) = 0$, then $\log d^*(t) = \min(\log K, \log d_B(t) - 1 + \lambda_1(t) + \lambda_2(t) - c)$.

Proof: (a) If W(t) > 0 and I(t) = 0, by Lemma 10, $x^*(t) = s^*(t) = K$, therefore $\frac{dH}{dd} = \frac{\partial H}{\partial d} = \xi(\log d_B(t) - \log d(t) - 1) + \lambda_1(t) + \xi \lambda_4(t)$ by equation (2.11). Setting it to zero, we have $\log d^*(t) = \log d_B(t) - 1 + \frac{1}{\xi} \lambda_1(t) + \lambda_4(t)$.

(b) If
$$W(t) = 0$$
 and $I(t) > 0$, by Lemma 10, $x^*(t) = K$ and $s^*(t) = d^*(t)$, therefore
$$\frac{dH}{dd} = \frac{\partial H}{\partial d} + \frac{\partial H}{\partial s} = \log d_B(t) - \log d(t) - 1 + \lambda_1(t) + \lambda_2(t) - \lambda_3(t) \text{ by equation (2.11)}.$$
 Setting it to zero, we have $\log d^*(t) = \log d_B(t) - 1 + \lambda_1(t) + \lambda_2(t) - \lambda_3(t)$.

(c) If W(t) = 0 and I(t) = 0, by Lemma 10, $x^*(t) = s^*(t) = d^*(t)$, therefore $\frac{dH}{dd} = \frac{\partial H}{\partial d} + \frac{\partial H}{\partial s} + \frac{\partial H}{\partial x} = \log d_B(t) - \log d(t) - 1 + \lambda_1(t) + \lambda_2(t) - c$ by equation (2.11). Setting it to zero, we have $\log d^*(t) = \log d_B(t) - 1 + \lambda_1(t) + \lambda_2(t) - c$. Since $0 \le x(t) \le K$, $\log d^*(t) = \min\{\log K, \log d_B(t) - 1 + \lambda_1(t) + \lambda_2(t) - c\}$.

A.15 Lemma for the Proof of Theorem 3

Lemma 21 (a) $d_2(t)$ and $d_3(t)$ are unimodal on [0, T].

(b) If
$$\frac{d \log d_B(t)}{dt} \geq 0$$
 whenever $\frac{d \log d_1(t)}{dt} = 0$, then $d_1(t)$ is unimodal on $[0,T]$.

(c) $\pi_1(t)$, $\pi_2(t)$ and $\pi_3(t)$ are unimodal whenever they are optimal.

Proof:

1. $d_3(t)$ is unimodal: Since I(t) = W(t) = 0 and $d^*(t) = x^*(t) = s^*(t)$,

$$\frac{d\lambda_1(t)}{dt} = r\lambda_1(t) + \frac{d(t)}{m - D(t)}$$

$$\frac{d\lambda_2(t)}{dt} = r\lambda_2(t) - \frac{q/md(t)}{p + qS(t)/m}$$

$$\frac{d\log d_3(t)}{dt} = \frac{d\log d_B(t)}{dt} + \frac{d\lambda_1(t)}{dt} + \frac{d\lambda_2(t)}{dt}$$

$$= \frac{d\log d_B(t)}{dt} + r\lambda_1(t) + \frac{d(t)}{m - D(t)} + r\lambda_2(t) - \frac{q/md(t)}{p + qS(t)/m}$$

$$= r\lambda_1(t) + r\lambda_2(t)$$

Since $\frac{d(t)}{m-D(t)} - \frac{q/md(t)}{p+qS(t)/m}$ is positive-negative, and $\lambda_1(T) = \lambda_2(T) = 0$, by similar argument as in the proof of Lemma 1, there exists at most one $t \in [0,T)$ such that $\lambda_1(t) + \lambda_2(t) = 0$, and $\lambda_1(t) + \lambda_2(t)$ is positive on [0,t) and negative on (t,T). Therefore $d_3(t)$ is increasing on [0,t) and decreasing on (t,T).

2.
$$\underline{\pi_3(t)}$$
 is unimodal: $\underline{\pi_3(t)} = \log d_B(t) - \log d_3(t) = 1 + c - \lambda_1(t) - \lambda_2(t)$. If $\frac{d\pi_3(t)}{dt} = 0$ for some t , then $\frac{d\pi_3(t)}{dt} = -\frac{d(\lambda_1 + \lambda_2)}{dt} = -r(\lambda_1(t) + \lambda_2(t)) - \frac{d(t)}{m - D(t)} + \frac{d(t)}{m - D(t)}$

$$\frac{q/md(t)}{p+qS(t)/m} = -r(\lambda_1(t)+\lambda_2(t)) + (-(p+qS(t))/m + q/m(m-D(t))e^{-\pi_3(t)} = 0,$$
 thus

$$\frac{d^2\pi_3(t)}{dt^2} = -r\frac{d(\lambda_1 + \lambda_2)}{dt} + (-(p+qS(t))/m + q/m(m-D(t))e^{-\pi_3(t)}(-\frac{d\pi_3(t)}{dt}) + (-qs(t)/m - qd(t)/m)e^{-\pi_3(t)}$$

Since the first two terms of the last equation are both zero, we have $\frac{d^2\pi_3(t)}{dt^2} \leq 0$, therefore $\pi_3(t)$ is unimodal.

- 3. $\underline{d_2(t)}$ and $\underline{\pi_2(t)}$ are unimodal: Since I(t) > 0, W(t) = 0, $x^*(t) = K$, and $s^*(t) = d^*(t)$, the unimodality of $d_2(t)$ and $\underline{\pi_2(t)}$ can be shown with the same logic as in the proof of Theorem 2.
- 4. $\underline{d_1(t)}$ is unimodal under the condition of the theorem: If $\frac{d \log d_B(t)}{dt} \geq 0$ whenever $\frac{d \log d_1(t)}{dt} = 0$, note that s(t) = K when W(t) > 0, then we examine $\frac{d^2 \log d_1(t)}{dt^2}$ at $\frac{d \log d_1(t)}{dt} = 0$:

$$\frac{d\log d_{1}(t)}{dt} = \frac{d\log d_{B}(t)}{dt} + \frac{1}{\xi} \frac{d\lambda_{1}(t)}{dt} + \frac{d\lambda_{4}(t)}{dt}
= \frac{-d(t)}{m - D(t)} + \frac{qc/m}{p + qS(t)/m} + \frac{1}{\xi} r\lambda_{1}(t) + \frac{d(t)}{m - D(t)} + r\lambda_{4}(t) + b
= \frac{qc/m}{p + qS(t)/m} + \frac{1}{\xi} r\lambda_{1}(t) + r\lambda_{4}(t) + b$$

If
$$\frac{d\log d_1(t)}{dt} = 0$$
, then

$$\frac{d^{2} \log d_{1}(t)}{dt^{2}} = -\left(\frac{qc/m}{p+qS(t)/m}\right)^{2} + \frac{1}{\xi}r\frac{d\lambda_{1}(t)}{dt} + r\frac{d\lambda_{4}(t)}{dt}
= -\left(\frac{qc/m}{p+qS(t)/m}\right)^{2} + r\left(\frac{1}{\xi}r\lambda_{1}(t) + \frac{d(t)}{m-D(t)} + r\lambda_{4}(t) + b\right)
= -\left(\frac{qc/m}{p+qS(t)/m}\right)^{2} + r\left(\frac{d(t)}{m-D(t)} - \frac{qc/m}{p+qS(t)/m}\right)
= -\left(\frac{qc/m}{p+qS(t)/m}\right)^{2} - r\frac{d\log d_{B}(t)}{dt} \le 0$$

Therefore $\log d_1(t)$ is unimodal.

5.
$$\frac{\pi_{1}(t) \text{ is unimodal: } \pi_{1}(t) = \log d_{B}(t) - \log d_{1}(t) = 1 - \frac{1}{\xi}\lambda_{1}(t) - \lambda_{4}(t). \text{ Let } \lambda(t) = \frac{1}{\xi}\lambda_{1}(t) + \lambda_{4}(t). \text{ We show that } \frac{d^{2}\lambda(t)}{dt^{2}} \geq 0 \text{ whenever } \frac{d\lambda(t)}{dt} = 0.$$

$$\frac{d\lambda(t)}{dt} = \frac{1}{\xi}\frac{d\lambda_{1}(t)}{dt} + \frac{d\lambda_{4}(t)}{dt}$$

$$= r\left(\frac{1}{\xi}\lambda_{1}(t) + \lambda_{4}(t)\right) + \frac{d(t)}{m - D(t)} + b$$

$$= r\lambda(t) + (p + qS(t)/m)e^{-\pi(t)} + b$$

$$\frac{d^{2}\lambda(t)}{dt^{2}} = r\frac{d\lambda(t)}{dt} + (p + qs(t)/m)e^{-\pi(t)} + (p + qS(t)/m)e^{-\pi(t)} \left(-\frac{d\pi(t)}{dt}\right)$$
At
$$\frac{d\lambda(t)}{dt} = -\frac{d\pi(t)}{dt} = 0, \quad \frac{d^{2}\lambda(t)}{dt^{2}} = (p + qs(t)/m)e^{-\pi(t)} \geq 0. \text{ Therefore } \pi_{1}(t) \text{ is unimodal.} \blacksquare$$

A.16 Proof of Theorem 3

Proof: We first show that if W(t) = 0 and d(t) < K for an open interval, then s(t) = d(t) on this interval. Note that s(t) < d(t) for an open interval (t', t''), implies W(t) > 0 and, by Lemma 10, we have s(t) = K > d(t), resulting in a contradiction.

The optimal demand trajectory can be characterized as follows. Since W(0) = 0, if $d_3(0) \leq K$, then $d^*(t) = \min(d_3(t), K)$ until some t_1^C , when similar to the case in Theorem 2, demand switches to $d^*(t) = d_2(t)$. As we have shown in Lemma 10, W(t) = 0, when I(t) > 0, demand cannot switch to $d_1(t)$. Therefore $d^*(t) = d_2(t)$ until t_2^C , where $I(t_2^C) = 0$.

While I(t) = 0 and W(t) = 0, we have $d^*(t) = \min(d_3(t), K)$. If $d^*(t_2^C) = d_3(t_2^C) \le K$, since $d_3(t)$ is unimodal, it will continue to be smaller than K. We have shown in Theorem 2 that demand cannot switch back to $d_2(t)$ (i.e., I(t) > 0), and we have

also shown that s(t) = d(t) as long as W(t) = 0 and d(t) < K, therefore demand cannot switch to $d_1(t)$. Thus $d^*(t) = d_3(t)$ for $t \in (t_2^C, T)$.

If $d_3(t_2^C) > K$, then $d^*(t) = K$ until t_3^C , where demand may switch to $d_1(t)$. As shown in Lemma 21, $d_1(t)$ is unimodal. At t_4^C , $W(t_4^C) = 0$, and demand switches back to $\min(d_3(t), K)$. As backlogging is cleared, $d_3(t_4^C) \leq K$, and by the argument above, demand also cannot switch back to $d_1(t)$.

