Evaluation of Supply Chain Strategies for Mass Customization

A Dissertation
Presented to
The Academic Faculty

By

Jack C. P. Su

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in the
School of Management

Georgia Institute of Technology
August 2004

Evaluation of Supply Chain Strategies for Mass Customization

Approved by:

Dr. Yih-Long Chang, Advisor

Dr. Mark Ferguson

Dr. Cheryl Gaimon

Dr. Soumen Ghosh

Dr. Paul Griffin

Date Approved: July 6th 2004

ACKNOWLEDGEMENT

I would like to acknowledge several people for their guidance, encouragement, and friendship. Dr. Yih-Long Chang, who served as the chairman of this dissertation, provided me considerable advice, motivation, and observations. Without his constructive assistance, this research would not be finished. I would like to express my great appreciation to him.

A special thanks and appreciation are extended to Dr. Mark Ferguson, Dr. Cheryl Gaimon, Dr. Souman Ghosh, and Dr. Paul Griffin for serving as my committee members. Their comments and supports for this research are substantial.

I would like to give special thanks to my parents and my sisters for their unconditional support and to my friends in ATPC and AATA for making my life in Atlanta memorable. Finally, I would like to thank God's guidance. May glory belong to the Lord.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

Product proliferation is a common challenge for firms providing customized products. To cope with this challenge, firms usually incorporate strategies such as component commonality, postponement, and/or delayed differentiation in their supply chains. In this dissertation, we study the effectiveness of these strategies.

Component commonality (CC) is one of the most popular supply chain strategies to cope with challenges of product proliferation such as difficulties in estimating demand, controlling inventory, and providing high service levels for customers. It advocates using a common component to replace a number of distinctive components in various products so that the safety stock can be reduced due to risk pooling.

In this dissertation, we first evaluate three component commonality strategies in supply chain environment: Distinctive Part (DP), Pure Component Commonality (PCC), and Mixed Component Commonality (MCC) strategies. DP is where all products consist of distinctive parts and no common component is used, while PCC is where one or more parts from different products are completely replaced by common components. In MCC, unlike PCC, it allows partial substitution of distinctive parts with common components. We develop models to analyze these strategies for both the constant and stochastic demands. The solution to minimize the total inventory cost is presented.

For constant demand, MCC is the worst choice and PCC is the best for the case of low common component price, high ordering cost, or high interest rate. For stochastic demand, PCC is the best for the case of low common component price, high demand variation, high ordering cost, long lead time, or high interest rate. However, MCC can be used to reduce inventory cost if the demand variation is high. Furthermore, we conclude

that when demand variation is moderate, unit shortage cost is not a significant factor in the choice of component commonality strategies. In the case of high demand variation, the PCC strategy is preferred when shortage cost is high, and the MCC strategy can be adopted for a range of moderate shortage cost.

Second, we study the performance of two postponement strategies and their relationship with product proliferation. In order to meet increasing customer demands for more diverse product offerings, firms are revising their supply chain strategies to accommodate mass customization. The revised strategies often involve delaying the delivery of the products until after the customer orders arrive, termed Time Postponement (TP), or delaying the differentiation of the products until later production stages, termed Form Postponement (FP). We develop models representing the TP and FP strategies and compare their performance in total supply chain cost and expected customer waiting times. We find that once the number of different products increases above some threshold level, the TP strategy is preferred under both performance metrics.

For our most general model, we design a numerical experiment to investigate how different factors affect the performance of the TP and FP strategies. Through this experiment we show that higher arrival time and process time variations make the FP strategy more favorable while increases in the number of products and higher interest rates make the TP strategy more favorable. We also offer guidance to managers using either strategy on where to allocate resources for performance improvement. For example, to improve the customer waiting times under the FP strategy, increasing the coverage of the generic component and reducing the number of products provide larger benefits than reducing the variability of the arrival and process times.

Third, we analyze the relevant costs and benefits of implementing delayed differentiation, an implementation of the FP, in a make-to-order environment and provide insights for managers choosing the optimal point along the supply chain they should differentiate their products.

To achieve mass customization, many firms are shifting their supply chain structures from make-to-stock to make-to-order. A make-to-order strategy comes at a price however, as customers must wait longer for their customized products. Incorporating delayed differentiation in a make-to-order environment offers potential to reduce the customer's wait since the generic part/component of the products can be made before the customer order is received. In the third part of our study, we quantify the tradeoffs involved in implementing delayed differentiation in a make-to-order environment using both customer waiting time and cost as performance metrics. We show that under common assumptions, the introduction of delayed differentiation results in shorter waiting times and higher cost over a pure make-to-order strategy. However, we derive reasonable conditions where the introduction of delayed differentiation results in shorter customer waiting times *and* lower cost, thus dominating a pure make-to-order strategy on both performance dimensions. We also provide insights to firms seeking the optimal place in the production process to differentiate their products. For example, we show that the expected customer waiting time is convex in the point of the process that differentiation takes place. Thus, for firms seeking to minimize customer waits, differentiating the product in the middle of the process may result in shorter customer waits than waiting until the end of the process than waiting until the end of the process.

CHAPTER 1

INTRODUCTION

Rapid changes in technology and increased globalization are two common trends of today's business environment. One of the immediate responses to this new environment is increased product proliferation (Lee 1996). Companies experiencing product proliferation face increasing problems in forecasting demand, controlling their inventory, and providing high service levels for their customers. To deal with these problems, companies usually incorporate strategies such as component commonality, postponement, and/or delayed differentiation in their supply chains. In this dissertation, we study the effectiveness of these strategies.

Component commonality advocates using a common component to replace a number of distinctive components in various products so that the safety stock can be reduced due to risk pooling. In this dissertation, we compare three component commonality strategies and evaluate their impact on inventory systems for both the constant and stochastic demand scenarios.

In addition to the component commonality strategy, various supply chain strategies have been explored to provide a wide range of product varieties in a cost efficient way (also referred to as mass customization). Many of these strategies involve either delaying the delivery of the products until after receiving the customer orders or delaying the differentiation of the products until later stages of the supply chain. Zinn and Bowersox (1988) label the former as Time Postponement (TP) and the later as Form Postponement (FP). Employing TP involves delaying the manufacturing and shipping of the product until after the customer order is received. Production and distribution of the product is most often centralized in a single facility. An example of a company using TP is the Danish company Bang & Olufsen, a high-end television and stereo system manufacturer. All of Bang & Olufsen's products are made-to-order at a centralized plant and shipped directly to their customers. The need for holding safety stock is eliminated when using TP but customers must be willing to wait the entire manufacturing lead-time for their customized products.

In contrast to TP, employing FP involves shipping the products in a semi-finished state from the manufacturing facility to a downstream facility where final customization occurs. In order to delay the final customization of the product, the firm stocks a generic (semi-finished) component from which it draws upon for final assembly. A classic example of a company using FP is Hewlett-Packard's (HP) postponement of the final assembly of their DeskJet printers to their local distribution centers (Lee et al. 1993).

Although the viability of the postponement strategies has been discussed, the environments where one type of postponement strategy is outperforms the other have not received sufficient attention. Also, despite the fact that increasing product proliferation is often a major factor behind a firm's decision to incorporate a postponement strategy, its impact on the choice of strategy has not been addressed. In this dissertation, we seek to fill these gaps.

After comparing TP and FP, we study the costs and benefits of a new strategy which combines TP and FP, and compare it to a pure make-to-order strategy. We first examine a popular FP strategy called Delayed Differentiation which advocates that a firm redesigns its products so that differentiation is postponed until later stages of the supply chain (Lee and Tang 1997). In the past, research in delayed differentiation mainly focuses on its implications in make-to-stock environment. However, more and more companies start to implement delayed differentiation in make-to-order environments. Consider the retail market for household paints as an example. Retailers such as Home Depot and Lowe's have shifted from stocking a wide variety of premixed colors to stocking paint in a neutral color (generic component) and mixing the final color only after receiving a specific customer order (the point of differentiation is delayed from the production site to the retail site). Since delayed differentiation can also provide substantial benefits for companies choosing a make-to-order strategy, we analyze the costs and benefits of implementing delayed differentiation in a make-to-order environment and provide insights for managers choosing where along the supply chain they should differentiate their products.

The remainder of this dissertation is organized as follows. The review of the literature is present in Section 2. Sections 3, 4 and 5 present the problems and results of our study in component commonality, postponement, and delayed differentiation, respectively. Section 6 summarizes this dissertation and discusses potential areas for future research.

CHAPTER 2

LITERATURE REVIEW

In this section, we review the literature in component commonality, postponement, and delayed differentiation. The research questions of this dissertation are motivated and identified.

2.1 Component Commonality

Much research has been done in component commonality. Collier (1981) defines an index to measure the degree of component commonality. He finds that higher degree of component commonality is significantly associated with the reduction in manufacturing cost. Baker et al (1986) present a two-product, two-level, single period inventory model to study the effect of commonality on the number of units in stock. Their model minimizes the number of units in stock with a specified service level under the normally distributed demand scenario. They show that by introducing commonality the total number of units in inventory is reduced and the inventory level of the common component is lower than the total inventory level of the two components it has replaced. Gerchak et al (1988) extend Baker et al's work and minimize the total material acquisition cost under the general demand distribution case. Eynan and Rosenblatt (1996) extend Baker et al's work by allowing the price of the common component to exceed the price of the common component that it replaces and analyze cases in which commonality is still economically justified. Hillier (1999) extends the model of Eynan and Rosenblatt (1996) to consider the multiple-period case, and concludes that benefits of commonality are lessened in the multiple-period case.

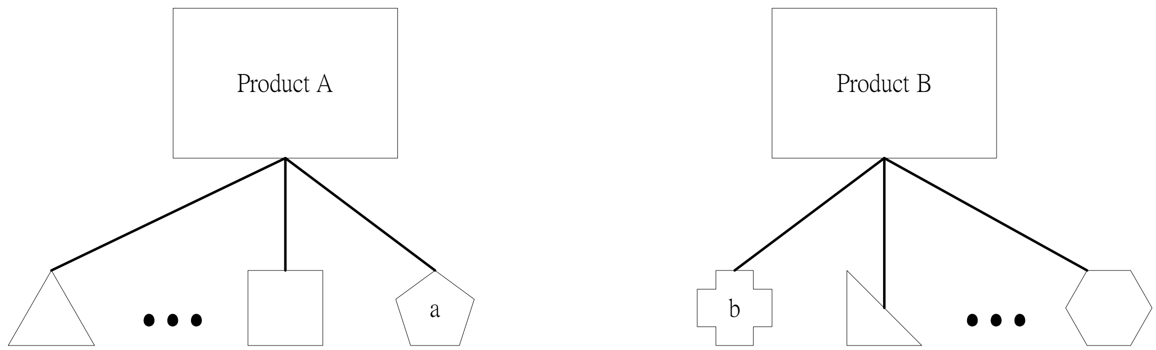

Figure 1: Distinctive Part Strategy.

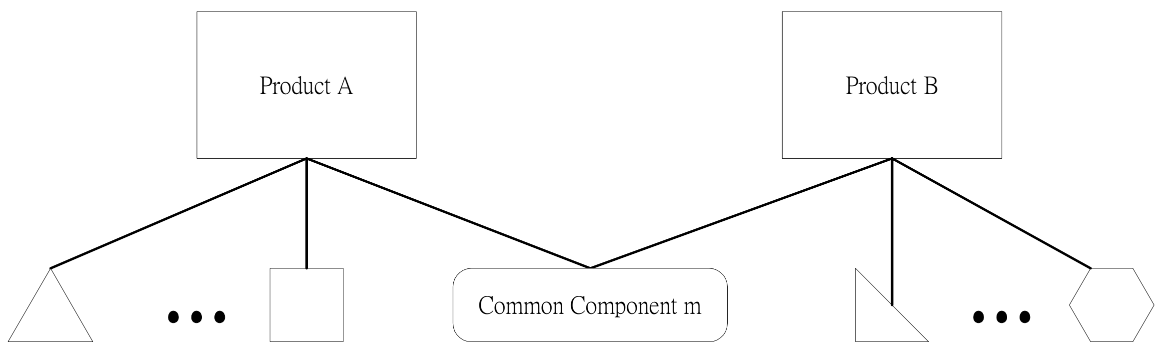

Figure 2: Pure Component Commonality Strategy

Our study departs from the previous research in two aspects. First, all of the previous research focuses on the comparison between Distinctive Part (DP) strategy and Pure Component Commonality (PCC) strategy. The DP strategy is where all products consist of distinctive parts and no common component is used (see **Figure 1**). The Pure PCC strategy is where one or more distinctive parts are completely replaced by a common component (see **Figure 2**). We call those parts which can be replaced by the common component as the replaceable parts (e.g. parts a and b).

In this dissertation, we propose partial substitution of replaceable parts by a common component and call this approach the Mixed Component Commonality (MCC) Strategy as shown in **Figure 3**. An example of the MCC strategy can be observed from the PC industry. Most of the PC manufacturers utilize the Pentium CPU in their high-end PCs and the Celeron CPU in their low-end PCs. There are generally three types of motherboard chip sets available: the chip set that supports only the Pentium CPU, the chip set that supports only the Celeron CPU, and the chip set that supports both CPUs. PC manufacturers often use combinations of these three types of chip sets in their PCs.