For the optimal price trajectory, Lemma 21 has shown that $\pi_1(t)$, $\pi_2(t)$ and $\pi_3(t)$ are unimodal. Therefore, $\pi^*(t)$ is unimodal on (t_3^C, t_4^C) and (t_4^C, T) . Using similar logic as in the proof of Theorem 2, we can show that $\pi^*(t)$ is also unimodal on $(0, t_3^C)$.

In a make-to-order environment, $h=+\infty$, therefore, it is not possible to carry inventory and $t_1^C=t_2^C$ in Theorem 3. Let $t_t^D=t_1^C=t_2^C$, $t_2^D=t_3^C$, and $t_3^D=t_4^C$, then the case for make-to-order environment and partial backlogging is proved.

APPENDIX B

Main Proofs in Chapter 3

B.1 Proof of Theorem 4

Proof: To prove the structure of optimal policy, we use value iteration.

Let $H^0(i, D, N) = 0$ for all (i, D, N). For any $k \ge 1$, define $L^{k-1}(i, N) = N *$ $W + W_{new} + c_R/\mu + \sum_{N'=0}^{\infty} \operatorname{Prob}(N') \sum_{j=N}^{N+N'-1} H^{k-1}(i, \bar{D}, j) \text{ and } K^{k-1}(i, Q, D, N) =$ $c_M Q + \sum_{N'=0}^{\infty} \operatorname{Prob}(N') \int_{aQ}^{D} H^{k-1}(i+1, D-y, N+N') dF(Q, y), \text{ then } H^k(i, D, N) =$ $L^{k-1}(i, N) + \min_{Q \in \Pi(D)} K^{k-1}(i, D, N, Q), \text{ and } H(i, D, N) = \lim_{k \to \infty} H^k(i, D, N).$

We prove the theorem by induction. The theorem holds obviously for H^0 , and Q^{*0} exists and increases in D and N. Assume it is true for H^{k-1} :

- 1. $H^{k-1}(i, D, N)$ is nonnegative;
- 2. $H^{k-1}(i, D, N)$ increases in D;
- 3. $H^{k-1}(i, D, N)$ increases in N.

We take first and second derivatives of K with respect to Q, and show the second derivative is positive under conditions (a) and (b), therefore there exists an optimal lot size Q.

$$K^{k-1}(i,Q,D,N) = c_M Q + \sum_{N'=0}^{\infty} \text{Prob}(N') \int_{aQ}^{D} H^{k-1}(i+1,D-y,N+N') dF(i,Q,y)$$
(B.1)

$$\begin{split} \frac{\partial K}{\partial Q} &= c_M + \sum_{N'=0}^{\infty} \operatorname{Prob}(N') [\int_{aQ}^{D} H^{k-1}(i+1,D-y,N+N') dF_Q'(i,Q,y) \\ &- H^{k-1}(i+1,D-aQ,N+N') F_y'(i,Q,aQ) a] \end{split} \tag{B.2}$$

$$&= c_M + \sum_{N'=0}^{\infty} \operatorname{Prob}(N') [H^{k-1}(i+1,D-y,N+N') F_Q'(i,Q,y)]_{aQ}^{D} \\ &+ \int_{aQ}^{D} H_D^{k-1'}(i+1,D-y,N+N') F_Q'(i,Q,y) dy \\ &- H^{k-1}(i+1,D-aQ,N+N') F_y'(i,Q,aQ) a] \end{aligned} \tag{B.3}$$

$$&= c_M + \sum_{N'=0}^{\infty} \operatorname{Prob}(N') [H^{k-1}(i+1,0,N+N') F_Q'(i,Q,D) \\ &- H^{k-1}(i+1,D-aQ,N+N') F_Q'(i,Q,aQ) \\ &+ \int_{aQ}^{D} H_D^{k-1'}(i+1,D-y,N+N') F_Q'(i,Q,y) dy \\ &- H^{k-1}(i+1,D-aQ,N+N') F_y'(i,Q,aQ) a] \end{aligned} \tag{B.4}$$

$$&= c_M + \sum_{N'=0}^{\infty} \operatorname{Prob}(N') [H^{k-1}(i+1,0,N+N') F_Q'(i,Q,y) dy \\ &- H^{k-1}(i+1,D-aQ,N+N') F_Q'(i,Q,y) dy \\ &- H^{k-1}(i+1,D-aQ,N+N') F_Q'(i,Q,y) dy \\ &- H^{k-1}(i+1,D-aQ,N+N') F_Q'(i,Q,y) dy \end{aligned}$$

Since F(i,Q,aQ) = 0 for any $i,Q, F'_Q(i,Q,aQ) + aF'_q(i,Q,aQ) = 0$, thus

$$\frac{\partial K^{k-1}}{\partial Q} = c_M + \sum_{N'=0}^{\infty} \text{Prob}(N')[H^{k-1}(i+1,0,N+N')F_Q'(i,Q,D)
+ \int_{aQ}^{D} H_D^{k-1'}(i+1,D-y,N+N')F_Q'(i,Q,y)dy]
\frac{\partial^2 K}{\partial Q^2} = \sum_{N'=0}^{\infty} \text{Prob}(N')[H^{k-1}(i+1,0,N+N')F_Q''(i,Q,D)
+ \int_{aQ}^{D} H_D^{k-1'}(i+1,D-y,N+N')F_Q''(i,Q,y)dy
-H_D^{k-1'}(i+1,D-aQ,N+N')F_Q'(i,Q,aQ)a]$$
(B.7)

Since H^{k-1} is nonnegative and increasing in D, and F is decreasing and convex on Q by condition (a), $\frac{\partial^2 K^{k-1}}{\partial Q^2} \geq 0$. So K is a convex function of Q and $K^{k-1} \to \infty$ as

 $Q \to \infty$, and thus there exists an optimal $Q^{*k-1}(i, D, N) \in \Omega(D)$.

We then show that the cross derivatives of K^{k-1} with respect to (Q, D) and (Q, N) are both negative under condition (a) and (b), and thus the optimal policy increases in D and N. From (B.2)

$$\frac{\partial K^{k-1}}{\partial Q} = c_M + \sum_{N'=0}^{\infty} \text{Prob}(N') \int_{aQ}^{D} H^{k-1}(i+1, D-y, N+N') F_{Qy}''(i, Q, y) dy
-H^{k-1}(i+1, D-aQ, N+N') F_y'(i, Q, aQ) a
(B.8)$$

$$\frac{\partial^2 K^{k-1}}{\partial Q \partial D} = \sum_{N'=0}^{\infty} \text{Prob}(N') \int_{aQ}^{D} H_D^{k-1'}(i+1, D-y, N+N') F_{Qy}''(i, Q, y) dy
+H(i+1, 0, N+N') F_{Qy}''(i, Q, D)$$

$$-H_D^{k-1'}(i+1, D-aQ, N+N') F_y'(i, Q, aQ) a \le 0$$
(B.9)
$$\frac{\partial^2 K^{k-1}}{\partial Q \partial N} = \sum_{N'=0}^{\infty} \text{Prob}(N') \int_{aQ}^{D} H_N^{k-1'}(i+1, D-y, N+N') F_{Qy}''(i, Q, y) dy
-H_N^{k-1'}(i+1, D-aQ, N+N') F_y'(i, Q, aQ) a \le 0$$
(B.10)

Therefore Q^* increases in N. For D, since $\Omega(D) = \left[\frac{D}{b}, \frac{D}{a}\right]$ is a convex set increasing in D, Q^* also increases in D.

Then we prove that the three hypotheses also hold for H^k . Clearly $H^k(i, D, N)$ is nonnegative for all (i, D, N).

 $L^{k-1}(N)$ increases in N. Since H^{k-1} is nonnegative and increasing in N, H^k is still increasing in N after taking expectation and minimization.

 $L^{k-1}(N)$ is constant relative to D. K^{k-1} is still increasing in D after taking expectation. Therefore $\min_{Q \in [0,+\infty)} K^{k-1}$ is still increasing in D. We will show $\min_{Q \in [\frac{D}{h},\frac{D}{a}]} K^{k-1}$ is also increasing in D.

Since any $Q > \frac{D}{a}$ is dominated by $\frac{D}{a}$, K^{k-1} will not take minimum at $\frac{D}{a}$. So we

restrict our attention to the case when Q is bounded below by $\frac{D}{h}$.

- 1. For $D_1 < D_2$, if both $K^{k-1}(Q, i, D_1, N)$ and $K^{k-1}(Q, i, D_2, N)$ are constrained by the lower bound, since K^{k-1} is convex in Q, then $K^{k-1}(Q, i, D_1, N)$ is increasing for any $Q \ge D_1$. Since $K^{k-1}(Q, i, D, N)$ is increasing in D, we have $K^{k-1}(D_1, i, D_1, N) \le K^{k-1}(D_2, i, D_1, N) \le K^{k-1}(D_2, i, D_2, N)$.
- 2. If only $K^{k-1}(Q, i, D_2, N)$ is constrained by the lower bound, then $K^{k-1}(Q^*(D_1), i, D_1, N) \le K^{k-1}(Q^*(D_2), i, D_2, N) \le K^{k-1}(D_2, i, D_2, N)$.
- 3. If only $K^{k-1}(Q, i, D_1, N)$ is constrained by the lower bound, as K^{k-1} is convex in $Q, K^{k-1}(Q, i, D_1, N)$ is increasing for any $Q \ge D_1$. Since $D_1 \le D_2 \le D_2/b \le Q^*(D_2)$ and $K^{k-1}(Q, i, D, N)$ is increasing in D, we have $K^{k-1}(D_1, i, D_1, N) \le K^{k-1}(Q^*(D_2), i, D_1, N) \le K^{k-1}(Q^*D_2, i, D_2, N)$.

Therefore $H^k(i, D, N)$ is increasing in D.

By induction a unique Q^* exists and increases in D and N.

B.2 Proof of Theorem 5

Proof: We prove the theorem by induction. $H^0 = 0$ is concave in D. Assuming that H^{k-1} is concave in D, we prove the results in three steps. First we show that if the optimal lot size satisfies the first-order condition (FOC), then H^k is also concave in D. Then we show Q^*/D is decreasing in D, therefore as D increases, Q^* changes from D/a to FOC solution to D. Finally we show that if $Q^* = D/a$ or $Q^* = D$, H^k is still concave, and the first order derivative of H^k at the transition points is continuous, therefore H^k is concave. In each step, we first prove for r being non-integer, and then

complete the proof for integer r. For the purpose of clarity we omit the production run number i here, and the proof will not change if i is added.

1. If the optimal Q^* satisfies FOC, and r is not an integer, then H^k is concave: Assume n < r < n+1, where n is an integer, $n \ge 1$.

$$H^{k}(D, N) = L^{k-1}(N) + \min_{Q \in \Pi(D)} K^{k-1}(Q, D, N)$$

$$K^{k-1}(Q, D, N) = c_{M}Q + E_{N'} \int_{0}^{D} H^{k-1}(D - y, N + N') \frac{ry^{r-1}}{Q^{r}} dy$$

The first-order condition is:

$$\frac{\partial K^{k-1}}{\partial Q} = c_M - \frac{r}{Q^{r+1}} E_{N'} \int_0^D H^{k-1}(D - y, N + N') r y^{r-1} dy = 0$$

Thus the optimal quantity is

$$Q^* = \left(\frac{r^2}{c_M} E_{N'} \int_0^D H^{k-1}(D - y, N + N') r y^{r-1} dy\right)^{\frac{1}{r+1}}$$

And the optimal cost difference function H^k is

$$H^{k^*} = L^{k-1}(N) + (c_M(r^2 + r^{-2})E_{N'} \int_0^D H^{k-1}(D - y, N + N')ry^{r-1}dy)^{\frac{1}{r+1}}$$

Denote $P_r = E_{N'} \int_0^D H^{k-1}(D-y,N+N')y^r dy$, we will show in the following that $(P_{r-1})^{r+1}$ is concave, and thus H^k is concave in D.

Notice that for any non-integer r, P_r has the following properties: (a) $P_r \geq 0$,

(b) $\lim_{D\to 0} P_r = 0$, and $(c)\frac{\partial P_r}{\partial D} = rP_{r-1}$. Property (c) is obtained in the following way:

$$\frac{\partial P_r}{\partial D} = \frac{\partial}{\partial D} (E_{N'} \int_0^D H^{k-1}(y, N + N')(D - y)^r dy)
= E_{N'} \int_0^D H^{k-1}(y, N + N')r(D - y)^{r-1} dy
= E_{N'} \int_0^D H^{k-1}(D - y, N + N')ry^{r-1} dy = rP_{r-1}$$

Take the second derivative of $(P_{r-1})^{\frac{1}{r+1}}$:

$$\frac{\partial}{\partial D}(P_{r-1})\frac{1}{r+1} = \frac{1}{r+1}(P_{r-1})\frac{-r}{r+1}(r-1)P_{r-2}$$

$$\frac{\partial^2}{\partial D^2}(P_{r-1})\frac{1}{r+1} = \frac{r-1}{r+1}\left[-\frac{r}{r+1}(P_{r-1})^{-\frac{r}{r+1}-1}P_{r-2}(r-1)P_{r-2}\right]$$

$$+(P_{r-1})^{-\frac{r}{r+1}}(r-2)P_{r-3}$$

$$= \frac{r-1}{r+1}\left[\frac{r}{r+1}(P_{r-1})^{-\frac{r}{r+1}-1}\right]$$

$$*[(r-2)(r-1)P_{r-1}P_{r-3} - r(r-1)P_{r-2}P_{r-2}]$$

The first two terms in the above equation is nonnegative. Denote the third term by $S^{(0)}=(r-2)(r-1)P_{r-1}P_{r-3}-r(r-1)P_{r-2}P_{r-2}$, and we will show $S^{(0)}\leq 0$. If 1< r<2, then clearly $S^{(0)}\leq 0$;

If 2 < r < 3, then $\lim_{D \to 0} S^{(0)} = 0$, and

$$\frac{\partial S^{(0)}}{\partial D} = (r+1)(r-2)(r-3)P_{r-1}P_{r-4} + (r-2)(r+1)(r-1)P_{r-2}P_{r-3}$$
$$-2r(r-1)(r-2)P_{r-2}P_{r-3}$$
$$= (r+1)(r-2)(r-3)P_{r-1}P_{r-4} - (r-2)(r+2)(r-1)P_{r-2}P_{r-3} \le 0$$

In general, for n < r < n+1, the (n-1)th derivative of $S^{(0)}$ is:

$$\frac{\partial^{(n-1)}S^{(0)}}{\partial D^{(n-1)}} = x_1 P_{r-n-2} P_{r-1} + S^{(1)}$$

where $S^{(1)}$ is a linear combination of terms: $P_{r-n-1}Pr-2$, $P_{r-n}Pr-3$, $\lim_{D\to 0} S^{(1)} = 0$, and $x_1 = (r+1)(r-2) \prod_{i=3}^{n+1} (r-i) < 0$. If $S^{(1)} \leq 0$, then $\frac{\partial^{(n-1)}S^{(0)}}{\partial D^{(n-1)}} \leq 0$. Since for any $i \leq n-1$, $\lim_{D\to 0} \frac{\partial^{(i)}S^{(0)}}{\partial D^{(i)}} = 0$, we have $\frac{\partial^{(i)}S^{(0)}}{\partial D^{(i)}} \leq 0$ and $S^{(0)} \leq 0$. To show that $S^{(1)} \leq 0$, we take first derivative again over $S^{(1)}$: $\frac{\partial S^{(1)}}{\partial D} = x_2 P_{r-2} P_{r-n-2} + S^{(3)}$. Iteratively we take $\frac{\partial S^{(k-1)}}{\partial D} = x_k P_{r-k} P_{r-n-2} + S^{(k)}$ for any $k \leq n$. Then we can show that $x_k \leq 0$ for any $k \leq n$:

$$x_{k} = (r+1)(r-2)(C_{k+n-2}^{k-1} + C_{k+n-2}^{k-3}) \prod_{i=1}^{k-1} (r-i) \prod_{i=3}^{n+1} (r-i)$$

$$-r(r-1)2C_{k+n-2}^{k-2} \prod_{i=2}^{k-1} (r-i) \prod_{i=2}^{n+1} (r-i)$$

$$= \prod_{i=1}^{k-1} (r-i) \prod_{i=2}^{n+1} (r-i)((r+1)(C_{k+n-2}^{k-1} + C_{k+n-2}^{k-3}) - 2rC_{k+n-2}^{k-2})$$

$$= \prod_{i=1}^{k-1} (r-i) \prod_{i=2}^{n+1} (r-i)((r+1)(\frac{(k+n-2)!}{(k-1)!(n-1)!} + \frac{(k+n-2)!}{(k-3)!(n+1)!}) - 2r\frac{(k+n-2)!}{(k-2)!n!})$$

$$= \prod_{i=1}^{k-1} (r-i) \prod_{i=2}^{n+1} (r-i) \frac{(k+n-2)!}{(k-1)!(n+1)!}$$

$$((r+1)(n(n+1) + (k-1)(k-2)) - 2r(k-1)(n+1))$$

For r = n, $((r+1)(n(n+1) + (k-1)(k-2)) - 2r(k-1)(n+1)) = (n+1)(n+1-k)(n+2-k) \ge 0$ for $k \le n$.

For r = n + 1,

$$((r+1)(n(n+1)+(k-1)(k-2))-2r(k-1)(n+1))$$

$$= ((n+2)(n(n+1)+(k-1)(k-2))-2(n+1)(k-1)(n+1))$$

$$= (n+2)(k-1)^2 - (2(n+1)^2 + (n+2))(k-1) + n(n+1)(n+2)$$

The above expression can be viewed as a function of k. At k = n, the value is $3 \ge 0$. The first derivative at k = n is -3n - 8 < 0, therefore it is positive for all $k \le n$.

Since ((r+1)(n(n+1)+(k-1)(k-2))-2r(k-1)(n+1)) is a linear function of r and is positive at both r=n and r=n+1, it is positive. Since r-n-1<0,

we conclude that $x_k \leq 0$.

For k = n, $\frac{\partial S^{(n-1)}}{\partial D} = x_n P_{r-n} P_{r-n-2} + x_{n+1} P_{r-n-1} P_{r-n-1}$. We have shown that $x_n \leq 0$. Now we show that $x_{n+1} \leq 0$.

$$x_{n+1} = (r+1)(r-2)C_{2n-2}^{n} \prod_{i=1}^{n} (r-i) \prod_{i=3}^{n} (r-i) - r(r-1)C_{2n-2}^{n-1} \prod_{i=2}^{n} (r-i) \prod_{i=2}^{n} (r-i) \prod_{i=1}^{n} (r-i) \prod_{i=1}^{n} (r-i) \prod_{i=1}^{n} (r-i) \prod_{i=1}^{n} (r-i) \prod_{i=1}^{n} (r-i) \prod_{i=1}^{n} (r-i) \frac{(2n-2)!}{n!(n-1)!} ((r+1)(n-1) - rn)$$

$$= \prod_{i=2}^{n} (r-i) \prod_{i=1}^{n} (r-i) \frac{(2n-2)!}{n!(n-1)!} (n-1-r) < 0$$

So $\lim_{D\to 0} S^{(n-1)} = 0$ and $\frac{\partial S^{(n-1)}}{\partial D} \leq 0$, thus $S^{(n-1)} \leq 0$. Similarly, for any k if $S^{(k+1)} \leq 0$, then $\frac{\partial S^{(k)}}{\partial D} = x_{k+1} P_{r-k} P_{r-n-2} + S^{(k+1)} \leq 0$. Since $\lim_{D\to 0} S^{(k)} = 0$, we have $S^{(k)} \leq 0$. By induction $S^{(0)} \leq 0$. Therefore $\frac{\partial^2 V^k}{\partial D^2} \leq 0$ and V^k is concave in D.

2. If the optimal Q^* satisfies FOC, and r is an integer, then H^k is concave:

Define $P_{-1} = E_{N'}H^{k-1}(D, N + N')$, $P_{-2} = E_{N'}H_D^{k-1}(D, N + N')$, and $P_{-3} = E_{N'}H_D^{k-1}(D, N + N')$, then the following properties of P_r are different:

- (a) $\lim_{D\to 0} P_r = 0$ for $r \ge 0$.
- (b) $P_r \ge 0$ for $r \ge -2$, and $P_{-3} \le 0$ by induction hypothesis.

(c)
$$\frac{\partial P_r}{\partial D} = rP_{r-1}$$
 for $r >= 1$, and $\frac{\partial P_r}{\partial D} = P_{r-1}$ for $r = 0, -1, -2$.

And the proof needs to be modified in the following way:

$$\frac{\partial^{(n-1)}S^{(0)}}{\partial D^{(n-1)}} = x_1 P_{-3} P_{r-1} + S^{(1)}$$

where $x_1 = (r+1)(r-2) \prod_{i=1}^r (r-i) \ge 0, P_{-3} \le 0$. For any $k \le n$, $\frac{\partial S^{(k-1)}}{\partial D} =$

 $x_k P_{r-k} P_{-3} + S^{(k)}$, where

$$x_{k} = (r+1)(r-2)(C_{k+r-2}^{k-1} + C_{k+r-2}^{k-3}) \prod_{i=1}^{k-1} (r-i) \prod_{i=3}^{r+1} (r-i)$$

$$-r(r-1)2C_{k+r-2}^{k-2} \prod_{i=2}^{k-1} (r-i) \prod_{i=2}^{r+1} (r-i)$$

$$= \prod_{i=1}^{k-1} (r-i) \prod_{i=2}^{r+1} (r-i) \frac{(k+r-2)!}{(k-1)!(r+1)!} (r+1)(r+1-k)(r+2-k) \ge 0$$

For $k \leq r$ we have $\lim_{D\to 0} S^{(k)} = 0$.

$$\frac{\partial S^{(r+1)}}{\partial D} = x_{r+1}P_0P_{-2} + x_{r+2}P_{-1}P_{-1}$$
, where

$$x_{r+2} = (r+1)(r-2)C_{2r-2}^{r} \prod_{i=1}^{r-1} (r-i) \prod_{i=3}^{r-1} (r-i) - r(r-1)C_{2r-2}^{r-1} \prod_{i=2}^{r-1} (r-i) \prod_{i=2}^{r-1} (r-i) \prod_{i=2}^{r-1} (r-i) \prod_{i=1}^{r-1} (r-i) \prod_{i=1}^{r-1} (r-i) (r-i) \prod_{i=1}^{r-1} (r-i) (r-i) \prod_{i=1}^{r-1} (r-i) \prod_{i=1}^{r-1} (r-i) \frac{(2r-2)!}{r!(r-1)!} (r-1) (r-1) - r^2 = 0$$

So
$$\lim_{D\to 0} S^{(r+1)} \leq 0$$
, and $\frac{\partial S^{(r+1)}}{\partial D} = x_{r+3} P_{-1} P_{-2}$, where

$$x_{r+3} = (r+1)(r-2)(C_{2r-1}^r + C_{2r-1}^{r-2}) \prod_{i=1}^{r-1} (r-i) \prod_{i=3}^{r-1} (r-i)$$

$$-r(r-1)2C_{2r-1}^{r-1} \prod_{i=2}^{r-1} (r-i) \prod_{i=2}^{r-1} (r-i)$$

$$= \prod_{i=2}^{r-1} (r-i) \prod_{i=1}^{r-1} (r-i) \frac{(2r-1)!}{(r+1)!(r-1)!} ((r+1)(r+1+r-1) - 2r(r+1)) = 0$$

So $S^{(r+1)} \leq 0$. By similar argument $\frac{\partial^2 V^k}{\partial D^2} \leq 0$ and V^k is concave in D.