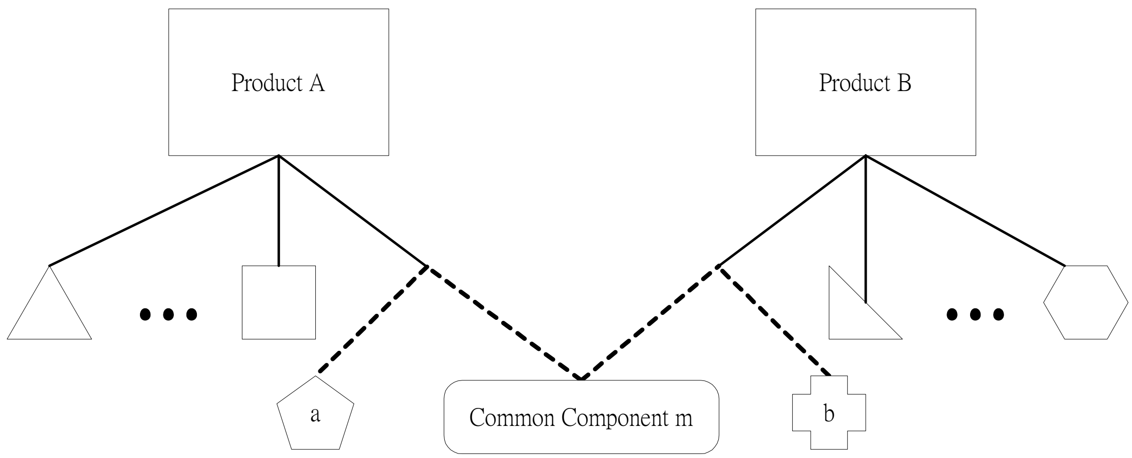

Figure 3: Mixed Component Commonality Strategy

An inevitable outcome of the MCC strategy is an increase in the number of parts being inventoried. In the past, it was very expensive to design, produce, and manage an extra part and was therefore impractical to adopt the MCC strategy. However, the trend towards outsourcing has made the MCC strategy more feasible today. A company can easily order different parts from its suppliers. In addition, the adoption of the information technology in inventory management has reduced the cost and complexity of managing a very large number of parts. For these reasons, it is important to consider the MCC strategy in this study.

The second aspect distinguishing this study from the previous research is that we extend the previous models by considering all inventory related costs including the ordering cost and shortage cost. The detailed description of our models is presented in Section 3.

2.2 Postponement Strategies

Research on the concept of postponement originated from Bucklin (1965), who was the first to use the term "postponement" but did not provide any analytical results. Christopher (1992) provides case studies describing how postponement worked in the European market and Lee et al. (1993) presents the HP DeskJet printer case involving multiple international markets. In both cases, the authors found that significant supply chain savings could be achieved in shorter lead-time and lower safety stocks by redesigning the product or process to delay the differentiation decision. Feitzinger and Lee (1997) and Grag and Tang (1997) provide analytical models measuring the costs and benefits of delayed differentiation - a type of Form Postponement. They show that reductions in safety stock levels due to risk-pooling is the key benefit while the cost of designing and manufacturing the generic component is the main drawback.

Zinn and Bowersox (1988), Cooper (1993), and Pagh and Cooper (1998) overview different types of postponement strategies and discuss their potential benefits but do not provide models to compare the strategies analytically. Although the viability of various postponement strategies has been discussed, the environments where one type of postponement strategy outperforms the other have not received sufficient attention. Also, despite the fact that increasing product proliferation is often a major factor behind a firm's decision to incorporate a postponement strategy, its impact on the selection of the optimal strategy to implement has not been addressed. In this paper, we seek to fill these gaps.

We compare the TP strategy and the FP strategy by using queuing models and derive conditions under which each strategy is preferred. In addition, we show how product proliferation affects the supply chain performance of both strategies. Two performance measures are employed in the evaluation. The first is the total supply chain cost, which includes the amortized fixed cost and the periodic operating cost. The second is the expected customer waiting time, i.e., the time to fulfill the orders. These two measures are important evaluation criteria for most supply chain managers. The model detail and results are presented in Section 4.

2.3 Delayed Differentiation

Previous research on delayed differentiation focuses mainly on make-to-stock (MTS) strategies, where the main benefit comes from savings in inventory holding cost while the main drawback is the cost of designing the generic component (Lee and Tang 1997, Garg and Tang 1997, and Swaminathan and Tayur 1998). Shown in the earlier customized paint example, delayed differentiation can also provide substantial benefits for companies choosing a make-to-order strategy. Hence, we investigate the costs and benefits of implementing delayed differentiation in a make-to-order environment. To do so, we model two supply chain strategies for meeting customized product demand: a pure make-to-order (MTO) strategy and a configure-to-order (CTO) strategy where delayed differentiation is adopted in a make-to-order environment. Both strategies are fully described in the next section.

Lee (1996) also presents a make-to-order model where the point of product differentiation can be delayed. Lee shows that if the unit holding cost of the generic component is not increasing with respect to the degree of delayed differentiation, the total inventory holding cost may decrease the further down the differentiation takes place in the supply chain. We extend Lee's work in several ways. First, production lead-times are an output of our model as opposed to Lee's model where they are assumed to be constant and to be independent of the demand, the production capacity, and the structure of the supply chain. Second, since the fixed cost of designing and manufacturing the generic component is often cited as the main drawback of delayed differentiation, we include this important factor in our cost function. Third, we model all pertinent costs including the fixed cost, unit inventory holding cost, unit work-in-process (WIP) holding cost, and the unit production cost, as functions of the degree of delayed differentiation. The detailed description of models and results are presented in Section 5.

CHAPTER 3

EVALUATION of COMPONENT COMMONALITY

We develop two-product, two-level, and multiple period inventory models for both the constant and stochastic demand scenarios, to present the solution to minimizing the total inventory cost, and to derive managerial insight from our analysis. In this section, the model, solution procedure, and the managerial insights for constant and stochastic demand scenarios are presented.

Suppose that a firm manufactures a product family consisting of two products, *A* and *B*. Product *A* is oriented for higher end market while product *B* is aimed at lower end market. We use *a* to denote a subsystem of *A* and *b* to denote a subsystem of *B*. Subsystems *a* and *b* may be replaced by a common component, *m*. The common component, *m*, generally equips with extended functions such as additional conjunction to attach to two different end products and has to meet the quality standard of the higher end product, *A*. Hence, it is reasonable to assume the unit price of *m* is more expensive than the replaceable parts. That is $P_m \geq P_a \geq P_b$, where P_j, j= *m*, *a*, *b* are the unit price for replaceable parts *a* and *b*, and common component, *m*, respectively.

In this study, we focus on the internal choice of component commonality strategies; hence we assume that all components are outsourced from the same or similar suppliers and have the same constant delivery lead-time and ordering cost. Since the components are outsourced, there is no additional cost involved in the design, development, and manufacturing of the components. Therefore, inventory related costs - unit cost, ordering cost, inventory holding cost, and shortage cost – are the only ones of interest.

We study both the constant demand and stochastic demand problems. The following notation is defined for use in this paper.

D_i	Annual demand for end product i, $i = A, B$
P_j	Unit price of subsystem j, $j=a, b, m$
A	Ordering cost
L	Lead time
x_j	Random variable representing demand for subsystem j over lead time with mean μ_j and standard deviation σ_j, $j=a, b, m$
e_i	Degree of commonality, i.e. percentage of the product i produced by using the common component, $i= A, B$
Q_j	Order quantity of subsystem j, $j=a, b, m$
I	Annual interest rate
$h_j = iP_j$,	Unit inventory holding cost per period of time of subsystem j, $j=a, b, m$
$\bar{b}_j$	Expected unit of shortage during lead time of subsystem j, $j=a, b, m$
π_j	unit shortage cost of subsystem j, $j=a, b, m$

3.1 Constant Demand Problem

In this basic model we assume the demands of products A and B are constant. The total material cost is:

$$P_a(1-e_a)D_A + P_b(1-e_b)D_B + P_m(e_a D_A + e_b D_B) \quad \textbf{(3.1)}$$

The first two terms represent the material cost of acquiring a and b; and the third terms denotes the material cost of acquiring m. (1-e_a) and (1-e_b) are the percentages of demands fulfilled by replaceable parts a and b. ($e_aD_A + e_bD_B$) is the quantity of the common component m ordered.

Since all the components are outsourced, the only setup cost is ordering cost. The total ordering cost is

$$\frac{A(1-e_a)D_A}{Q_a} + \frac{A(1-e_b)D_B}{Q_b} + \frac{A(e_a D_A + e_b D_B)}{Q_m} \quad \textbf{(3.2)}$$

The total inventory holding cost is

$$h_a\left[\frac{Q_a}{2}\right]+h_b\left[\frac{Q_b}{2}\right]+h_m\left[\frac{Q_m}{2}\right] \qquad \textbf{(3.3)}$$

The long-term average inventory position for part j is $\frac{Q_j}{2}$, $j=a,b,m$.

The total cost of the constant demand problem is the summation of Equations **(3.1)** **(3.2)**, and **(3.3)**. We will solve for Q_a, Q_b, Q_m, e_a, and e_b to minimize the total inventory cost in Section

3.2 Solution Procedure of Constant Demand Problem

We present the procedure to minimize the inventory cost of the constant problems in this section.

Recall that the total cost function of this problem is:

$$\begin{aligned}
&TC = \\
&P_a(1-e_a)D_A + P_b(1-e_b)D_B + P_m(e_aD_A+e_bD_B)+ \\
&\frac{A_a(1-e_a)D_A}{Q_a}+\frac{A_b(1-e_b)D_B}{Q_b}+\frac{A_m(e_aD_A+e_bD_B)}{Q_m}+ \\
&h_a\left[\frac{Q_a}{2}\right]+h_b\left[\frac{Q_b}{2}\right]+h_m\left[\frac{Q_m}{2}\right]
\end{aligned} \qquad \textbf{(3.4)}$$

To find the optimal order quantity, we first set

$\frac{\partial TC}{\partial Q_a}=0$ and solve for Q_a, and we obtain

$$Q_a^{\ *}=\sqrt{\frac{2A(1-e_a)D_A}{h_a}} \qquad \textbf{(3.5)}$$

Similarly, we obtain

$$Q_b^{\,*} = \sqrt{\frac{2A(1-e_b)D_B}{h_b}} \quad \textbf{(3.6)}$$

$$Q_m^{\,*} = \sqrt{\frac{2A(e_a D_A + e_b D_B)}{h_m}} \quad \textbf{(3.7)}$$

Substitute Q_a^*, Q_b^*, and Q_m^* into Equation **(3.4)**. Now the total cost function is reduced to a function of e_a and e_b, i.e., $TC = f(e_a, e_b)$.

To find the optimal degree of commonality, we set $\frac{\partial TC}{\partial e_a} = 0$ and get

$$D_A(P_m - P_a) - \sqrt{\frac{(2h_a A D_A)}{2((1-e_a)}} + \sqrt{\frac{h_m A}{2(e_a D_A + e_b D_B)}} D_A = 0 \quad \textbf{(3.8)}$$

Again, we set $\frac{\partial TC}{\partial e_b} = 0$ and get

$$D_B(P_m - P_b) - \sqrt{\frac{(2h_b A D_B)}{2((1-e_b)}} + \sqrt{\frac{h_m A}{2(e_a D_A + e_b D_B)}} D_B = 0 \quad \textbf{(3.9)}$$

We solve Equations **(3.8)** and **(3.9)** simultaneously to determine the optimal degree of commonality, e_a^* and e_b^*. However, there is no simple closed form solution.

To find an alternative way to solve this problem analytically, we now exploit the concavity property of the cost function.

Lemma 3.1: *Total Cost Function is concave in e_a and e_b.*

Proof:

$$\frac{\partial^2 TC(e_a,e_b)}{\partial e_a{}^2}\frac{\partial^2 TC(e_a,e_b)}{\partial e_b{}^2}-\left[\frac{\partial^2 TC(e_a,e_b)}{\partial e_a \partial e_b}\right]^2 =$$

$$(2(1-e_a))^{-\frac{3}{2}}(2(1-e_b))^{-\frac{3}{2}}(h_a AD_A)^{\frac{1}{2}}(h_b AD_B)^{\frac{1}{2}} + \tag{3.10}$$

$$(2(1-e_b))^{-\frac{3}{2}}(h_b AD_B)^{\frac{1}{2}}(2(e_a D_A + e_b D_B))^{-\frac{3}{2}}(h_m A)^{\frac{1}{2}} D_A{}^2 +$$

$$(2(1-e_a))^{-\frac{3}{2}}(h_a AD_A)^{\frac{1}{2}}(2(e_a D_A + e_b D_B))^{-\frac{3}{2}}(h_m A)^{\frac{1}{2}} D_B{}^2 \geq 0$$

$$\frac{\partial^2 TC(e_a,e_b)}{\partial e_a{}^2} = \tag{3.11}$$

$$-(2(1-e_a))^{-\frac{3}{2}}(h_a AD_A)^{\frac{1}{2}} - (2(e_a D_A + e_b D_B))^{-\frac{3}{2}}(h_m A)^{\frac{1}{2}} D_A{}^2 \leq 0$$

$$\frac{\partial^2 TC(e_a,e_b)}{\partial e_b{}^2} = \tag{3.12}$$

$$-(2(1-e_b))^{-\frac{3}{2}}(h_b AD_B)^{\frac{1}{2}} - (2(e_a D_A + e_b D_B))^{-\frac{3}{2}}(h_m A)^{\frac{1}{2}} D_B{}^2 \leq 0$$

From Equations **(3.10)**, **(3.11)**, and **(3.12)**, the total cost function is concave in e_a and e_b. ■

Since the total cost function is concave in e_a and e_b, the optimal degree of commonality e_a^* and e_b^* may only assume either 0 or 1. For this reason, the minimum cost must be among $f(0,0)$, $f(0,1)$, $f(1,0)$, or $f(1,1)$.