3. If r is not an integer, then Q^*/D decreasing in D:

We show Q^*/D is decreasing in D by showing $\frac{\partial K^{k-1}(Q, D, N)}{\partial D}|_{Q=u*D}$ is increasing in D for any constant u. If so then for $D_1 < D_2$, and $Q^*(D_1) = D_1$, we have $\frac{\partial K^{k-1}(Q, D_1, N)}{\partial D}|_{Q=D_1} > 0$, and thus $\frac{\partial K^{k-1}(Q, D_2, N)}{\partial D}|_{Q=D_2} > 0$, and $Q^*(D_2) = D_2$.

Similarly, if $Q^*(D_1) = u_1D_1 > D_1$, then $\frac{\partial K^{k-1}(Q, D_2, N)}{\partial D}|_{Q=u_1D_2} \ge \frac{\partial K^{k-1}(Q, D_1, N)}{\partial D}|_{Q=u_1D_1} = 0$, and $Q^*(D_2) \le u_1D_2$. Therefore $\frac{Q^*(D_2)}{D_2} \le \frac{Q^*(D_1)}{D_1}$.

If $Q^*(D_1) = D_1/a$, then $\frac{Q^*(D_2)}{D_2} \le \frac{Q^*(D_1)}{D_1}$ automatically holds.

Thus we need to show $\frac{\partial K^{k-1}(Q, D, N)}{\partial D}|_{Q=u*D}$ is increasing in D for any constant u.

$$\frac{\partial K^{k-1}}{\partial Q} = c_M - \frac{r^2}{Q^{r+1}} P_{r-1}$$

$$\frac{\partial K^{k-1}}{\partial Q}|_{Q=uD} = c_M - \frac{r^2}{u^{r+1}D^{r+1}} P_{r-1}$$

We will show that $\frac{P_{r-1}}{D^{r+1}}$ decreases in D.

$$\frac{\partial}{\partial D} \frac{P_{r-1}}{D^{r+1}} = -(r+1) \frac{P_{r-1}}{D^{r+2}} + (r-1) \frac{P_{r-2}}{D^{r+1}}$$
$$= \frac{1}{D^{r+2}} ((r-1)DP_{r-2} - (r+1)P_{r-1})$$

Let $S^{(0)}=(r-1)DP_{r-2}-(r+1)P_{r-1}$, we will show $S^{(0)}\leq 0$ using similar techniques used in the first step. For n< r< n+1, $\frac{\partial^{(n)}S^{(0)}}{\partial D^{(n)}}=x_1DP_{r-2-n}+x_2P_{r-1-n}$, where $x_1=\prod_{i=1}^{n+1}(r-i)\leq 0$, and $x_2=\prod_{i=1}^{n}(r-i)-(r+1)\prod_{i=1}^{n}(r-i)=-r\prod_{i=1}^{n}(r-i)\leq 0$. Furthermore, $\lim_{D\to 0}\frac{\partial^{(k)}S^{(0)}}{\partial D^{(k)}}=0$ for all $k\leq n-1$. So $S^{(0)}\leq 0$ and $\frac{\partial K^{k-1}}{\partial D}(Q,D,N)|_{Q=D}$ is increasing in D.

4. If r is an integer, then Q^*/D decreasing in D:

The proof above needs to be modified as follows: for r = 1,

$$\frac{\partial}{\partial D} \frac{P_{r-1}}{D^{r+1}} = -(r+1) \frac{P_0}{D^{r+2}} + \frac{P_{-1}}{D^{r+1}} = \frac{1}{D^{r+2}} ((DP_{-1} - (r+1)P_0))$$

Let $S^{(0)} = DP_{-1} - (r+1)P_0$, then $\lim_{D\to 0} S^{(0)} \le 0$.

$$\begin{split} \frac{\partial S^{(0)}}{\partial D} &= DP_{-2} - P_{-1}, \text{ and } \lim_{D \to 0} \frac{\partial S^{(0)}}{\partial D} \leq 0. \\ \frac{\partial^2 S^{(0)}}{\partial D^2} &= DP_{-3} - P_{-2} + P_{-2} = DP_{-3} \leq 0. \text{ Thus } S^{(0)} \leq 0. \\ \text{For } r \geq 2, \ S^{(0)} &= (r-1)DP_{r-2} - (r+1)P_{r-1}. \\ \frac{\partial^{(r+1)}S^{(0)}}{\partial D^{(r+1)}} &= \prod_{i=1}^{r-1} (r-i)DP_{-3} + r \prod_{i=1}^{r-1} (r-i)P_{-2} \leq 0, \text{ and } \lim_{D \to 0} \frac{\partial^{(k)}S^{(0)}}{\partial D^{(k)}} \leq 0 \text{ for all } k \leq n. \text{ So } S^{(0)} \leq 0 \text{ and } \frac{\partial K^{k-1}}{\partial D}(Q, D, N)|_{Q=D} \text{ is increasing in } D. \end{split}$$

5. If r is not an integer, and if $Q^* = D$ or $Q^* = D/a$, then H^k is concave:

If
$$Q^* = D/a$$
, then $H^k = L^{k-1}(N) + c_M D/a$, which is clearly concave in D .

If
$$Q^* = D$$
, then $H^k = L^{k-1}(N) + c_M D + \frac{r}{D^r} E_{N'} \int_0^D H^{k-1}(D-y, N+N') y^{r-1} dy$.

It is sufficient to show $\frac{P_{r-1}}{D^r}$ is concave in D.

For n < r < n + 1,

$$\frac{\partial}{\partial D} \frac{P_{r-1}}{D^r} = c_M + \frac{1}{D^{r+1}} ((r-1)DP_{r-2} - rP_{r-1})$$

$$\frac{\partial^2}{\partial D^2} \frac{P_{r-1}}{D^r} = \frac{1}{D^{r+2}} ((r-1)(r-2)D^2P_{r-3} - (2r+1)(r-1)DP_{r-2} + r(r+1)P_{r-1})$$

Let
$$S^{(0)} = (r-1)(r-2)D^2P_{r-3} - (2r+1)(r-1)DP_{r-2} + r(r+1)P_{r-1}$$
, we will

show $S^{(0)} \leq 0$.

$$\frac{\partial^{(n-1)}S^{(0)}}{\partial D^{(n-1)}} = x_1 D^2 P_{r-2-n} + x_2 D P_{r-1-n} + x_3 P_{r-n}, \text{ where }$$

$$x_{1} = \prod_{i=1}^{n+1} (r-i) \le 0$$

$$x_{2} = 2(r-1)(r-2) \prod_{i=3}^{n} (r-i)C_{n-1}^{1} - (2r+1)(r-1) \prod_{i=2}^{n} (r-i)$$

$$= \prod_{i=1}^{n} (r-i)(2n-3-2r) \le 0$$

$${n-1}$$

$$n-1$$

$$x_3 = -(2r+1)(r-1) \prod_{i=2}^{n-1} (r-i)C_{n-1}^1 + r(r+1) \prod_{i=1}^{n-1} (r-i)$$
$$= \prod_{i=1}^{n-1} (r-i)(-(n+1)(2r+1) + r + 1) \le 0$$

6. If r is an integer, and if $Q^* = D$ or $Q^* = D/a$, then H^k is concave:

$$\frac{\partial^{(r)} S^{(0)}}{\partial D^{(r)}} = x_1 D^2 P_{-3} + x_2 D P_{-2} + x_3 P_{-1}, \text{ where}$$

$$x_1 = \prod_{i=1}^{r-1} (r-i) \ge 0, P_{-3} \le 0$$

$$x_2 = 2(r-1)(r-2) \prod_{i=3}^{r-1} (r-i) C_r^1 - (2r+1)(r-1) \prod_{i=2}^{r-1} (r-i)$$

$$= \prod_{i=1}^{r-1} (r-i)(2r-1-2r) \le 0$$

$$x_3 = -(2r+1)(r-1) \prod_{i=2}^{r-1} (r-i) C_r^1 + r(r+1) \prod_{i=1}^{r-1} (r-i)$$

$$= \prod_{i=1}^{r-1} (r-i)(-r) \le 0$$

7. The first derivative of K^k is continuous:

If there exists \underline{D} such that $Q^*(D) = D/a$ for $D \leq \underline{D}$, and $Q^*(D)$ satisfies the first-order condition for $D \geq \underline{D}$, $\frac{\partial H^k}{\partial D}|_{\underline{D}_-} = \frac{\partial K^k}{\partial Q}|_{Q=\underline{D}} + \frac{\partial K^k}{\partial D}|_{Q=\underline{D}} = \frac{\partial K^k}{\partial D}|_{Q=\underline{D}} = \frac{\partial K^k}{\partial D}|_{Q=\underline{D}} = 0.$

Similarly, if there exists \overline{D} such that $Q^*(D) = D$ for $D \ge \overline{D}$, and $Q^*(D)$ satisfies the first-order condition for $D \le \overline{D}$, then $\frac{\partial H^k}{\partial D}|_{\overline{D}_-} = \frac{\partial H^k}{\partial D}|_{\overline{D}_+}$.

By steps 1-7 the theorem is proved.