From Equation **(3.4)** we obtain

$$f(0,0) = P_a D_A + P_b D_B + \sqrt{2AD_A h_a} + \sqrt{2AD_B h_b} \tag{3.13}$$

$$f(0,1) = P_a D_A + P_m D_B + \sqrt{2AD_A h_a} + \sqrt{2AD_B h_m} \tag{3.14}$$

$$f(1,0) = P_m D_A + P_b D_B + \sqrt{2AD_A h_m} + \sqrt{2AD_B h_b} \tag{3.15}$$

$$f(1,1) = P_m (D_A + D_B) + \sqrt{2A(D_A + D_B)h_m} \tag{3.16}$$

Since $P_m \geq P_a \geq P_b$, then $f(0,1)$ and $f(1,0)$ must be greater than or equal to $f(0,0)$. Hence the candidates for an optimal solution are reduced to only two, i.e., $f(0,0)$ or $f(1,1)$.

In addition, numerical examples can be used to demonstrate that there is no definitive inequality relationship between $f(0,0)$ and $f(1,1)$. Therefore, the optimal solution is either e_a=e_b=0 or e_a=e_b =1. In other words we only need to compute $f(0,0)$ and $f(1,1)$ and select the one with a lower cost as an optimal solution. We summarized the above finding in the following lemma.

> ***Lemma 3.2:*** *Assume $P_m \geq P_a \geq P_b$, the optimal degree of commonality is either e_a=e_b=0 or e_a=e_b =1.*

The optimal order quantity for *a*, *b*, and *m* can be calculated by substituting the optimal degree of commonality found by ***Lemma 3.2*** into Equations **(3.5)**, **(3.6)**, and **(3.7)**.

From ***Lemma 3.2***, since the minimum cost can be achieved by using DP strategy (e_a=e_b=0) or PCC strategy (e_a=e_b =1), so we can now conclude that MCC strategy is not useful in the constant demand problem.

3.3 Strategy Comparison in Constant Demand Environments

Besides solving the problem, it is important to learn which component commonality strategy is preferred under various conditions. To do so, we first calculate the cost difference between the PCC strategy and DP strategy as follows:

$$\begin{aligned}&\Delta TC = f(1,1) - f(0,0) = \\ &(D_A)(P_m - P_a) + (D_B)(P_m - P_b) + \\ &\sqrt{2Ah_m(D_A + D_B)} - \sqrt{2AD_A h_a} - \sqrt{2AD_B h_b}\end{aligned} \quad \textbf{(3.17)}$$

We assume the demands of products *A* and *B* are proportional to a total demand *D*; hence, $D_A = K_A D$ and $D_B = K_B D$. Since the unit holding cost is equal to the price times interest rate, i.e., $h_j = P_j \cdot i$, $j = a, b, m$. Equation **(3.17)** becomes

$$
\begin{aligned}
&\Delta TC = f(1,1) - f(0,0) = \\
&(K_A D)(P_m - P_a) + (K_B D)(P_m - P_b) + \\
&\sqrt{2AP_m i(K_A D + K_B D)} - \sqrt{2AK_A DP_a i} - \sqrt{2AK_B DP_b i}
\end{aligned} \tag{3.18}
$$

We first study how change of P_m impacts the choice of the strategies.

Lemma 3.3: *ΔTC is an increasing function with respect to P_m.*

Proof:

$$
\frac{\partial \Delta TC}{\partial P_m} = K_A D + K_B D + (2P_m)^{-\frac{1}{2}}\left(Ai(K_A D + K_B D)\right)^{\frac{1}{2}} > 0 . \ \blacksquare
$$

From ***Lemma 3.3***, when *m* is more expensive, the cost of PCC strategy is higher. Thus, the DP strategy is preferred. Next, we discuss how the ordering cost will impact the choice of the strategy.

Lemma 3.4: *ΔTC is a decreasing function with respect to A.*

Proof:

$$
\begin{aligned}
\frac{\partial \Delta TC}{\partial A} &= (2A)^{-\frac{1}{2}}\left(P_m i(K_A D + K_B D)\right)^{\frac{1}{2}} - (2A)^{-\frac{1}{2}}\left(P_a iK_A D\right)^{\frac{1}{2}} - (2A)^{-\frac{1}{2}}\left(P_b iK_B D\right)^{\frac{1}{2}} \\
&= (2A)^{-\frac{1}{2}}(Di)^{\frac{1}{2}}\left\{\left(P_m(K_A + K_B)\right)^{\frac{1}{2}} - \left(P_a K_A\right)^{\frac{1}{2}} - \left(P_b K_B\right)^{\frac{1}{2}}\right\}
\end{aligned}
$$

Because $P_m \geq P_a \geq P_b$, the above equation is less than or equal to

$$
(2A)^{-\frac{1}{2}}(Di)^{\frac{1}{2}}\left\{\left(P_a(K_A + K_B)\right)^{\frac{1}{2}} - \left(P_a K_A\right)^{\frac{1}{2}} - \left(P_a K_B\right)^{\frac{1}{2}}\right\}
$$

$$
= (2A)^{-\frac{1}{2}}(P_a Di)^{\frac{1}{2}}\left\{\sqrt{(K_A + K_B)} - \sqrt{K_A} - \sqrt{K_B}\right\} < 0 \ \blacksquare
$$

From ***Lemma 3.4***, when the ordering cost is higher, the cost of the DP strategy is higher. Thus, the PCC strategy is preferred. The rationale for the PCC strategy is that order pooling resulting from having to stock only a common component *m* helps lower the ordering cost. As a consequence, the larger that the ordering cost is, the larger the saving is.

Lemma 3.5: *ΔTC is a decreasing function with respect to i.*

Proof:

$$\frac{\partial \Delta TC}{\partial i} = (2i)^{-\frac{1}{2}}\left(P_m A(K_A D + K_B D)\right)^{\frac{1}{2}} - (2i)^{-\frac{1}{2}}\left(P_a A K_A D\right)^{\frac{1}{2}} - (2i)^{-\frac{1}{2}}\left(P_b A K_B D\right)^{\frac{1}{2}}$$
$$= (2i)^{-\frac{1}{2}}(DA)^{\frac{1}{2}}\left\{\left(P_m(K_A + K_B)\right)^{\frac{1}{2}} - \left(P_a K_A\right)^{\frac{1}{2}} - \left(P_b K_B\right)^{\frac{1}{2}}\right\}$$

Similar to the proof of ***Lemma 3.4***, where $P_m \geq P_a \geq P_b$, the above equation is less than or equal to

$$(2i)^{-\frac{1}{2}}(DA)^{\frac{1}{2}}\left\{\left(P_a(K_A + K_B)\right)^{\frac{1}{2}} - \left(P_a K_A\right)^{\frac{1}{2}} - \left(P_a K_B\right)^{\frac{1}{2}}\right\}$$

$$= (2i)^{-\frac{1}{2}}(P_a DA)^{\frac{1}{2}}\left\{\sqrt{(K_A + K_B)} - \sqrt{K_A} - \sqrt{K_B}\right\} < 0 \quad \blacksquare$$

Theoretically, a higher interest rate increases the unit holding cost. Since the price of the common component is more expensive, so in a high interest rate environment, the unit holding cost for the common component increases more than that of the replaceable parts. Consequently, the DP strategy is preferred. On the other hand, a higher unit holding cost reduces the optimal order quantity which in turn reduces the average inventory level. Since it is more expensive to hold the common component, lower inventory levels benefit the PCC strategy. The bottom line is that without proof it remains unclear how the interest rate will impact the choice between these two strategies. From ***Lemma 3.5*** we prove that when the interest rate is higher, the cost of the DP strategy is higher. Therefore, the PCC strategy is preferred.

In short, we conclude that the PCC strategy is preferred when the price of common component is lower, the order cost is higher, or the interest rate is higher.

3.4 Stochastic Demand Problem

In real life the demand is rarely constant. For this reason, it is important to study how the different component commonality strategies perform in the stochastic demand environment. In this model, we assume the demands over lead time for components j, j = a, b, m, are normally distributed with mean μ_j and standard deviation σ_j.

Since the demand over lead-time is stochastic, it is possible that there exists shortage during lead time. The expected shortage during lead time is defined as

$$\overline{b}_j = \int_{s_i}^{\infty} (x_j - s_j) f_j(x_j) dx_j \text{ , } j = a, b, m$$

where $f_j(x_j)$ are probability density functions. The shortage cost is calculated as

$$\frac{\pi_a \overline{b}_a D_A}{Q_a} + \frac{\pi_b \overline{b}_b D_B}{Q_b} + \frac{\pi_m \overline{b}_m (e_a D_A + e_b D_B)}{Q_m} \qquad \textbf{(3.19)}$$

Material cost and ordering cost are the same as the constant demand problem given as follows.

Material cost:

$$P_a(1 - e_a)D_A + P_b(1 - e_b)D_B + P_m(e_a D_A + e_b D_B) \qquad \textbf{(3.20)}$$

Ordering cost:

$$\frac{A(1 - e_a)D_A}{Q_a} + \frac{A(1 - e_b)D_B}{Q_b} + \frac{A(e_a D_A + e_b D_B)}{Q_m} \qquad \textbf{(3.21)}$$

To calculate the inventory holding cost, we first estimate the long-term average inventory position. Obtaining an exact expression for average inventory level is difficult and an approximation result will be used here, which is good if the time the system is in a backorder condition during a cycle is small compared to the cycle length (Johnson and Montgomery 1974). In most reality, this is the case. However, when backorder time is long, the approximation used here may lead to high error.

Since the maximum inventory position during lead time is the order quantity plus safety stock or re-order point minus average demand over the lead time, i.e., $Q_j + s_j - \mu_j$ and the minimum inventory position during lead-time is safety stock or re-order point subtract average demand over lead time, i.e., $s_j - \mu_j$. Hence, the expected long-term average inventory position is:

$$\frac{Q_j}{2} + s_j - \mu_j \quad j = a, b, m$$

The inventory holding costs are estimated as:

$$h_a\left[\frac{Q_a}{2} + s_a - \mu_a\right] + h_b\left[\frac{Q_b}{2} + s_b - \mu_b\right] + h_m\left[\frac{Q_m}{2} + s_m - \mu_m\right] \quad \textbf{(3.22)}$$

The total cost of the stochastic demand problem is the summation of Equations **(3.19)**, **(3.20)**, **(3.21)**, and **(3.22)**. We will solve for Q_a, Q_b, Q_m, s_a, s_b, s_m, e_a, and e_b to minimize the total inventory cost in the next section.

3.5 Solution Procedure of Stochastic Demand Problem

From Equations **(3.19)**, **(3.20)**, **(3.21)**, and **(3.22)**, the total cost function for stochastic demand is:

$$TC = P_a(1-e_a)D_A + P_b(1-e_b)D_B + P_m(e_aD_A + e_bD_B) +$$
$$\frac{A(1-e_a)D_A}{Q_a} + \frac{A(1-e_b)D_B}{Q_b} + \frac{A(e_aD_A + e_bD_B)}{Q_m} +$$
$$h_a\left[\frac{Q_a}{2} + s_a - \mu_a\right] + h_b\left[\frac{Q_b}{2} + s_b - \mu_b\right] + h_m\left[\frac{Q_m}{2} + s_m - \mu_m\right] + \qquad \textbf{(3.23)}$$
$$\frac{\pi_a\bar{b}_aD_A}{Q_a} + \frac{\pi_b\bar{b}_bD_B}{Q_b} + \frac{\pi_m\bar{b}_m(e_aD_A + e_bD_B)}{Q_m}$$

Unlike the constant demand problem, the stochastic demand problem does not have a nice closed form solution. We therefore solve this problem numerically.

The stochastic demand problem carries two important properties:

1. There is no singular point with respect to e_a and e_b when $0 < e_a < 1$ and $0 < e_b < 1$.
2. Given e_a and e_b, solving the stochastic demand problem is equivalent to solving three independent (s, Q) problems.

The first property ensures that there will be no sudden jump in total cost when we search along e_a and e_b. For this reason, we can search our optimal solution along e_a and e_b in a discrete manner. The interval chosen for each search is 0.01. So the complete search consists of 10,000 possible combinations of e_a and e_b. The solution of (s, Q) problem is based on the procedure discussed in Johnson and Montgomery (1974).

3.6 Strategy Comparison in Stochastic Demand Environments

Since there is no closed form solution to the stochastic demand problem, the managerial insights are derived by numerical studies. Our study is based on a numerical example taken from Johnson and Montgomery (1974) in which D_A = 10,000, D_B = 20,000; the standard deviations of demands are set to 20%; the ordering cost is \$70; the interest rate is 20%; P_m = \$5.05, P_a = \$5.00, and P_b = \$4.95; the unit shortage cost is \$1.50; and the lead time is two weeks. To test the effect of a given factor, we fix the other parameters and vary the factor to cover a very broad range. We first study how the price of the common component impacts the choice of the strategy.