B.3 Proof of Theorem 6

Proof: If a = 0,

$$\begin{split} H(i,D,N) + g/\mu &= L^{k-1}(i,N) + \min_{Q \in \Omega(D)} K^{k-1}(Q,i,D,N) \\ &= N*W + W_{new} + \sum_{N'=0}^{\infty} \operatorname{Prob}(N') \sum_{j=N}^{N+N'-1} H^{k-1}(1,\bar{D},j) \\ &+ \min_{Q \in \Omega(D)} \{c_M Q + \sum_{N'=0}^{\infty} \operatorname{Prob}(N') \int_0^D H^{k-1}(i+1,D-y,N+N') dF(i,Q,y)\} \end{split}$$

 $H^0 = 0$ is constant relative to c_W, λ, T and Y. Assume H^{k-1} increases in c_W, λ , stochastically increases in T, and stochastically decreases in Y.

$$\frac{\partial K^{k-1}}{\partial Q} = c_M + \sum_{N'=0}^{\infty} \text{Prob}(N') \int_0^D H^{k-1}(i, D - y, N + N') dF_Q'(i, Q, y) (B.11)$$

Note that $dF'_Q(i, Q, y) = F''_{Qy}(i, Q, y)dy \le 0$.

- 1. If c_W increases: H^{k-1} is larger, $\int_0^D H^{k-1}(i, D-y, N+N') dF_Q'(i, Q, y)$ is smaller, and thus $\frac{\partial K^{k-1}}{\partial Q}$ is smaller.
- 2. If λ increases: then H^{k-1} is larger, and N' is stochastically larger. As $\int_0^D H^{k-1}(i, D y, N+N')dF_Q'(i, Q, y)$ is an decreasing function of N', $\sum_{N'=0}^{\infty} \operatorname{Prob}(N') \int_0^D H^{k-1}(i, D y, N+N')dF_Q'(i, Q, y)$ is smaller. Therefore $\frac{\partial K^{k-1}}{\partial Q}$ is smaller.
- 3. If production time T is stochastically larger, then H^{k-1} is larger, W and W_{new} are larger, and N' is stochastically larger. As $\int_0^D H^{k-1}(i, D-y, N+N')dF_Q'(i,Q,y)$ is an decreasing function of N', $\sum_{N'=0}^{\infty} \operatorname{Prob}(N') \int_0^D H^{k-1}(i,D-y,N+N')dF_Q'(i,Q,y)$ is smaller. $\frac{\partial K^{k-1}}{\partial Q}$ is smaller.
- 4. If the yield Y is stochastically smaller, then H^{k-1} is larger. If $F''_{Qy}(i,Q,y)$ is smaller, thus $\int_0^D H^{k-1}(i,D-y,N+N')dF'_Q(i,Q,y)=\int_0^D H^{k-1}(i,D-y,N+N')F''_{Qy}(i,Q,y)dy$ is smaller. Thus $\frac{\partial K^{k-1}}{\partial Q}$ is smaller.

Since $K^{k-1}(Q, i, D, N)$ is convex in Q, if $\frac{\partial K^{k-1}}{\partial Q}$ decreases, then the optimal Q^{*k-1} increases.

We then prove H^k also increases if c_W , λ increases, or if the time for each production run T stochastically increases, or if the yield Y stochastically decreases.

- 1. If c_W increases: then H^{k-1} is larger, W and W_{new} are larger, thus H^k is larger.
- 2. If λ increases: then H^{k-1} is larger, W_{new} is larger, and N' is stochastically

larger. As $\int_0^D H^{k-1}(i, D-y, N+N') dF(i, Q, y)$ is an increasing function of N', thus $\sum_{N'=0}^{\infty} \operatorname{Prob}(N') \int_0^D H^{k-1}(i, D-y, N+N') dF(i, Q, y)$ is larger. Therefore H^k is larger.

- 3. If production time T is stochastically larger, then H^{k-1} is larger, W and W_{new} are larger, and N' is stochastically larger. As $\int_0^D H^{k-1}(i, D y, N + N')dF(i, Q, y)$ is an increasing function of N', thus $\sum_{N'=0}^{\infty} \operatorname{Prob}(N') \int_0^D H^{k-1}(i, D y, N + N')dF(i, Q, y)$ is larger. Therefore H^k is larger.
- 4. If the yield Y is stochastically smaller, H^{k-1} is larger. As $H^{k-1}(i, D-y, N+N')$ is an decreasing function of y, $\int_0^D H^{k-1}(i, D-y, N+N')dF(i, Q, y)$ is larger. Therefore H^k is larger.

Therefore by induction Q^* increases.

B.4 Proof of Theorem 7

Proof: Let $\tilde{H}(i, D - y, N + N') = \min[H(i + 1, D - y, N + N'), C(D - y)]$. Let $H^0(i, D, N) = 0$ for all (i, D, N). Assume it is true for H^{k-1} and each i:

- 1. $H^{k-1}(i, D, N)$ is nonnegative;
- 2. $H^{k-1}(i, D, N)$ increases in D;
- 3. $H^{k-1}(i, D, N)$ increases in N.

Then \tilde{H}^{k-1} is nonnegative and increases in D and N for each i. Follow similar steps as in the proof of Theorem 4, we can show that a unique optimal Q_i^* exists and increases in D, N.

Since H(i+1, D-y, N+N') is increasing in N, and C(D-y) is constant with respect to N, then there exists a threshold $\bar{N}(D-y)$, such that the firm outsources

if and only if $N + N' > \bar{N}(D - y)$.

B.5 Proof of Theorem 8

Proof: Define $\tilde{H}(i, D-y, N+N') = \min[H(i+1, D-y, N+N'), A_S + c_S(D-y)],$ then \tilde{H} is nonnegative and increasing in D. If the yield is stochastically proportional power distribution, follow similar steps in the proof of Theorem 5, we can show that \tilde{H} is increasing and concave in D.

Since $C_S(D-y)=A_S+c_S(D-y)$ is linearly increasing in D-y, let $\bar{A}=H(i+1,0,N+N')$, then if $A_S\leq \bar{A}$, H(i+1,D-y,N+N') and $A_S+c_S(D-y)$ interact at most once (as a function of D-y), and thus there exists a threshold \bar{D} such that subcontracting achieves minimum if and only if $D-y\leq \bar{D}$.

Furthermore, when the queue length N increases, H(i+1, D-y, N+N') increases, while $C_S(D-y) = A_S + C_S(D-y)$ remains constant. Therefore the threshold increases with larger N.

If $A > \bar{A}$, then H(i+1, D-y, N+N') and $A_S + c_S(D-y)$ may interact at most twice (as a function of D-y), and thus there exist two thresholds $\underline{D} \leq \bar{D}$ such that subcontracting achieves minimum if and only if $\underline{D} \leq D-y \leq \bar{D}$.

Furthermore, when the queue length N increases, H(i+1, D-y, N+N') increases, while $C_S(D-y) = A_S + C_S(D-y)$ remains constant. Therefore the threshold \underline{D} decreases and \overline{D} increases with larger N.

Since the minimum of a linear function and a concave function function is still concave, we have $\tilde{H}(i, D-y, N+N')$ is still concave, which completes the induction.

B.6 Proof of Theorem 9

Proof: For each N, we need to compare H(i, D, N) and $C_O(D)$ to make the overtime decisions. Note that $\lim_{D\to 0} H(i, D, N) = N * W + W_{new} + c_R/\mu$, and $\lim_{D\to 0} C_O(D) = c_O/\mu$. Let $\underline{c}_O = \mu(N*W+W_{new}+c_R/\mu)$. Also note that $\frac{\partial H}{\partial D} \geq \frac{\partial C_O}{\partial D}$.

Then if $c_O \leq \underline{c}_O$, H(i, D, N) and $C_O(D)$ never intersect as a function of D. Thus the firm will run overtime production for any D.

If $c_O \geq \underline{c}_O$, then it is possible that H(i, D, N) and $C_O(D)$ intersect. Thus the subcontracting policy takes a threshold structure.

However, let $\bar{c}_O = \arg\max\{c_O|H(i,\bar{D},N) = C_O(\bar{D})\}$. Then if $c_O \geq \bar{c}_O$, H(i,D,N) always takes minimum for any $D \in (0,\bar{D}]$, thus the firm should not run overtime production for any demand.

APPENDIX C

Main Proofs in Chapter 4

C.1 Appendix

C.1.1 Proof of Lemma 1

Proof: Given wholesale price ω_2 and demand Q_2 , Manufacturer chooses $q_2^M \in [0, Q_2]$ to maximize profit:

$$R_2^M = (P_2 - c_2^M)q_2^M + (P_2 - \omega_2)(Q_2 - q_2^M) = (P_2 - \omega_2)Q_2 + (\omega_2 - c_2^M)q_2^M$$

So Manufacturer produces $q_2^M=Q_2$ if $\omega_2>c_2^M$, and $q_2^M=0$ if $\omega_2\leq c_2^M$. Then Supplier's expected profit is

$$R_2^S = E_{Q_2}(\omega_2 - c_2^S)Q_2 1_{\{\omega_2 \le c_2^M\}} = (\omega_2 - c_2^S)\bar{Q}_2 1_{\{\omega_2 \le c_2^M\}}$$

Therefore Supplier charges $\omega_2 = c_2^M$ and produces all if he can make positive profit, i.e., if $c_2^S \leq \omega_2 = c_2^M$, otherwise he will not produce anything. The outcome is as described in the lemma.

C.1.2 Proof of Lemma 2

Proof: First we examine Manufacturer's effort decision. If his production quantity exceeds the minimum requirement, he chooses an effort level to maximize his expected

second-period profit subtracted by effort cost.

$$\begin{aligned} & \max_{\epsilon^{M}} \{ -\epsilon^{M} + E_{c_{2}^{M}}(P_{2} - c_{2}^{M})\bar{Q}_{2} \} = \max_{\epsilon^{M}} \{ -\epsilon^{M} + E_{\delta^{M}}(P_{2} - c_{1}^{M} + g^{M}(\epsilon^{M}, \delta^{M}))\bar{Q}_{2} \} \\ & = \max_{\epsilon^{M}} \{ \epsilon^{M} + E_{\delta^{M}}g^{M}(\epsilon^{M}, \delta^{M})\bar{Q}_{2} \} + (P_{2} - c_{1}^{M})\bar{Q}_{2} \\ & = \max_{\epsilon^{M}} \{ \epsilon^{M} + \bar{g}^{M}(\epsilon^{M})\bar{Q}_{2} \} + (P_{2} - c_{1}^{M})\bar{Q}_{2} \end{aligned}$$

Therefore Manufacturer chooses ϵ^{M^e} to maximize $-\epsilon + \bar{g}^M(\epsilon)\bar{Q}_2$ independent of Supplier's effort decision. Since \bar{g}^M is concave, ϵ^{M^e} is unique.