> **Observation 3.1:** *The DP strategy is preferred when the price of the common component is high.*

We vary P_m from \$4.00 to \$5.50. From **Figure 4**, when the price of the common component is higher, the DP strategy is preferred. This observation is consistent with our analysis in ***Lemma 3.3***. However, **Figure 4** also shows that it is not worthwhile to use the common component even when it is just a few cents more expensive than the replaceable parts. The reason is that unlike the single period model, the lower safety stock implies a firm can buy fewer components. For a multiple period model, although the PCC strategy allows fewer safety stock inventories, the total number of components purchased is unchanged since all demands are eventually met.

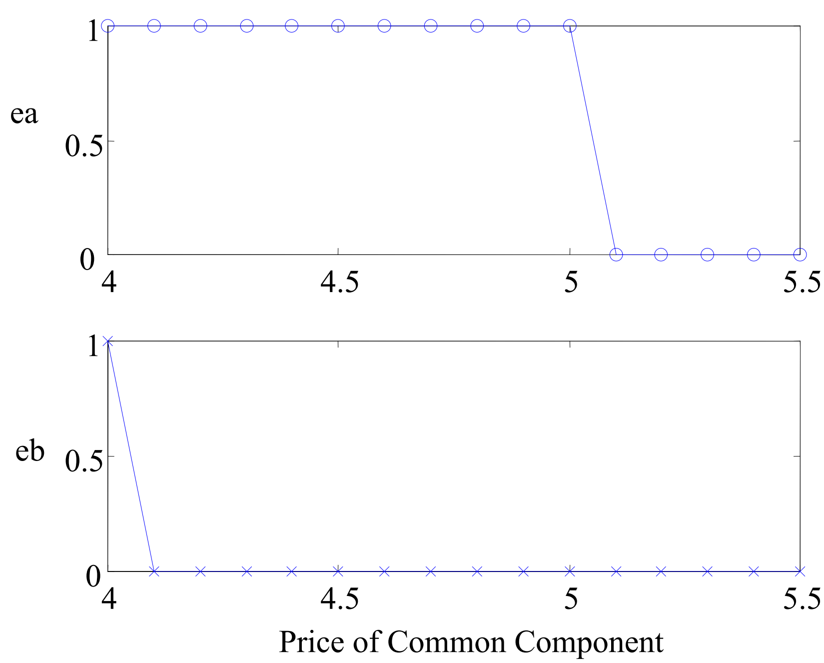

Figure 4: The Effect of the Price of Common Component

Observation 3.2: *The PCC strategy is preferred when the demand variation is high.*

To test the effect of demand variation, we fix the values of other factors and vary the standard deviations of the total demand from 10% to 100%. From **Figure 5**, when the demand variation is high, the PCC strategy is preferred because the benefit of using the common component to pool demand variation outweighs the extra cost of buying it. It is also worth noting that for product *B*, when the demand variation is more than 70%, the minimal cost is achieved by using the MCC strategy; i.e., use both replaceable part *b* and common component *m*. This leads to our next observation.

Observation 3.3: *The MCC strategy could be used to reduce total cost when demand variation is high.*

Observation 3.4: *The PCC strategy is preferred when the ordering cost is high.*

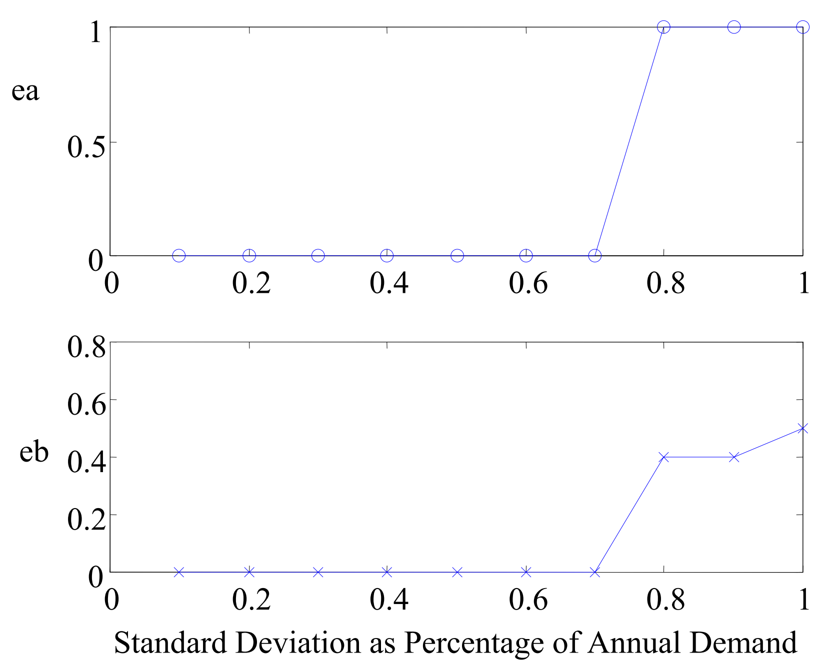

Figure 5: The Effect of Demand Variation

To test the effect of ordering cost, we vary the ordering cost from $70 to $1,400. From **Figure 6**, when the ordering cost is high, the PCC strategy is preferred. This observation is consistent with our finding in ***Lemma 3.5***.

Observation 3.5: *The PCC strategy is preferred when the lead time is long.*

To test the effect of lead time, we vary the lead time from two weeks to forty weeks. From **Figure 7**, when the lead time is longer, the PCC strategy is preferred. The reason is that when lead time is longer, the unit of demand variation during lead time becomes larger and the benefit of using the common component to pool demand variation becomes larger.

Observation 3.6: *The PCC strategy is preferred when the interest rate is high.*

To test the effect of interest rate, we vary the interest rate from 20% to 100%. From **Figure 8**, when the interest rate is high, the PCC strategy is preferred. This observation is consistent with our finding in ***Lemma 3.5***.

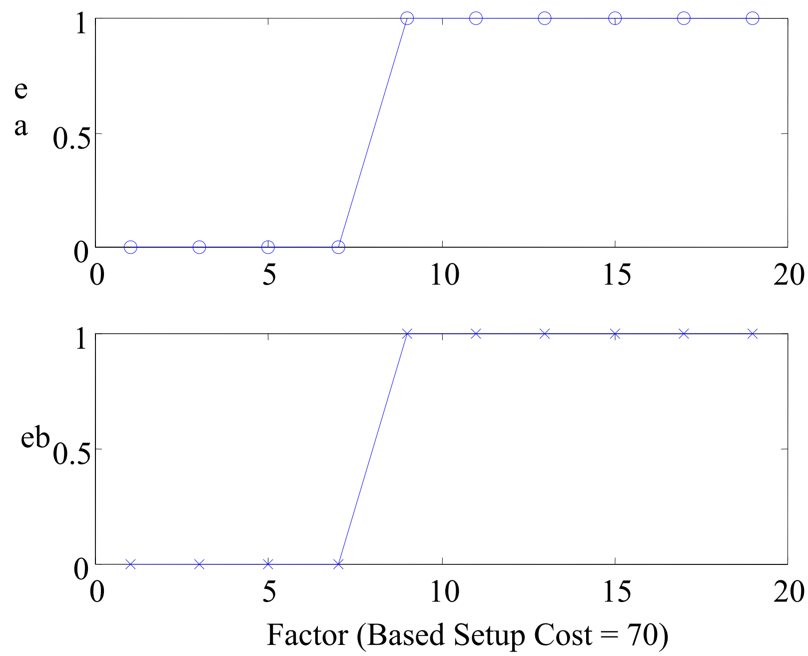

Figure 6: The Effect of Ordering Cost

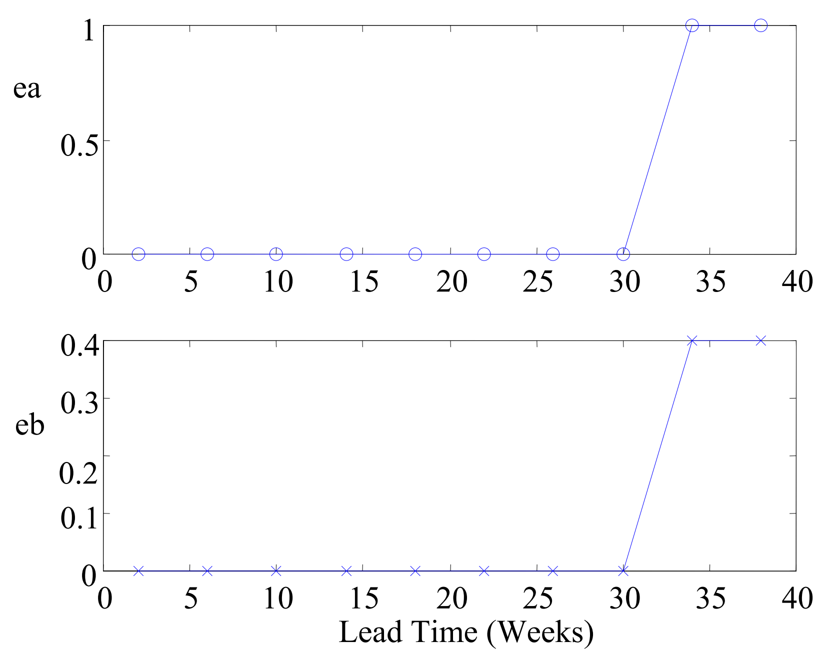

Figure 7: The Effect of Lead Time

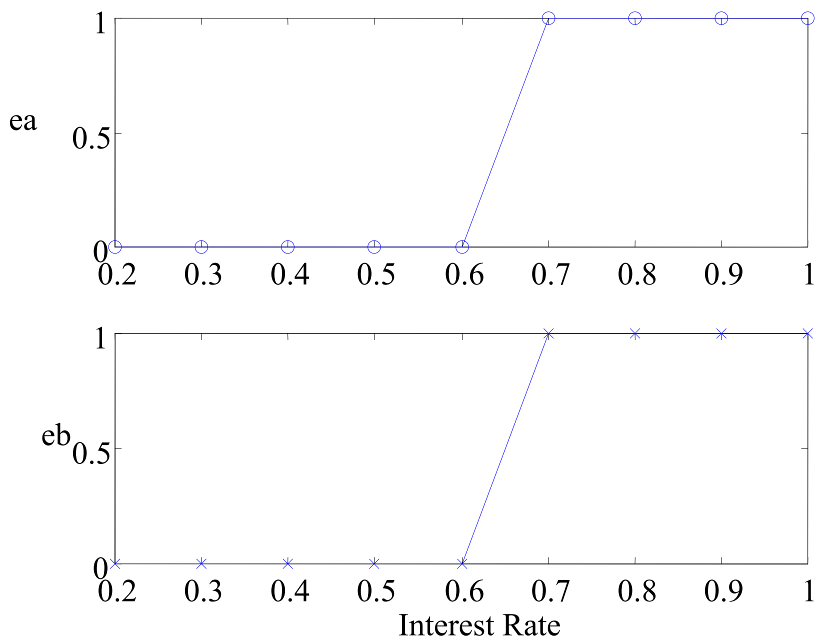

Figure 8: The Effect of Interest Rate

> ***Observation 3.7:*** *In a moderate demand variation environment, the unit shortage cost is not a significant factor in the choice of the strategies.*

To test the effect of shortage cost, we vary the unit shortage cost from $1.50 to $30.00 dollars. However, from **Figure 9**, the effect of shortage cost is not significant in this scenario. The DP strategy is always preferred no matter how high the shortage cost is. To further verify this finding, we set up another scenario by reducing D_A from 10,000 to 1,000 and D_B from 20,000 to 2,000. The PCC strategy is preferred when the unit shortage costs vary from $1.50 to $30.00. **Figure 10** again shows that the effect of shortage cost is not significant. The PCC strategy is always preferred no matter how high the shortage cost is. We therefore conclude that the unit shortage cost is not a significant factor in the choice of the strategies when the demand variation is moderate (20%).

Theoretically, when the shortage cost is high, the (*s*, *Q*) system will automatically set a higher re-order point, *s*, to prevent the expensive shortages; consequently this will increase the inventory level. Since it is more expensive to carry the common component, this phenomenon makes the DP strategy more attractive. On the other hand, when the shortage cost is high, it is also important to pool the demand variation to minimize the number of shortages. This makes the PCC strategy more attractive. From our observation, in a moderate demand variation environment, none of these two effects dominates the other; hence, the unit shortage cost becomes an insignificant factor.