Supplier also chooses an optimal effort level ϵ^{S^e} given his anticipated Manufacturer's optimal effort, i.e.,

$$\max\{-\epsilon^{S} + E_{\delta^{M},\delta^{S}}(c_{1}^{M} - c_{1}^{S} + g^{S}(\epsilon^{S}, \delta^{S}) - g^{M}(\epsilon^{M^{e}}, \delta^{M})1_{\{q_{1}^{M} \geq q_{min}^{M}\}})^{+}\bar{Q}_{2}\}$$

There may be multiple ϵ^S that maximize the equation above, but since Supplier's actual cost in the second period does not affect Manufacturer's profit, this will not affect the analysis. \blacksquare

C.1.3 Proof of Lemma 3

Proof: Given a wholesale price ω_1 and demand Q_1 , Manufacturer allocate production in order to maximize profit:

$$\begin{split} R_1^M &= (P_1 - c_1^M)q_1^M + (P_1 - \omega_1)(Q_1 - q_1^M) + (P_2 - c_1^M)\bar{Q}_2 + G^M(\bar{Q}_2)\mathbf{1}_{\{q_1^M \geq q_{min}^M\}} \\ &= (P_1 - \omega_1)Q_1 + (\omega_1 - c_1^M)q_1^M + (P_2 - c_1^M)\bar{Q}_2 + G^M(\bar{Q}_2)\mathbf{1}_{\{q_1^M \geq q_{min}^M\}} \\ &= (P_1 - \omega_1)Q_1 + (P_2 - c_1^M)\bar{Q}_2 + \begin{cases} (\omega_1 - c_1^M)q_1^M & \text{if } q_1^M < q_{min}^M \\ (\omega_1 - c_1^M)q_1^M + G^M(\bar{Q}_2) & \text{else} \end{cases} \end{split}$$

So the optimal quantity allocation is obtained by choosing q_1^M from $\{0, q_{min}^M, Q_1\}$, and the lemma follows.

C.1.4 Proof of Theorem 1

Proof: Let $R^S(0)$, $R^S(q_{min}^M)$, and $R^S(Q_1)$ be Supplier's expected profits if Manufacturer produces $q_1^M = 0$, $q_1^M = q_{min}^M$, and $q_1^M = Q_1$, respectively, then

$$R^{S}(0) = (c_{1}^{M} - c_{1}^{S} - \frac{G^{M}(\bar{Q}_{2})}{q_{min}^{M}})\bar{Q}_{1} + R^{Se}$$

$$R^{S}(q_{min}^{M}) = (c_{1}^{M} - c_{1}^{S})(\bar{Q}_{1} - q_{min}^{M}) + R^{MSe}$$

$$R^{S}(Q_{1}) = 0 + R^{Me}$$

If we can show $R^S(0)$, $R^S(q^M_{min})$, and $R^S(Q_1)$ are all increasing in $c^M_1 - c^S_1$, $R^S(0)$ increases faster than $R^S(q^M_{min})$, and $R^S(q^M_{min})$ increases faster than $R^S(Q_1)$, then there exist two thresholds x_1 and x_2 , as shown in Figure 4.1. So it is sufficient to show that R^{S^e} , R^{MS^e} , and R^{M^e} are all increasing in $c^M_1 - c^S_1$, which is straightforward to show, and $\frac{\partial R^{S^e}}{\partial (c^M_1 - c^S_1)} \ge \frac{\partial R^{M^e}}{\partial (c^M_1 - c^S_1)} \ge \frac{\partial R^{M^e}}{\partial (c^M_1 - c^S_1)}$.

Let $h(x) = \max\{-\epsilon^S + E(x + g^S(\epsilon^S, \delta^S))\bar{Q}_2\}$, then h(x) is convex since the maximization and expectation of convex functions are still convex. Then $R^{S^e} = h(c_1^M - c_1^S)$, and $R^{MS^e} = h(c_1^M - c_1^S - g^M(\epsilon^{M^e}, \delta^M))$, so $\frac{\partial R^{S^e}}{\partial (c_1^M - c_1^S)} \ge \frac{\partial R^{MS^e}}{\partial (c_1^M - c_1^S)}$.

For the comparison between R^{MS^e} and R^{M^e} , let $h_1(\epsilon^S) = -\epsilon^S + E(c_1^M - c_1^S + g^S(\epsilon^S, \delta^S) - g^M(\epsilon^{M^e}, \delta^M))^+ \bar{Q}_2$, and $h_2 = E(c_1^M - c_1^S - g^M(\epsilon^{M^e}, \delta^M))^+ \bar{Q}_2$. Then for any positive ϵ^S , $\frac{\partial h_1}{\partial (c_1^M - c_1^S)} \ge \frac{\partial h_2}{\partial (c_1^M - c_1^S)}$. Therefore $\frac{\partial R^{MS^e}}{\partial (c_1^M - c_1^S)} = \frac{\partial h_1}{\partial (c_1^M - c_1^S)}|_{\epsilon^{S^*}} \ge \frac{\partial h_2}{\partial (c_1^M - c_1^S)} = \frac{\partial R^{M^e}}{\partial (c_1^M - c_1^S)}$. $x_1 < 0$ since at $c_1^M - c_1^S = 0$, $R^S(q_{min}^M) = R^{MS^e}$ and $R^S(Q_1) = R^{M^e} \le R^S(q_{min}^M)$, thus it is never optimal for Manufacturer to produce all demand in house.

The condition for $x_2 \ge 0$ is obtained by setting $R^S(q_{min}^M) \ge R^S(0)$ at $c_1^M - c_1^S = 0$.

C.1.5 Proof of Lemma 4

Proof:
$$\epsilon_1^{M^e} = \frac{1}{4}a^2Q_2^2$$
, $\epsilon_1^{S^e} = \frac{1}{4}b^2Q_2^2$, $G^{M^*}(Q_2) = \frac{1}{4}a^2Q_2^2$, $G^{S^*}(Q_2) = \frac{1}{4}b^2Q_2^2$, $g^M(\epsilon_1^{M^e}) = \frac{1}{2}a^2Q_2$.

Supplier makes wholesale price decisions by comparing the following four options and the equilibrium outcome follows.

$$R^{S}(Q_{1}, Q_{2}) = 0$$

$$R^{S}(q_{min}^{M}, Q_{2}) = (c_{1}^{M} - c_{1}^{S})(Q_{1} - q_{min}^{M})$$

$$R^{S}(q_{min}^{M}, 0) = (c_{1}^{M} - c_{1}^{S}(Q_{1} - q_{min}^{M}) + (c_{1}^{M} - c_{1}^{S} - \frac{1}{2}a^{2}Q_{2}))Q_{2} + \frac{1}{4}b^{2}Q_{2}^{2}$$

$$= (c_{1}^{M} - c_{1}^{S}(Q_{1} + Q_{2} - q_{min}^{M}) + (\frac{1}{4}b^{2} - \frac{1}{2}a^{2})Q_{2}^{2}$$

$$R^{S}(0, 0) = (c_{1}^{M} - c_{1}^{S} - \frac{1}{q_{min}^{M}} \frac{1}{4}a^{2}Q_{2})Q_{1} + (c_{1}^{M} - c_{1}^{S})Q_{2} + \frac{1}{4}b^{2}Q_{2}^{2}$$

$$= (c_{1}^{M} - c_{1}^{S})(Q_{1} + Q_{2}) + (\frac{1}{4}b^{2} - \frac{1}{4}a^{2}\frac{Q_{1}}{q_{min}^{M}})Q_{2}^{2}$$

C.1.6 Proof of Lemma 6

Proof: (a) Based on the two-period results, Manufacturer is never hurt by Supplier's improvement, so $(Q_1 - q_{min}^S)_+$ is dominated by $Q_1 - q_{min}^S$ or Q_1 .

(b) Based on the two-period results, Manufacturer's profit is independent of c^S for certain production allocation. Manufacturer benefits from Supplier's learning only when $(q_2^M, q_3^M) = (0, 0)$ if Supplier improves and $(q_2^M, q_3^M) \neq (0, 0)$ otherwise. If Supplier does not improve or his improvement does not benefit Manufacturer, $(Q_1 - q_{min}^S)$ is dominated by q_{min}^M or Q_1 .

C.1.7 Proof of Lemma 7

Proof: $\varepsilon^{S^*}(Q_2 + Q_3) = \arg\max\{-\epsilon_1^S + (c_2^M - c_2^S + f^S(\epsilon_1^S))(Q_2 + Q_3) + G^{S^*}(Q_3)\}$ and the profit is positive if $c_2^M - c_1^S \ge A - \frac{G^{S^*}(Q_2 + Q_3)}{Q_2 + Q_3} + N$.

C.1.8 Proof of Lemma 8

Proof: If $c_1^M - c_2^S < A$, then $c_2^M - c_2^S < A$ for any effort, and the problem is identical to the benchmark case.

If $c_1^M - c_2^S > A + E$, then $c_2^M - c_2^S > A$ if $\epsilon_1^M = \varepsilon^{M^*}(Q_2 + Q_3)$. By investing more effort the marginal benefit can only be smaller than $f^{M'}(\epsilon^S)(Q_2 + Q_3)$.

If $A \leq c_1^M - c_2^S \leq A + E$, Manufacturer has two options: either improve by $\varepsilon^{M^*}(Q_2 + Q_3)$ or just improve to make $c_2^M - c_2^S = A$. The first option is better when $c_1^M - c_1^S < B$.

C.1.9 Proof of Theorem 2

Proof:

- (a) For $c_1^M c_1^S < A D$, $c_2^S = c_1^S$ by Lemma 18, and $c_1^M c_1^S = c_1^M c_2^S < A D$. For $c_1^M - c_1^S \ge A + E$, $c_1^M - c_2^S \ge A + E$, Supplier will fully participate no matter he learns or not. In both cases Manufacturer does not benefit from Supplier's learning. By Lemma 19, the problem is identical to the benchmark case.
- (b), (c) For $A-D \leq c_1^M-c_1^S < A$, if only Supplier can learn, by Lemma 18 $c_2^S=c_1^S-\Delta c^S(Q_2+Q_3)$ and $c_2^M-c_2^S=c_1^M-c_2^S\geq A$ since Supplier earns positive profit. This brings Manufacturer extra profit $(\frac{Q_2}{q_{min}^M}-1)G^{M^*}(Q_3)$. If both can learn, $c_2^S=c_1^S-\Delta c^S(Q_2+Q_3)$ and Manufacturer invests no more than ε^{M^*} . This brings

Manufacturer extra profit of no more than $(\frac{Q_2}{q_{min}^M}-1)G^{M*}(Q_3)+G^{M*}(Q_2+Q_3)$. If only Manufacturer can learn, $c_1^M-c_2^S < A$ and he invests $\varepsilon^{M*}(Q_2+Q_3)$. The extra profit is $G^{M*}(Q_2+Q_3)$.

For $A \leq c_1^M - c_1^S < A + B$, if only Supplier can learn, Manufacturer does not benefit. If both can learn, Manufacturer benefits by no more than $G^{M^*}(Q_2 + Q_3)$. This benefit must be higher than $G^{M^*}(Q_2 + Q_3) - (\frac{Q_2}{q_{min}^M} - 1)G^{M^*}(Q_3)$ since the latter can always be achieved by investing $\varepsilon^{M^*}(Q_2 + Q_3)$. If only Manufacturer can learn, he invests ε^{M^*} and $c_2^M - c_2^S < A$. The extra profit is $G^{M^*}(Q_2 + Q_3) - (\frac{Q_2}{q_{min}^M} - 1)G^{M^*}(Q_3)$.