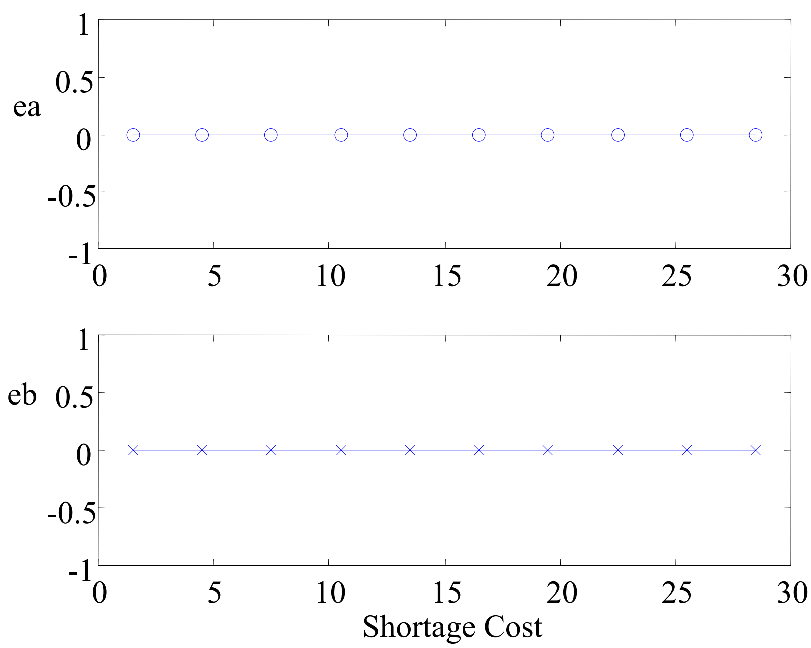

Figure 9: The Effect of Unit Shortage Cost in Moderate Demand Variation Environments (DP is preferred)

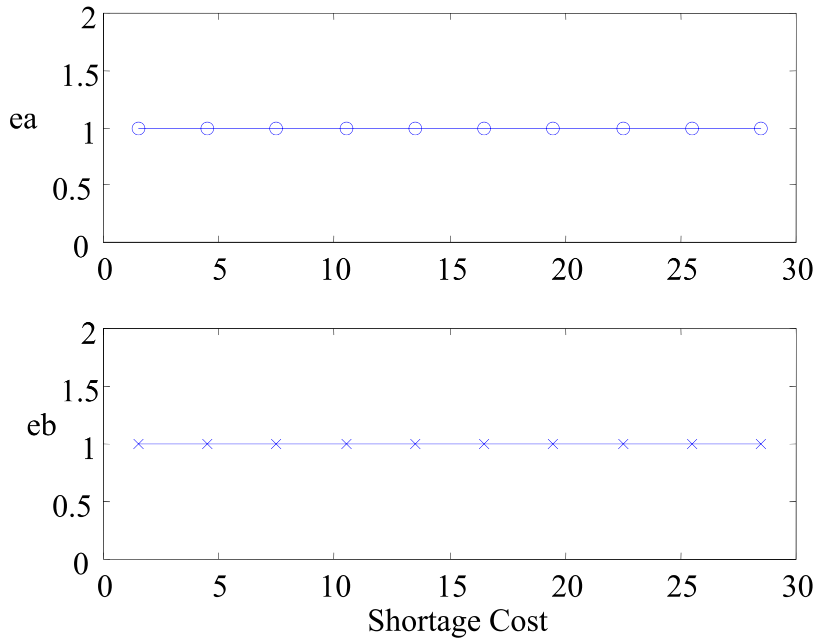

Figure 10: The Effect of Unit Shortage Cost in Moderate Demand Variation Environments (PCC is preferred)

Observation 3.8: *In a high demand variation environment, the PCC strategy is preferred when the unit shortage cost is high.*

In this experiment, we setup a high demand variation scenario (60%). When the unit shortage cost is high, the benefit of using the common component to pool demand variation outweighs the cost of higher inventory level. As shown in **Figure 11**, when the unit shortage cost is between \$4.50 and \$10.50, the MCC strategy is able to minimize the cost of product *B*. The optimal strategy switches to the PCC strategy when the unit shortage cost is larger than \$10.50. Lastly, Table 1 summarizes the results of all eight observations.

Table 1: A Summary of Observations.

Preferred Strategy	Price of Common Component	Demand Variation	Setup Cost	Lead Time	Interest Rate	Shortage Cost (moderate demand variation)	Shortage cost (high demand variation)
PCC	Low	High	High	Long	High	Not significant	High
DP	High	Low	Low	Short	Low		Low

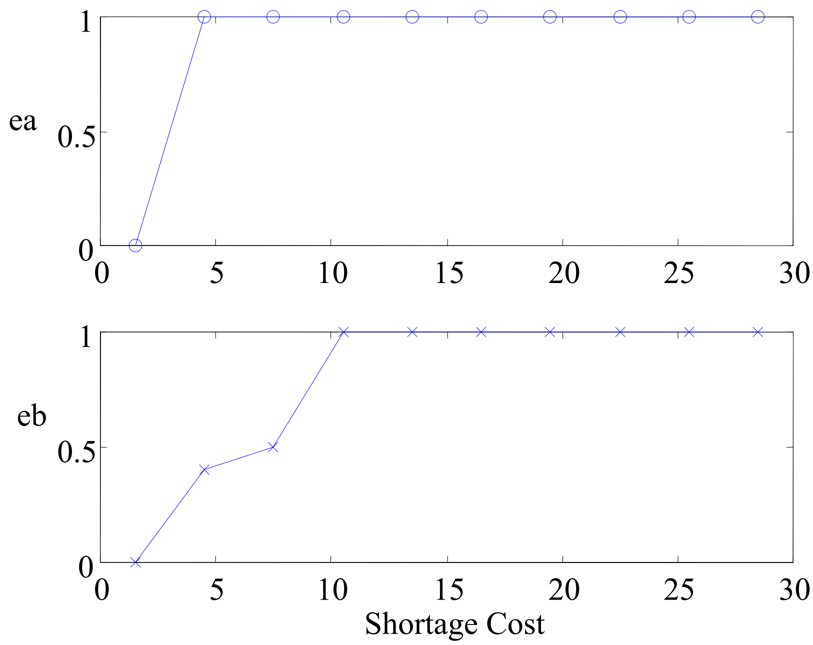

Figure 11: The Effect of Unit Shortage Cost in High Demand Variation Environments

CHAPTER 4

EVALUATION OF POSTPONEMENT STRATEGIES

To study the effectiveness of postponement strategies, we develop analytical models which present TP and FP postponement strategies under both M/M/1 and G/G/1 assumptions.

4.1 M/M/1 Models

Consider a firm that supplies a product family consisting of N different customized products. In the TP strategy, the products are manufactured in a make-to-order (MTO) fashion and shipped directly to customers from a centralized facility following the order receipts. We assume that for both TP and FP strategies, the demand arrivals and the production process both follow a random Poisson process. Thus, we model the TP strategy as a multi-class single server queuing system with exponential interarrival times and exponential service times i.e. a multi-class M/M/1 system. There are N types of customer orders (N different product types) of size one arriving at the centralized facility where the service rule is First Come First Serve. The arrival processes are assumed to be independent and the interarrival times for type k orders, $1 \le k \le N$, come from a Poisson process with a mean arrival rate of λ_k. The processing rates for all products are *iid* random variables from a Poisson distribution with a mean rate of μ. This strategy is illustrated in **Figure 12**.

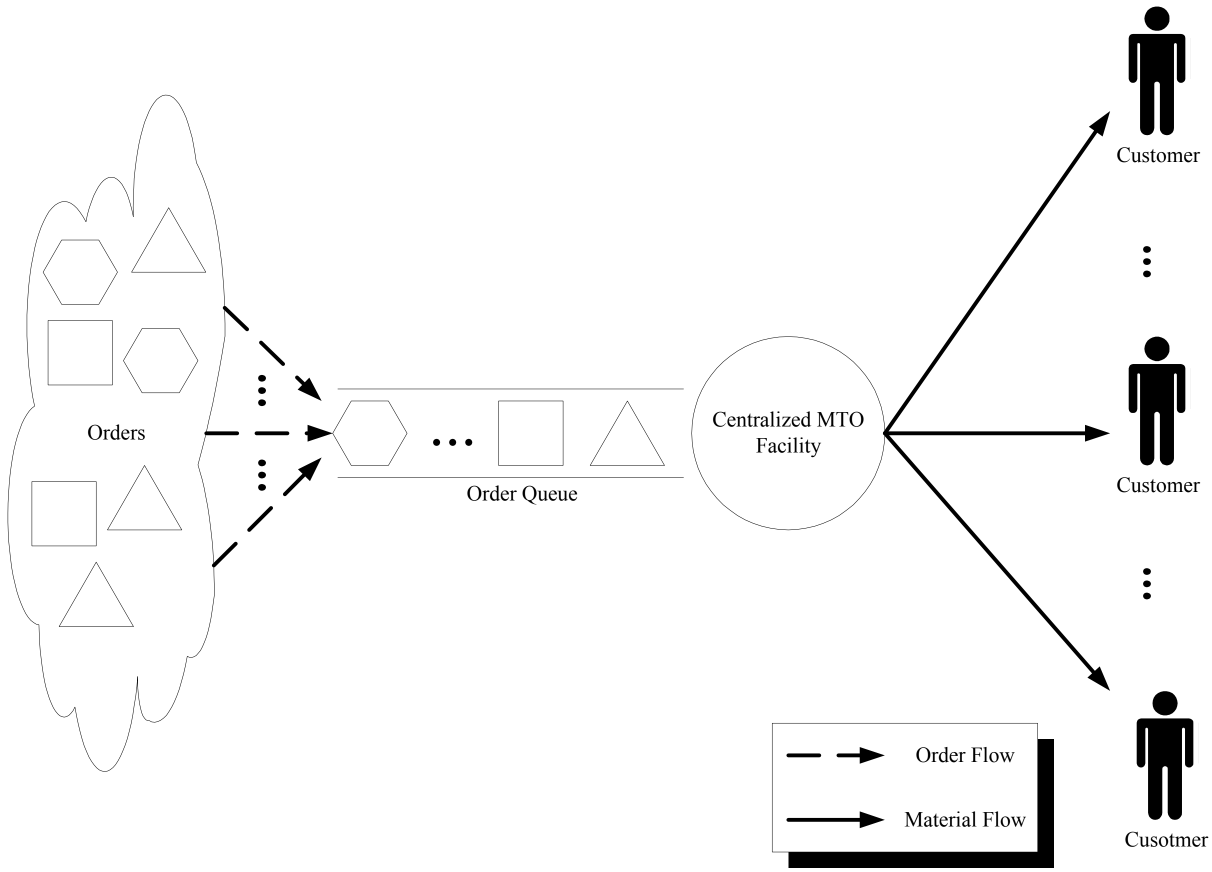

Figure 12: The TP Strategy.

The FP strategy consists of two general stages. At Stage 1, a single generic component is made to stock at a centralized facility. Thus, there is no setup cost at this stage, and a base-stock control policy is optimal for managing the generic component inventory (Zipkin 2000). Stage 1 is therefore analyzed as a single class, single server base-stock system, i.e., an M/M/1 base-stock system (Buzacott and Shanthikumar 1993). Final customizations are then made to stock at Stage 2, consisting of a dedicated facility for each of the *N* different product configurations. Our motivation for this supply chain strategy comes from the HP DeskJet Printer postponement example where the production line at a regional distribution center is dedicated to the product distributed in that region and the final product is made to stock. The FP strategy is illustrated in **Figure 13**.

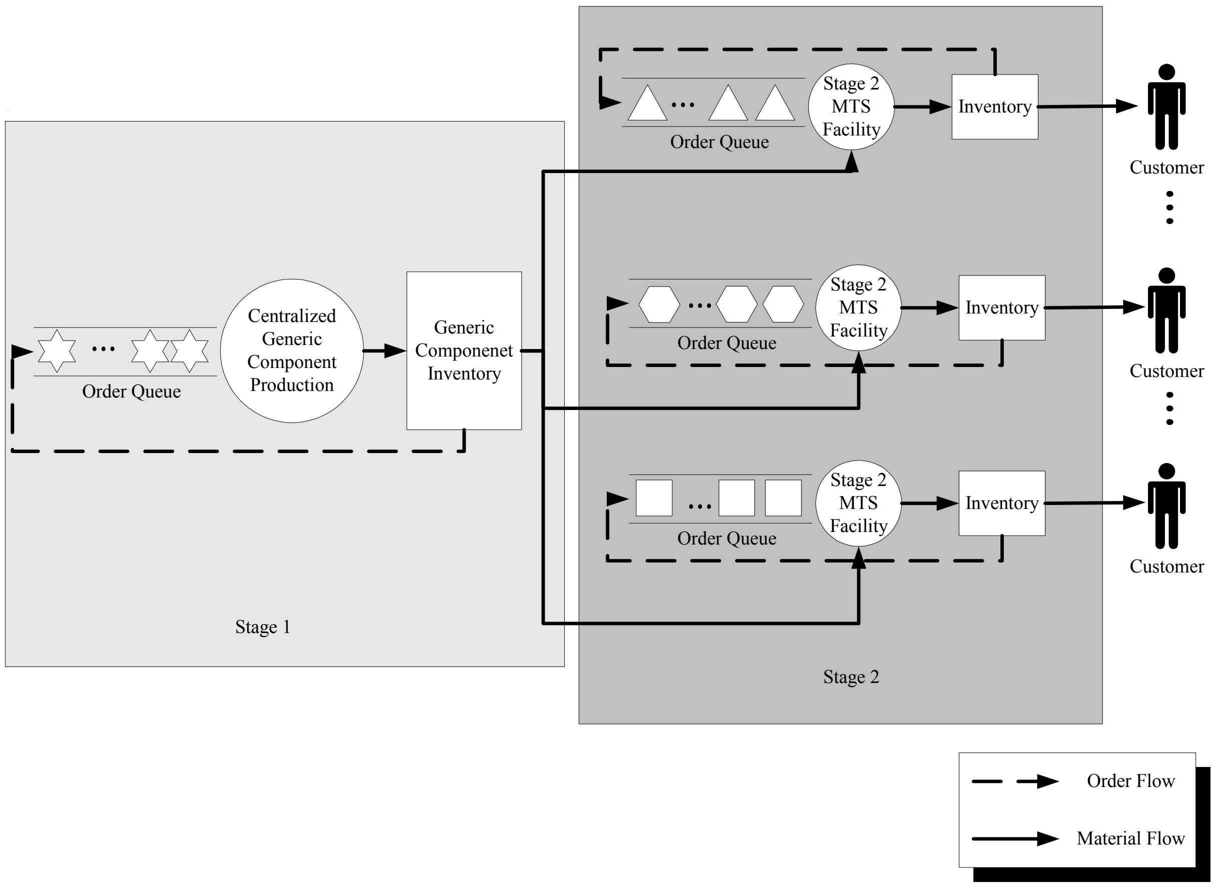

Figure 13: The FP Strategy.