For $A+B \leq c_1^M-c_1^S < A+D$, if only Supplier can learn, Manufacturer does not benefit. If both can learn, Manufacturer benefits by $G^{M*}(Q_2+Q_3)$ or $f^{M-1}(c_1^M-c_1^S+\Delta c^S(Q_2+Q_3)-A)+(c_1^M-c_1^S+\Delta c^S(Q_2+Q_3)-A)(Q_2+Q_3)\leq G^{M*}(Q_2+Q_3)$, i.e., he either improves by $\Delta c^M(Q_2+Q_3)$ or makes $c_2^M-c_2^S=A$, depending on which profit is larger. If only Manufacturer can learn, $c_2^M-c_2^S=A$, and Manufacturer benefits by $f^{M-1}(c_1^M-c_1^S-A)+(c_1^M-c_1^S-A)(Q_2+Q_3)$. In all the above three cases, Manufacturer is better off with both learning than with learning alone. So Manufacturer may produce $Q_1-q_{min}^S$ even if $\omega_1>c_1^M$. Let $\tilde{\omega}$ be the price such that Manufacturer is indifferent between producing $q_1^M=Q_1$ and $Q_1-q_{min}^S$, then $\tilde{\omega}>c_1^M$.

In the benchmark case, The difference of Manufacturer's benefits of both learning and of only Supplier learning is $G^{M*}(Q_2 + Q_3)$, while in the above three cases the difference is smaller. Let $\bar{\omega}$ be the price such that Manufacturer is

indifferent between producing $q_1^M = q_{min}^M$ and 0. By equating two profits we can obtain $\bar{\bar{\omega}} = c_1^M - \frac{R^M(\text{both learn}) - R^M(\text{only Supplier learns})}{q_{min}^M}$, then $\bar{\bar{\omega}} \geq \bar{\omega}$.

Manufacturer accepts a higher wholesale price than the one in the benchmark case to give up full quantity.

Since Manufacturer is indifferent between producing q_{min}^M and $Q_1 - q_{min}^S$ when $\omega_1 = c_1^M$, the equilibrium can be obtained by comparing Supplier's profits of four options: $(\omega_1 = \bar{\omega}, q_1^M = 0)$, $(\omega_1 = c_1^M, q_1^M = q_{min}^M)$, $(\omega_1 = \tilde{\omega}, q_1^M = Q_1 - q_{min}^S)$, and $(\omega_1 = \infty, q_1^M = Q_1)$.

C.1.10 Proof of Theorem 3

Proof: In the first period Manufacturer only considers four quantities: $Q_1, Q_1 - q_{min}^S, q_{min}^M, 0$. Since we have proved that only four types of quantity allocation can be optimal in two-period problem, there are totally 16 candidates, and we can rule out five of them:

- (Q_1, q_{min}^M, Q_3) : If $c_1^M > c_1^S$, Supplier can offer $\omega_1 = c_1^M$, then $q_1^M < Q_1$. Supplier can earn positive profit in the first period and both players can be better off because of smaller c_2^S ; If $c_1^M < c_1^S$, and Manufacturer produces everything in Period 1, then $c_2^M < c_2^S$, and by Theorem 1 (q_{min}^M, Q_3) cannot be optimal.
- $(0, Q_2, Q_3)$: Supplier's profit $R^S = (\omega_1 c_1^S)Q_1 \ge 0$, implying $c_1^S \le \omega_1 \le c_1^M$ and thus $c_2^S \le c_1^S \le c_1^M = c_2^M$. (Q_2, Q_3) is not optimal by Theorem 1.
- $(Q_1 q_{min}^S, Q_2, Q_3), (Q_1 q_{min}^S, q_{min}^M, Q_3), \text{ and } (Q_1 q_{min}^S, q_{min}^M, 0) \text{ are not possible}$ by Lemma 17.

Assuming that Demands across periods are stationary, three more candidates can be further ruled out using two-period results:

• $(Q_1, 0, 0)$ is dominated by $(q_{min}^M, 0, 0)$:

By Theorem 1, if (0,0) is the equilibrium quantity , then in scenario 1, $c_2^M - c_2^S \ge \frac{M}{2Q} \ge \frac{-G^{S^*}(Q)}{2Q} \ge \frac{-G^{S^*}(Q)}{Q - q_{min}^M}$. In scenario 1, $c_2^M - c_2^S \ge \frac{M - N}{q_{min}^M} \ge \frac{M}{2Q} \ge \frac{-G^{S^*}(Q)}{Q - q_{min}^M}$ by the conditions of scenario 2. In scenario 3 and 4, $c_2^M - c_2^S > 0 > \frac{-G^{S^*}(Q)}{Q - q_{min}^M}$. If (Q, 0, 0) is the equilibrium quantity, then $c_1^M - c_1^S \ge c_2^M - c_2^S \ge \frac{-G^{S^*}(Q)}{Q - q_{min}^M}$.

Compare Supplier's profits of (Q,0,0) and $(q_{min}^M,0,0)$. By producing $(q_{min}^M,0,0)$ Supplier can improve and earn $G^{S^*}(2Q)$ extra profit in the last two periods. In the first period, Supplier offers $\omega_1 = c_1^M$ and earns $(c_1^M - c_1^S)(Q - q_{min}^M) \geq \frac{-G^{S^*}(Q)}{Q - q_{min}^M}(Q - q_{min}^M) = -G^{S^*}(Q) \geq -G^{S^*}(2Q)$. Supplier earns nothing in the first period by producing Q. So by producing $(q_{min}^M, 0, 0)$ Supplier's profit is higher.

• $(Q_1, q_{min}^M, 0)$ is dominated by $(q_{min}^M, q_{min}^M, 0)$:

By Theorem 1, $(q_{min}^M, 0)$ can be the equilibrium quantity only in scenario 2 and 3. In scenario 2, $c_2^M - c_2^S \ge \frac{N}{2Q - q_{min}^M} \ge \frac{-G^{S^*}(Q)}{2Q - q_{min}^M} \ge \frac{-G^{S^*}(Q)}{Q - q_{min}^M}$. In scenario 3, $c_2^M - c_2^S > 0 > \frac{-G^{S^*}(Q)}{Q - q_{min}^M}$. If $(Q, q_{min}^M, 0)$ is the equilibrium quantity, then $c_1^M - c_1^S \ge c_2^M - c_2^S \ge \frac{-G^{S^*}(Q)}{Q - q_m^M}$.

Compare Supplier's profits of $(Q, q_{min}^M, 0)$ and $(q_{min}^M, q_{min}^M, 0)$. Since Manufacturer does not benefit from Supplier's learning, c_2^M are the same, so by producing $(q_{min}^M, q_{min}^M, 0)$ Supplier can improve and earn $G^{S^*}(2Q - q_{min}^M)$ extra profit

in the last two periods. In the first period, Supplier offers $\omega_1 = c_1^M$ and earns $(c_1^M - c_1^S)(Q - q_{min}^M) \ge \frac{-G^{S^*}(Q)}{Q - q_{min}^M}(Q - q_{min}^M) = -G^{S^*}(Q) \ge -G^{S^*}(2Q - q_{min}^M)$. Supplier earns nothing in the first period by producing Q. So by producing $(q_{min}^M, q_{min}^M, 0)$ Supplier's profit is higher.

• $(0, q_{min}^M, Q_3)$ is dominated by (q_{min}^M, Q_2, Q_3) : $(q_{min}^M, Q_3) \text{ is optimal in the last two periods only when } f^M, f^S \text{ satisfy conditions}$ in scenario 3 and 4. In scenario 3, $c_1^M - c_1^S \leq c_2^M - c_2^S \leq \frac{N}{Q_2} < \frac{M}{Q_2}$; In scenario 4, $c_1^M - c_1^S \leq c_2^M - c_2^S \leq \frac{M}{Q_3 + q_{min}^M} < \frac{M}{Q_3}$. Manufacturer does not benefit

from Supplier's learning and the equilibrium wholesale price is identical to the benchmark case, $\omega_1 = c_1^M - \frac{G^{M^*}(2Q)}{\sigma^M}$.

$$R_{S}(0, q_{min}^{M}, Q_{3}) = (c_{1}^{M} - \frac{G^{M^{*}}(2Q)}{q_{min}^{M}} - c_{1}^{S})Q + (c_{1}^{M} - c_{1}^{S})(Q - q_{min}^{M}) + G^{S^{*}}(Q - q_{min}^{M})$$

$$R^{S}(q_{min}^{M}, Q_{2}, Q_{3}) = (c_{1}^{M} - c_{1}^{S})(Q - q_{min}^{M})$$

$$R_{S}(0, q_{min}^{M}, Q_{3}) - R^{S}(q_{min}^{M}, Q_{2}, Q_{3}) = (c_{1}^{M} - \frac{G^{M^{*}}(2Q)}{q_{min}^{M}} - c_{1}^{S})Q + G^{S^{*}}(Q - q_{min}^{M})$$

$$\leq M - \frac{Q}{q_{min}^{M}}G^{M^{*}}(2Q) + G^{S^{*}}(Q - q_{min}^{M})$$

$$= \frac{Q}{q_{min}^{M}}G^{M^{*}}(Q) - G^{S^{*}}(Q) - \frac{Q}{q_{min}^{M}}G^{M^{*}}(2Q) + G^{S^{*}}(Q - q_{min}^{M}) \leq 0$$

C.1.11 Proof of Observation 1

Proof: First notice that in the second period Manufacturer never shares effort cost since he does not benefit from Supplier's improvement.

In the first period,

1. Supplier's effort decision: Given c_2^M and ε_1^{MS} , Supplier decides effort level ϵ_1^S .

$$R_2^S = \begin{cases} 0 & \text{if } (q_2^M, q_3^M) = (Q_2, Q_3) \\ (c_2^M - c_2^S)(Q_2 - q_{min}^M) & \text{if}(q_2^M, q_3^M) = (q_{min}^M, Q_3) \\ (c_2^M - c_2^S)(Q_2 + Q_3) + N & \text{otherwise} \end{cases}$$

Since $\varepsilon^{S^*}(Q_2 - q_{min}^M) = 0$,

$$\epsilon_1^S = \begin{cases} (\varepsilon^{S^*}(Q_2 + Q_3) - \varepsilon_1^{MS})^+, & \text{if } R^S(c_2^M, \varepsilon_1^{MS}, \varepsilon^{S^*}(Q_2 + Q_3)) \ge 0; \\ 0, & \text{otherwise.} \end{cases}$$

In other word, Supplier's cost improvement remains the same, but he only invests the effort not paid by Manufacturer. Therefore Supplier is more willing to improve compared with no cost sharing case.

2. Manufacturer's effort sharing decision:

Clearly $0 < \varepsilon_1^{MS} \le \varepsilon^{S^*}(Q_2 + Q_3)$ does not help Manufacturer. By Lemma 18, if $c_2^M - c_1^S > A - D$, then Supplier improves and produces full quantity in the next two periods even without effort shared by Manufacturer, so $\varepsilon^{MS} = 0$.

If $c_2^M - c_1^S < A - D$, Manufacturer only benefits if he can induce Supplier to fully participate in future periods and chooses the minimum possible effort that makes Supplier indifferent between full participation and not participating. If $\varepsilon^{MS} > 0$, Supplier always produces everything and gains zero profit. Let A - D' be the smallest cost difference $c_2^M - c_1^S$ such that $\varepsilon^{MS} > 0$.