For both strategies, we assume that the raw materials are always available. Because all products belong to the same product family, we also assume a common distribution for their process times along with negligible changeover times between products. For example, Dell Computer promises the same lead-time regardless of the computer configuration chosen (Dell.com). It is reasonable to assume that the cycle times required to install a larger or smaller hard drive come from the same distribution and the changeover times between assembling different configurations are minimal.

To evaluate the two supply chain strategies, two performance measurements are used: total cost (TC) and expected customer waiting time (ET). For the TP strategy, the total cost (TC_{TP}) for each period includes the amortized fixed cost (F_{TP}), the production cost, and a WIP holding cost. For the FP strategy, the Stage 1 cost ($C_{FP,\,1}$) for each period includes the production cost for the generic component and holding costs for both the WIP inventory and the finished generic component. The Stage 2 cost ($C_{FP,\,2}$) for each period includes the production cost for the final assembly as well as the associated holding costs for the WIP inventory and the finished customized products.

To implement the FP strategy, we assume the firm invests a fixed cost to develop and design the generic component and to set up machinery for both stages. This amortized fixed cost for each period is represented by F_{FP}. In general, F_{FP} is greater than F_{TP} because of the increased expense of redesigning the product for delayed differentiation.

4.1.1 The Model for the TP strategy

The expected waiting time for type k products in the TP strategy ($ET_{TP,k}$) can be derived using a birth-death process (pg. 77, Gross and Harris 1985), giving:

$$ET_{TP,k} = \frac{1}{\left(\mu - \sum_{k=1}^{N} \lambda_k\right)}. \qquad \textbf{(4.1)}$$

Because there is only one server, we use $\rho_k = {}^{\lambda_k}\!/_{\mu_k}$ to represent the average amount of WIP inventory for product k. The total cost of the TP strategy is the sum of fixed cost, WIP holding cost, and production cost as shown in Equation **(4.2)**.

$$TC_{TP} = F_{TP} + \sum_{k=1}^{N} \rho_k w_k + \sum_{k=1}^{N} \lambda_k c_k \ , \qquad \textbf{(4.2)}$$

where w_k and c_k are the unit WIP holding cost and the unit production cost for the k^{th} product respectively. We separate these two costs because the unit production cost is incurred for each unit produced but the holding cost varies with the average number of units in the system.

4.1.2 The Model for the FP Strategy

As with the TP strategy, we assume exponential interarrival times and exponential process times. Assume that orders for the generic component arrive with a mean rate of λ_g and are processed with a mean rate of μ_g. The average utilization of the Stage 1 server is then $\rho_g = {}^{\lambda_g}\!/_{\mu_g}$.

For the generic component, let h_g be the per unit holding cost, w_g be the unit holding cost of the WIP, c_g be the per unit production cost, and z_g be the base-stock level. The expected waiting time and expected inventory level for Stage 1 are based on the analysis of Buzacott and Shanthikumar (1993), pages 103-105.

The expected waiting time for Stage 1 is:

$$ET_{FP,1} = \frac{\rho_g^{\ z_g}}{\mu_g - \lambda_g} . \qquad \textbf{(4.3)}$$

Let $E[I]$ be the expected inventory level of the generic component, where

$$E[I] = z_g - \frac{\rho_g}{1-\rho_g}(1-\rho_g{}^{z_g}) \tag{4.4}$$

is made up of the base-stock level minus the expected production orders. The cost at Stage 1 is:

$$C_{FP,1} = h_g E[I] + \rho_g w_g + \lambda_g c_g \; . \tag{4.5}$$

The first term of Equation **(4.5)** is the inventory holding cost, the second term is the average WIP holding cost (for the M/M/1 system, the average WIP is equal to the utilization of the server; i.e., ρ_g), and the last term is the production cost.

At Stage 2, each product type is customized by a dedicated production line. We model this stage as *N* single-class M/M/1 base-stock systems, which are analyzed in the same way as Stage 1. For a type *k* product, let z_k be the base-stock level, λ_k be the mean arrival rate, and μ_k be the mean production rate. We assume that all product types have the same production rate since they belong to the same product family. The utilization of the server for the type *k* product is $\rho_k = {}^{\lambda_k}\!/_{\mu_k}$. The expected waiting time for product *k* at Stage 2 is:

$$ET_{FP,2,k} = \frac{\rho_k{}^{z_k}}{\mu_k - \lambda_k} \; . \tag{4.6}$$

Let $E[I_k]$ be the expected inventory level for product *k* at Stage 2, where

$$E[I_k] = z_k - \frac{\rho_k}{1-\rho_k}(1-\rho_k{}^{z_k}) \; . \tag{4.7}$$

For a type *k* product, let h_k be the unit holding cost, v_k be the unit WIP holding cost, and b_k the per unit production cost. The total cost at Stage 2 is:

$$C_{FP,2} = \sum_{k=1}^{N} \left(E[I_k] h_k + \rho_k v_k + b_k \lambda_k \right). \tag{4.8}$$

The first term of Equation **(4.8)** is the total holding cost for finished products, the second term represent the WIP holding cost, and the last term denotes the production cost. The total expected waiting time for product *k* under the FP strategy is the sum of the waiting times for both stages

$$ET_{FP,k} = ET_{FP,1} + ET_{FP,2,k} = \frac{\rho_g}{\mu_g - \lambda_g} + \frac{1}{\mu_k - \lambda_k} \qquad \textbf{(4.9)}$$

and the total cost for the FP strategy is:

$$TC_{FP} = F_{FP} + C_{FP,1} + C_{FP,2}\,. \qquad \textbf{(4.10)}$$

<u>4.1.3 Effect of Product Proliferation</u>

Product proliferation results when companies begin to customize their products for smaller customer groups or segments. In this section, we study how product proliferation affects the cost and the responsiveness of the TP and FP strategies.

We assume a constant overall utilization for the supply chain in order to isolate the impact of a change in the number of products from the impact of a change in the facility's utilization. To maintain a constant utilization, we use the throttle demand rate strategy (Gupta and Srinivansan 1998) where the total demand and total process capacity are held constant to maintain a constant system utilization rate, even though the total number of products may vary. In the absence of such control, an increase in the number of products may worsen the performance simply as a consequence of the increased load on the facility. Thus, the throttle demand rate is used to remove the effect of an increased utilization rate, allowing us to truly study the effect of increasing the number of products.

Without loss of generality, we normalize the total demand rate for the *N* products to equal 1. This allows us to simplify our notation and at the same time does not affect the performance of our models because our utilizations are adjusted accordingly. In order to isolate the effect of product proliferation, we also assume symmetric production, i.e., all the parameters of the different products are the same. Under this assumption, subscript *k* of all parameters disappears. For example, $\lambda_k=\lambda$, $\mu_k=\mu$, and $c_k=c$. Since the total demand rate for the *N* products is 1, the mean time between arrivals for each product is *N* or $\lambda = \frac{1}{N}$. In the TP strategy, setting the mean process time for each product to ρ or $\mu = \frac{1}{\rho}$ makes the system utilization rate $\sum_{k=1}^{N} \frac{\lambda_k}{\mu_k} = N \frac{1/N}{1/\rho} = \rho$.

To make a meaningful comparison between the TP and FP strategies, the same demand rate is employed for both strategies and we adjust the process times to ensure that both strategies have the same capacity. In the FP strategy, the mean time between arrivals at the generic component stage is 1 since it includes the demand of all *N* products. The mean time-between-arrivals for each product at the second stage is *N* or $\lambda = \frac{1}{N}$, since each product has a dedicated production line.

For the process time of the FP strategy, let ρ be the total process time of which a portion, $r: 0 \le r \le 1$, is consumed by the generic component. We call *r* the percentage of generic component coverage. At Stage 2, we divide the capacity available into *N* dedicated lines equally. Therefore, each line takes $N(1-r)\,\rho$ time units to finish the final customization. Based on the above assumptions and definitions, we derive the following lemmas before stating our main result on the effect of product proliferation.

Lemma 4.1*: As N increases, the expected waiting time of the TP strategy stays constant.*

Proof:

Applying the symmetric production system assumption to Equation **(4.1)**, $\lambda_k = \lambda$. The total expected waiting time for the TP strategy becomes:

$$ET_{TP} = \frac{1}{(\mu - N\lambda)} = \frac{1}{\left(\frac{1}{\rho} - 1\right)} = \frac{\rho}{1-\rho} \quad . \qquad \textbf{(4.11)}$$

Since ET_{TP} is not a function of N then as N increases, ET_{TP} stays the same.■

Lemma 4.2*: As N increases, the expected cost of the TP strategy stays constant.*

Proof:

Under symmetric production, $\lambda_k = \lambda$, $\mu_k = \mu$, and $c_k = c$. From Equation **(4.2)**, the total expected cost for the TP strategy becomes:

$$TC_{TP} = F_{TP} + N\frac{\lambda}{\mu}w + N\lambda c. \qquad \textbf{(4.12)}$$

Substituting $\lambda = \frac{1}{N}$ and $\mu = \frac{1}{\rho}$ into equation **(4.12)**, we get:

$$TC_{TP} = F_{TP} + \rho w + c. \qquad \textbf{(4.13)}$$

Thus, TC_{TP} is not a function of N. ■

Lemma 4.3*: The expected waiting time of the FP strategy increases monotonically in N.*

Proof:

Under symmetric production, from Equation **(4.9)**, the total expected waiting time for the FP strategy becomes:

$$ET_{FP} = \frac{\rho_g^{\ z_g}}{\mu_g - \lambda_g} + \frac{\left(\frac{\lambda}{\mu}\right)^z}{\mu - \lambda}. \quad \textbf{(4.14)}$$

From the throttle demand strategy and the equal capacity assumption, we get

$\rho_g = r\rho,\ \lambda_g = 1,\ \mu_g = \frac{1}{r\rho},\ \lambda = \frac{1}{N}$, and $\mu = \frac{1}{N(1-r)\rho}$. Substituting these values into Equation **(4.14)** gives:

$$ET_{FP} = \frac{(r\rho)^{z_g+1}}{1-r\rho} + \frac{N[(1-r)\rho]^{z+1}}{1-(1-r)\rho}. \quad \textbf{(4.15)}$$

Thus, as the number of products (N) increases, the expected waiting time of the FP strategy (ET_{FP}) increases monotonically. ■

Lemma 4.4*: The cost of Stage 1 of the FP strategy is constant with respect to N.*

Proof:

Substituting $\rho_g = r\rho,\ \lambda_g = 1,\ \mu_g = \frac{1}{r\rho},\ \lambda = \frac{1}{N}$, and $\mu = \frac{1}{N(1-r)\rho}$ into equation **(4.4)** gives the Stage 1 cost of the FP strategy as follows:

$$C_{FP,1} = \left[z_g - \frac{r\rho}{1-r\rho}\left(1-(r\rho)^z\right)\right]h_g + r\rho w_g + c_g. \quad \textbf{(4.16)}$$

Thus, $C_{FP,1}$ is not a function of N. ■

Lemma 4.5*: The cost of Stage 2 of the FP strategy increases monotonically in N.*

Proof:

From Equation **(4.8)**, assuming symmetric production gives

$$C_{FP,2} = NE[I]h + N\frac{\lambda}{\mu}v + Nb\lambda \qquad \textbf{(4.17)}$$

where $E[I] = z - \dfrac{\left(\frac{\lambda}{\mu}\right)}{1-\left(\frac{\lambda}{\mu}\right)}(1-\left(\frac{\lambda}{\mu}\right)^{z})$.

Substituting $\lambda = \dfrac{1}{N}$ and $\mu = \dfrac{1}{N(1-r)\rho}$ into equation **(4.17)** gives

$$C_{FP,2} = N\left[z - \frac{(1-r)\rho}{1-(1-r)\rho}\left(1-(1-r)\rho\right)^{z}\right]h + N(1-r)\rho v + b. \qquad \textbf{(4.18)}$$

Thus, as the number of products (N) increases, the Stage 2 cost of the FP strategy ($C_{FP,2}$) increases monotonically. ■

Lemma 4.6: *The total cost of the FP strategy increases monotonically in N.*

<u>Proof:</u>

Since $TC_{FP} = F_{FP} + C_{FP,1} + C_{FP,2}$, from ***Lemmas 4.4 and 4.5***, the total cost of the FP strategy (TC_{FP}) increases monotonically in N. ■

Based on the ***Lemmas*** given above, we conclude the following ***Proposition***.

Proposition: *As N increases, there exists a threshold value of N above which* $TC_{TP} < TC_{FP}$ *and* $ET_{TP} < ET_{FP}$.

<u>Proof:</u>

As N increases, ***Lemmas 4.1 and 4.2*** state that the expected waiting time (ET_{TP}) and expected cost (TC_{TP}) of the TP strategy stays constant. From ***Lemmas 4.3 and 4.6***, the expected waiting time (ET_{FP}) and expected total cost (TC_{FP}) of the FP strategy are monotonically increasing in N. Therefore, as N increases, there exists a threshold value of N above which $TC_{TP} < TC_{FP}$ and $ET_{TP} < ET_{FP}$.■

Example:

We now give a numerical example to demonstrate the result stated in the ***Proposition***. First, let $\rho = 0.5$, $r = 0.3$, $F_{TP} = 10$, and $F_{FP} = 12$. For the TP strategy, let $w = 0.05$, and $c = 1$. For Stage 1 of the FP strategy, let $z_g = 2$, $h_g = 0.03$, $w_g = 0.015$, and $c_g = 0.3$. For Stage 2 of the FP strategy, let $z = 1$, $h = 0.1$, $v = 0.065$, and $b = 0.7$. **Figure 14** is a plot of the total cost and customer waiting time as the number of products offered, N, increases from 1 to 8. An increase in the number of products greater than or equal to 6, results in the TP strategy requiring less cost *and* having a shorter customer waiting time than the FP strategy. Thus, the TP strategy dominates the FP strategy on both performance dimensions once the number of products exceeds five.

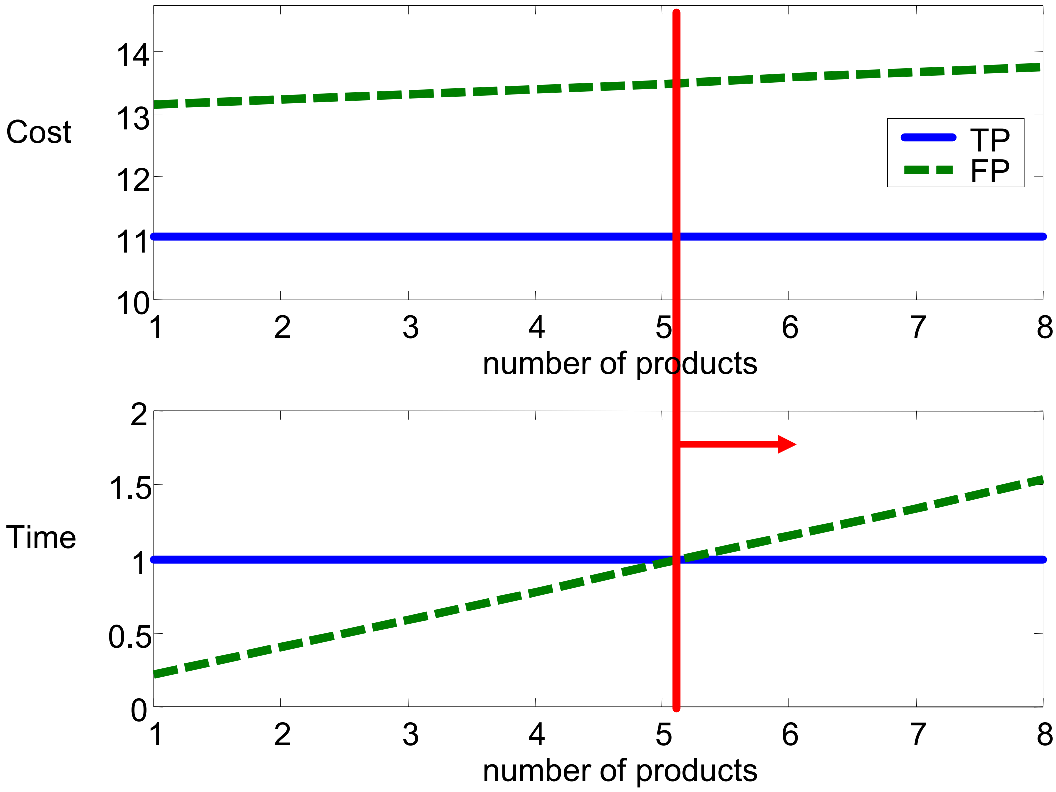

Figure 14: The Effect of Product Proliferation.

Intuitively, increasing the number of products could complicate the operation of the supply chain and worsen its performance. However, under the throttle demand and zero changeover time/cost assumptions, our analysis shows that the cost and time of the TP strategy stay constant as N increases. Gupta and Srinivasan (1998) also find that under the throttle demand assumption, there exist conditions where increasing the number of products decreases the number of back-orders and thus, reduces the expected customer waiting time.

In our models, increasing the number of products in the FP strategy requires that the capacity at Stage 2 be divided into equal amounts for each dedicated line. Thus, the pooling effect is lost in the FP strategy (resulting in an increase in the WIP and final product inventory levels as well as the average number of back-orders) but not in the TP strategy where all capacity is centralized. Therefore, as N increases the cost and time of the FP strategy increases. We expect that the proposition will still hold under a small positive setup time/cost for the TP strategy but the proof and specifications are left for future research.

4.2 G/G/1 Models

The M/M/1 models provide a means to study the effect of product proliferation on the choice between the TP and FP strategies. However, more detail is needed to perform sensitivity analysis on how changes in the interarrival time variation and process time variation affect the choice of strategy. The exponential distribution only has one parameter that determines both the mean and the variance. It does not allow us to change the variance without changing the mean. For this reason, we also model our supply chain strategies using G/G/1 queuing systems.

4.2.1 The G/G/1 Approximation of the TP Strategy

Let λ and μ be the mean arrival rate and mean process rate for a product, $\rho = \dfrac{N\lambda}{\mu}$ be the overall system utilization, and σ_d^2 and σ_p^2 be the variances of the interarrival times and process times respectively. There are several G/G/1 approximations available (Marchal 1976, and Shore 1988). We tested both and found their results to be very close. To our knowledge, there is no published study stating that one approximation is better than another. Therefore, we use the one from Shore (1998). From Shore's approximation, the expected number of customers in the system is:

$$L = \left\{ \frac{\rho^2 (1 + \mu^2 \sigma_p{}^2)}{1 + \rho^2 \mu^2 \sigma_p{}^2} \right\} \left\{ \frac{\lambda^2 \sigma_d{}^2 + \rho^2 \mu^2 \sigma_p{}^2}{2(1-\rho)} \right\} + \rho \,. \tag{4.19}$$

By applying Little's law, the expected customer waiting time using the TP strategy is:

$$ET_{TP} = \frac{1}{N\lambda} \left\{ \left\{ \frac{\rho^2 (1 + \mu^2 \sigma_p{}^2)}{1 + \rho^2 \mu^2 \sigma_p{}^2} \right\} \left\{ \frac{\lambda^2 \sigma_d{}^2 + \rho^2 \mu^2 \sigma_p{}^2}{2(1-\rho)} \right\} + \rho \right\} \,. \tag{4.20}$$

Similar to the M/M/1 model, the total cost of the TP strategy is:

$$TC_{TP} = F_{TP} + \rho w + N\lambda c \,. \tag{4.21}$$

4.2.2 The G/G/1 Approximation of the FP Strategy

The G/G/1 model for the FP strategy is based on the approximation developed by Buzacott and Shanthikumar (1993) on page 106. Like the M/M/1 model, we first analyze Stage 1 where the production and stocking of the generic component occurs.

Stage 1

Let λ_g and μ_g be the mean arrival rate and mean processing rate for the generic component, and $\sigma_{d,g}{}^2$ and $\sigma_{p,g}{}^2$ be variances of the interarrival times and process times. Define $\rho_g = \frac{\lambda_g}{\mu_g}$ to be the utilization of Stage 1, and following the notation of Buzacott and Shanthikumar (1993), let

$$\Lambda = \frac{L^* - \rho_g}{L^*} \tag{4.22}$$

where $$L^* = \left\{ \frac{\rho_g{}^2 (1 + \mu_g{}^2 \sigma_{p,g}{}^2)}{1 + \rho_g{}^2 \mu_g{}^2 \sigma_{p,g}{}^2} \right\} \left\{ \frac{\lambda_g{}^2 \sigma_{d,g}{}^2 + \rho_g{}^2 \mu_g{}^2 \sigma_{p,g}{}^2}{2(1-\rho_g)} \right\} + \rho_g \,. \tag{4.23}$$

The expected number of backorders of the generic component, $E[B]$, is:

$$\mathrm{E[B]} = \sum_{n=1}^{\infty} n\rho_g (1-\Lambda)\Lambda^{n-1+z} = \frac{\rho_g \Lambda^{z_g}}{(1-\Lambda)} . \qquad \textbf{(4.24)}$$

The expected waiting time for Stage 1, using Little's Law, is:

$$ET_{FP,1} = \frac{1}{\mu_g} E[B] . \qquad \textbf{(4.25)}$$

Let *E[I]* be the expected inventory level:

$$E[I] = \sum_{n=1}^{z_g-1} \left\{ n(1-\rho_g)\rho_g \Lambda^{z_g-n-1} \right\} + (1-\rho_g) z_g \qquad \textbf{(4.26)}$$

$$= \frac{(1-\rho_g)\rho_g \Lambda^{z_g-2} \left\{ 1 - z_g \left(\frac{1}{\Lambda} \right)^{z_g-1} + (z_g-1) \left(\frac{1}{\Lambda} \right)^{z_g} \right\}}{\left(1 - \frac{1}{\Lambda} \right)^2} + (1-\rho_g) z_g . \qquad \textbf{(4.27)}$$

Similar to **(4.4)**, the cost of Stage 1 is:

$$C_{FP,1} = E[I] h_g + \rho_g w_g + \lambda_g c_g . \qquad \textbf{(4.28)}$$

<u>Stage 2</u>

Let λ and μ' be the mean arrival rate and mean process rate for a product at each dedicated line of Stage 2 and σ_d^2 and σ_p^2 be the variances of the interarrival times and process times. Similar to Stage 1, the approximate expected waiting times for a product at Stage 2 is:

$$ET_{FP,2} = \frac{\rho \Lambda^z}{\mu'(1-\Lambda)} . \qquad \textbf{(4.29)}$$

where $\rho = \frac{\lambda}{\mu'}$, $\Lambda = \frac{L^* - \rho}{L^*}$, and

$$L^* = \left\{ \frac{\rho^2 (1 + \mu'^2 \sigma_p^{\ 2})}{1 + \rho^2 \mu'^2 \sigma_p^{\ 2}} \right\} \left\{ \frac{\lambda^2 \sigma_d^{\ 2} + \rho^2 \mu'^2 \sigma_p^{\ 2}}{2(1-\rho)} \right\} + \rho .$$

Similar to **(4.7)**, the cost of Stage 2 is:

$$C_{FP,2} = N\left(E\left[I\right]h + \rho v + b\lambda\right). \qquad \textbf{(4.30)}$$

Where

$$E[I] = \sum_{n=1}^{z-1}\left\{n(1-\rho)\rho\Lambda^{z-n-1}\right\} + (1-\rho)z$$

$$= \frac{(1-\rho)\rho\Lambda^{z-2}\left\{1 - z\left(\frac{1}{\Lambda}\right)^{z-1} + (z-1)\left(\frac{1}{\Lambda}\right)^{z}\right\}}{\left(1-\frac{1}{\Lambda}\right)^{2}} + (1-\rho)z$$

The total expected waiting time for a product under the FP strategy is

$$ET_{FP} = ET_{FP,1} + ET_{FP,2} \qquad \textbf{(4.31)}$$

and the total cost is

$$TC_{FP} = F_{FP} + C_{FP,1} + C_{FP,2}\ . \qquad \textbf{(4.32)}$$

4.3 Strategy Comparisons

Using the G/G/1 approximations, we can evaluate the impact of customer arrival time variation and process time variation on the cost and the waiting time of the two strategies. To compare the total costs and waiting times, we substitute in the parameter values for a particular scenario into equations **(4.19)** to **(4.32)** to see which strategy provides the lowest cost and/or shortest waiting time for those particular data values. However, there are many factors affecting the costs and waiting times and isolating the effect of each one analytically is intractable. Therefore, we design a numerical experiment to provide a comparison of the two strategies under a wide range of parameter values.

Six factors are included in the experiment: utilization rate, arrival time variation, process time variation, interest rate, percentage of generic component coverage, and the number of products. We choose the interest rate as a factor and assume that the holding costs for both the generic component and the WIP inventory are directly proportional to the interest rate. Any change in the interest rate changes the two holding cost proportionately, thus allowing us to analyze their impact through the use of a single factor. There are other factors that are not included because their effects on the costs and waiting times are straightforward. These include the fixed cost, the production cost, and the base-stock levels.