3. Quantity and wholesale price decisions:

Consider $A-D' < c_1^M - c_1^S < A-D$ and close to A-D. If Manufacturer cannot share effort cost, similar to Theorem 11, the equilibrium is identical to the benchmark case and the wholesale price for $q_1^M = 0$ is $\bar{\omega} = c_1^M - \frac{G^{M*}(Q_2 + Q_3)}{q_{min}^M}$,

where $G^{M*}(Q_2 + Q_3)$ is the difference of Manufacturer's profit in the last period if both learn and if only Supplier can learn.

If Manufacturer has the option of sharing effort cost, then if only supplier can learn, Manufacturer offers ε^{MS} to help Supplier and gains extra profit $-\varepsilon^{MS}+(\frac{Q_1}{q_{min}^M}-1)G^{M^*}(Q_3)$. If both firms can learn, since $(\frac{Q_1}{q_{min}^M}-1)G^{M^*}(Q_3)>G^{M^*}(Q_2+Q_3)$ by assumption, Manufacturer is better off by not fully improving. The optimal effort level keeps $c_2^M-c_1^S+\Delta c^S(Q_2+Q_3)=A$. So the extra profit gain from learning is between $-\varepsilon^{MS}+(\frac{Q_1}{q_{min}^M}-1)G^{M^*}(Q_3)$ and $-\varepsilon^{MS}+(\frac{Q_1}{q_{min}^M}-1)G^{M^*}(Q_3)+G^{M^*}(Q_2+Q_3)$. Thus the profit difference between both learning and only Supplier learning becomes smaller, resulting a higher wholesale price $\bar{\omega}$. Supplier need not offer a wholesale discount as deep as before.

If $c_1^M - c_1^S$ is close enough to A - D > 0, then the extra profit gain when both can learn is close to $-\varepsilon^{MS} + (\frac{Q_1}{q_{min}^M} - 1)G^{M^*}(Q_3)$, so $\bar{\omega}$ is close to c_1^M . If (0,0,0) is the equilibrium quantity, Supplier gains profit $R^S = (\bar{\omega} - c_1^S)Q_1 + 0 > 0$. Since $\bar{\omega} = c_1^M - \frac{1}{q_{min}^M}(R_2^M(\text{both learn}) - R_2^M(\text{only S learns}))$, Manufacturer's profit can be obtained by:

$$R^{M} = (P_{1} - c_{1}^{M} + \frac{1}{q_{min}^{M}} (R_{2}^{M}(\text{both learn}) - R_{2}^{M}(\text{only S learns})))Q_{1} + R_{2}^{M}(\text{only S learns})$$

$$= (P_{1} - c_{1}^{M})Q_{1} + \frac{Q_{1}}{q_{min}^{M}} (R_{2}^{M}(\text{both learn}) - R_{2}^{M}(\text{only S learns})) + R_{2}^{M}(\text{only S learns})$$

As mentioned above, $(R_2^M(\text{both learn}) - R_2^M(\text{only S learns}))$ becomes smaller with effort sharing option, Manufacturer's profit becomes smaller if Q_1 is large enough relative to Q_2 .

BIBLIOGRAPHY

BIBLIOGRAPHY

- [1] A. A. Akinola. An application of the bass model in the analysis of diffusion of coco-spraying chemicals among nigerian cocoa farmers. *Journal of Agricultural Economics*, 37(3):395–404, 1986.
- [2] S. Anily. Single machine lot sizing with uniform yields and rigid demands: Robustness of the optimal solution. *IIE Transactions*, 27:625–633, 1993.
- [3] A. Atamturk and D. Hochbaum. Capacity acquisition, subcontracting, and lot sizing. *Management Science*, 47(8):1081–1100, 2001.
- [4] F. M. Bass. A new product growth model for consumer durables. *Management Science*, 15(5):215–227, 1969.
- [5] A Beja. Optimal reject allowance with constant marginal production efficiency. Naval Research Logistics Quarterly, 24:21–33, 1997.
- [6] J. A. Dodson and E. Muller. Models of new product diffusion through advertising and word-of-mouth. *Management Science*, 24(11):1568–1578, 1978.
- [7] R. Dolan and A. P. Jeuland. Experience curves and dynamic demand models: Implications for optimal pricing strategies. *Journal of Marketing*, 45(1):52–62, 1981.
- [8] Y. Gerchak and A. Grosfeld-Nir. Lot-sizing for substitutable, production-toorder parts with random functionality yields. *The International Journal of Flex*-

- ible Manufacturering Systems, 11:371–377, 1999.
- [9] T. Grabill, D. Gross, and M. Magazine. A classified bibliography of research on optimal design and control of queues. *Operations Research*, 25(2):217–232, 1977.
- [10] J. V. Gray, B. T. Tomlin, and A. V. Roth. Outsourcing to a powerful contract manufacturer: The effect of learning-by-doing. working paper, 2008.
- [11] A. Grosfeld-Nir. Single bottleneck systems with proportional expected yields and rigid demand. European Journal of Operational Research, 80:297–307, 1995.
- [12] A. Grosfeld-Nir and Y. Gerchak. Multiple lotsizing with random common-cause yield and rigid demand. *Operations Research Letters*, (9):383–387, 1990.
- [13] A. Grosfeld-Nir and Y. Gerchak. Production to order with random yields: Single-stage multiple lot-sizing. IIE Transactions, 28:669–676, 1996.
- [14] A. Grosfeld-Nir and Y. Gerchak. Multiple lotsizing in production to order with random yields: Review of recent advances. Annals of Operations Research, 126:43–69, 2004.
- [15] A. Grosfeld-Nir, Y. Gerchak, and Q. He. Manufacturing to order with random yield and costly inspection. 48(5):761–767, 2000.
- [16] A. Grosfeld-Nir and B. Ronen. A single bottleneck system with binomial yields and rigid demand. *Management Science*, 39(5):650–653, 1993.
- [17] M. Henig and Y. Gerchak. The structure of periodic review policies in the presence of random yield. Operations Research, 38(4):634–643, 1990.
- [18] T. Ho, S. Savin, and C. Terwiesch. Managing demand and sales dynamics in new product diffusion under supply constraint. *Management Science*, 48(2):187–206,

2002.

- [19] D. Jain, V. Mahajan, and E. Muller. Innovation diffusion in the presence of supply restrictions. *Marketing Science*, 10(1):83–90, 1991.
- [20] S. Kalish. Monopolist pricing with dynamic demand and production cost. Marketing Science, 2(2):135–159, 1983.
- [21] S. Kalish. A new product adoption model with price, advertising, and uncertainty. *Management Science*, 31(12):1569–1585, 1985.
- [22] S. Kalish and G. L. Lilien. A market entry timing model for new technologie.
 Management Science, 32(2):194–205, 1986.
- [23] M. Kamien and L. Li. Subcontracting, coordination, flexibility, and production smoothing in aggregate planning. *Management Science*, 36(11):1352–1363, 1990.
- [24] M. Kamien, L. Li, and D. Samet. Bertrand competition with subcontracting. The RAND Journal of Economics, 20(4):553–567, 1989.
- [25] H. Krishnan, R. Kapuscinski, and D. Butz. Coordinating contracts for decentralized supply chains with promotional effort. *Management Science*, 50(1):48–63, 2004.
- [26] T. V. Krishnan, F. M. Bass, and D. C. Jain. Optimal pricing strategy for new products. *Management Science*, 45(12):1650–1663, 1999.
- [27] S. Kumar and J. M. Swaminathan. Diffusion of innovations under supply constraints. *Operations Research*, 51(6):866 –879, 2003.
- [28] A. Kurawarwala and H. Matsuo. Forecasting and inventory management of short lifecycle products. Operations Research, 44(1):131–150, 1996.

- [29] S. Lee and P. Zipkin. A dynamic lot-size model with make-or-buy decisions.

 Management Science, 35(4):447–458, 1989.
- [30] T. Lewis and D. Sappington. Technological change and the boundaries of the firm. The American Economic Review, 81(4):887–900, 1991.
- [31] H Luss. Operations researchand capacity expansion problems: a survey. *Operations Research*, 30(5):907–945, 1982.
- [32] V Mahajan, E. Muller, and F. M. Bass. New product diffusion models in marketing: A review and directions for research. *Journal of Marketing*, 54(1):1 –26, 1990.
- [33] K. Moinzadeh and H. L. Lee. A continuous review inventory model with constant resupply time and defective items. *Naval Research Logistics*, pages 457–468, 1987.
- [34] A. Parmigiani. Concurrent sourcing: When do firms both make and buy? *Ph.D.*Dissertation, The University of Michigan Business School, 2002.
- [35] C. Pascale, B. De Reyck, Z. Degraeve, and W. Chong. Research and development project valuation and licensing negotiations at phytopharm plc. *Interfaces*, 37:472–487, 2007.
- [36] E. Plambeck and T. Taylor. Sell the plant? the impact of contract manufacturing on innovation, capacity, and profitability. *Management Science*, 51(1):133–150, 2005.
- [37] E. L. Porteus. The impact of inspection delay on process and inspection lotsizing.
 Management Science, 36:999–1007, 1990.
- [38] E.L. Porteus. Optimal lotsizing, process quality improvement and setup cost

- reduction. Operations Research, 34:137–144, 1986.
- [39] M. L. Puterman. Markov Decision Processes: Discrete Stochastic Dynamic Programming. 1997.
- [40] B. Robinson and C. Lakhani. Dynamic price models for new-product planning. Management Science, 21(10):1113–1122, 1975.
- [41] M. Sawhney and J. Eliashberg. A parsimonious model for forecasting gross box-office revenues of motion pictures. *Marketing Science*, 15(2):113–131, 1996.
- [42] S. P. Sethi and G. L. Thompson. Optimal Control theory: applications to management science and economics. Kluwer Academic Publishers, 2000.
- [43] W. Shih. Optimal inventory policies when stockouts result from defective products. *International Journal of Production Research*, 18(6):677–685, 1980.
- [44] Y. Spiegel. Horizontal subcontracting. The RAND Journal of Economics, 24(4):133–150, 1993.
- [45] A. Stock and S. Balachander. The making of a "hot product": A signaling explanation of marketer's scarcity strategy. *Management Science*, 51(8):1181– 1192, 2005.
- [46] F. Sultan, J. U. Farley, and D. R. Lehmann. A meta-analysis of diffusion models. Journal of Marketing, 27(1):70–77, 1990.
- [47] C Van den Bulte. New product acceleration: measurement and analysis. Marketing Science, 19(4):366–380, 2000.
- [48] J. Van Miegham. Coordinating investment, production and subcontracting.

 Management Science, 45(7):954–971, 1999.

- [49] Y. Wang and Y. Gerchak. Periodic review production models with variable capacity, random yield, and uncertain demand. *Management Science*, 42(1):130– 137, 1996.
- [50] Y. Wang and Y. Gerchak. Input control in a batch production system with lead times, due dates and random yields. European Journal of Operational Research, 126:371–385, 2000.
- [51] C. Yano and H. Lee. Lot sizing with random yields: A review. *Operations Research*, 43(2):311–334, 1995.
- [52] A. Zhang and S. Guu. The multiple lot sizing problem with rigid demand and interrupted geometric yield. *IIE Transactions*, 30:427–431, 1998.