We measured performance through changes in the total costs and the expected waiting times. Initial tests showed that the utilization rate had the greatest impact of all the model parameters and significantly confounded the effects of the other factors. Therefore, we created three separate experiments corresponding to low, medium, and high utilization rates. Each experiment covers the other five factors, at three levels for each factor. A full factorial experiment would require 3^5 different runs. By using the Taguchi's $L_{18}(2^1 \text{ x } 3^7)$ orthogonal array (Phadke 1989), the number of experimental runs is significantly reduced to only 18 for each experiment. Taguchi's $L_{18}(2^1 \text{ x } 3^7)$ orthogonal array is designed to test the significance of up to eight different variables. Since our experiment only has five factors in each experiment, the remaining three are set as dummy factors.

Each factor has three levels: low, medium, and high. Values were chosen to cover the ranges of most realistic scenarios. The values selected for each factor are summarized at **Table 2**.

Table 2: Factors and Their Level Values.

Factors	Levels		
	Low	Medium	High
Utilization rate	0.5	0.75	0.97
Coefficient of variation of times between arrivals	10%	200%	400%
Coefficient of variation of process times	10%	50%	100%
Interest rate	5%	25%	50%
Percentage of generic component coverage (r)	30%	50%	70%
Number of products	2	8	16

Based on the factor levels, the other model parameters such as the demand rate (λ), process rate (μ), holding cost (h), and WIP cost (w) were set according to the throttle demand rate, symmetric production, and equal capacity assumptions described in Section 2.3. The two response variables are the expected customer waiting times given by **(4.18)** and **(4.31)** and the total costs given by **(4.19)** and **(4.32)**. ANOVAs were performed to test the significance of the factors and the results are summarized in **Table 3**.

Table 3: Results of ANOVA Analysis.

	Cost of FP			Time of FP			Cost of TP			Time of TP		
Utilization	L	M	H	L	M	H	L	M	H	L	M	H
Arrival time variation				↑(a) 0.030						↑(b) 0.000	↑(b) 0.000	↑(b) 0.000
Process time variation										↑(b) 0.000	↑(b) 0.000	↑(b) 0.000
Interest rate	↑(b) 0.001*	↑(b) 0.002	↑(b) 0.002				↑(b) 0.000	↑(b) 0.000	↑(b) 0.000			
Percentage of coverage				↓(a) 0.017	↓(a) 0.021	↓(a) 0.028						
Number of products	↑(b) 0.006	↑(b) 0.006	↑(b) 0.006	↑(a) 0.016	↑(a) 0.026	↑(a) 0.042						

↑: An increase in factor increases the value of the output
↓: An increase in factor decreases the value of the output
(a): P value less than 0.05 but greater than 0.01
(b): P value less than 0.01
* : Numbers represent p values

From the results in **Table 3**, we make observations to describe the impact of the design factors on strategy performances. The first four observations are intended to inform firms of the environments where one supply chain strategy may be more attractive than the other.

Observation 4.1*: Under medium to high utilization levels, higher arrival time variation significantly increases the expected waiting time of the TP strategy (ET_{TP}) but not that of the FP strategy (ET_{FP}).*

An increase in the arrival time variation significantly increases the expected waiting time of the TP strategy at all utilization levels but only increases the expected waiting time of the FP strategy at low utilization levels. The FP strategy is more robust to the increases of arrival time variation. This is due to the stock of generic components that provide a buffering effect. Thus, as the inter-arrival times become more variable, the FP strategy becomes more attractive.

Observation 4.2*: Higher process time variation significantly increases the expected waiting time of the TP strategy (ET_{TP}) but not that of the FP strategy (ET_{FP}).*

More variability in the process times significantly increases the expected waiting time of the TP strategy but has no significant impact on that of the FP strategy. Similar to ***Observation 1***, the FP strategy is more robust to increases in the variation because of its generic component inventory buffer. Hence, an increase in the process time variability makes the FP strategy more attractive. Combining ***Observations 1*** and ***2*** suggests that companies in a highly uncertain environment should consider an FP strategy.

Observation 4.3*: A higher percentage of generic component coverage (r) significantly decreases the expected waiting time of the FP strategy (ET_{FP}).*

A higher percentage of generic component coverage significantly reduces the expected waiting time of the FP strategy but has no significant effect on its cost. The magnitude of the reduction is increasing in the utilization rate. Hence, a company seeking to reduce its customer waiting times and operating in an FP strategy under high utilizations should consider delaying the differentiation of its products as long as possible. Of course, the reduction in waiting times must be balanced against the possible increase in the fixed cost for redesigning the product or process. The result is shown in **Figure 15**.

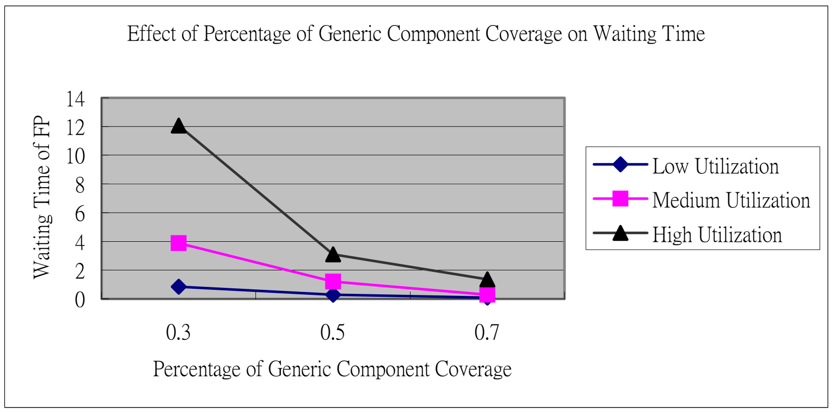

Figure 15: Effect of Generic Component Coverage on Waiting Time of FP.

> ***Observation 4.4***: *Increasing the number of products (N) significantly increases both the cost and the expected waiting time of the FP strategy but not those of the TP strategy.*

Increasing the number of products significantly increases the expected waiting time and expected cost of the FP strategy but has no significant impact on the TP strategy. Hence, increasing the number of products makes the TP strategy more attractive under both performance metrics. This result, shown in **Figure 16,** is consistent with our analysis of product proliferation (stated in the ***Proposition***) using the M/M/1 models.

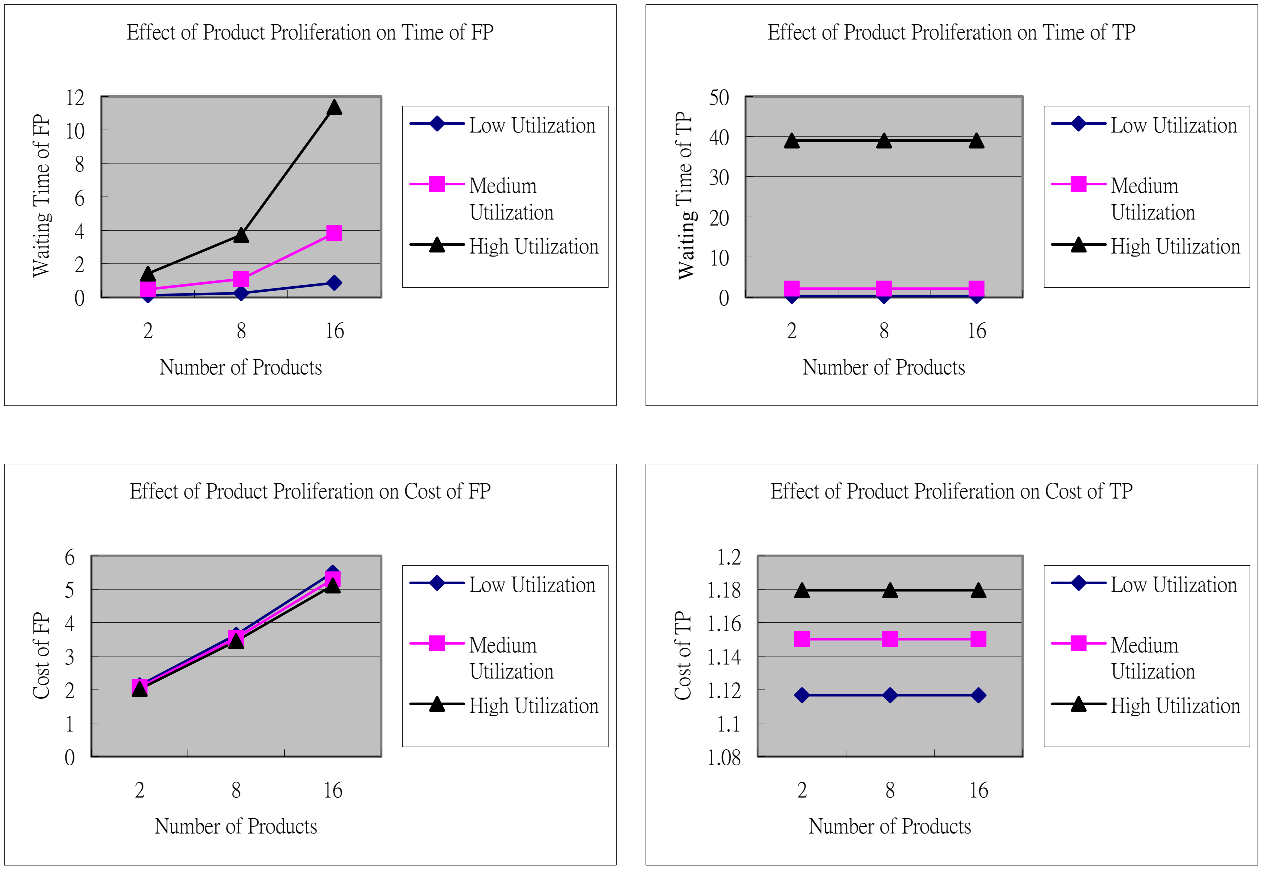

Figure 16: Effect of Product Proliferation.

Some companies may already be established in a particular supply chain strategy and are more interested in improving the performance of the strategy they currently have. The next observation provides guidance to firms, operating in an FP strategy, seeking to allocate resources in order to improve performance.

> ***Observation 4.5****: In the FP strategy, increasing the percentage of the generic component coverage (r) and reducing the number of products (N) significantly improve the expected waiting times. In contrast, lowering the arrival time and process time variations does not significantly improve the waiting times except when the utilization level is low.*

Increasing the generic component coverage and reducing the number of different products offers significant improvement to the customer waiting times of the FP strategy. Alternatively, a decrease in the arrival time variation does not significantly improve the waiting times except under low utilization levels. A decrease in the process time variation does not significantly improve the waiting times under any utilization level. Hence, if a firm desires to improve its responsiveness in the FP strategy, allocating resources to product improvement (i.e., increasing *r* by a better designed generic component or reducing *N* by designing a product suitable for multiple market segments/regions) is more effective than process improvement (i.e., reducing the variation in arrival times and process times).

The last observation compares the robustness of the two strategies to external shocks (in our case, increases in the firms' interest rates).

> ***Observation 4.6****: Increases in interest rate significantly increases both the cost of the TP and FP strategies. However, it has a larger impact on the FP strategy than on the TP strategy.*

An increase in the interest rate significantly impacts the costs of both the TP and FP strategies. However, by investigating the detailed results of its effect as shown in **Figure 17**, we see that the slope of the cost increase for the FP strategy is greater than the slope of the cost increase for the TP strategy. This is because the FP strategy (a make-to-stock strategy) has more inventory than the TP strategy (a make-to-order strategy). Therefore, higher interest rates make the TP strategy more favorable.

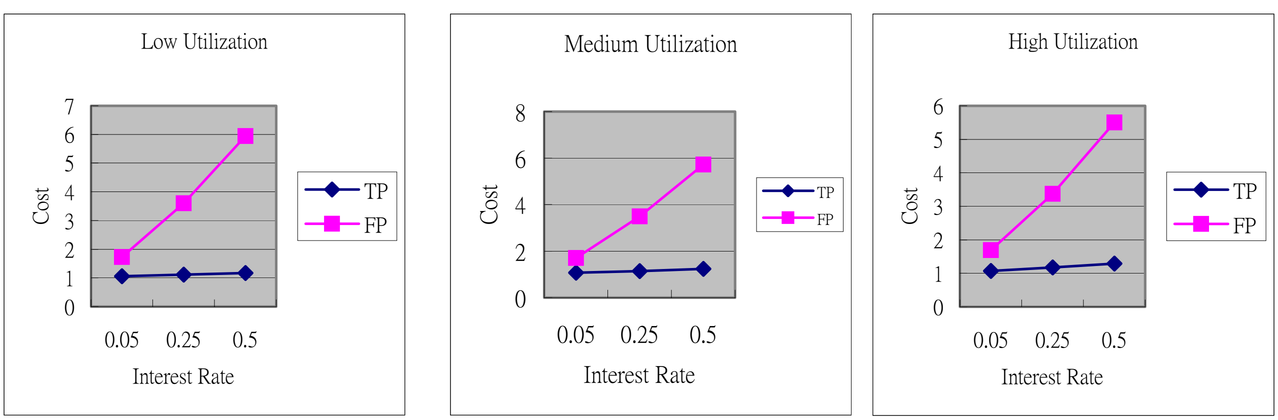

Figure 17: The Effect of Interest Rate on Strategy Costs.

CHAPTER 5

THE IMPACT OF DELAYED DIFFERENTIATION IN MAKE-TO-ORDER ENVIRONMENTS

To study the impact of delayed differentiation in make to order environments, we develop two queuing models to present the MTO and CTO strategies and derive the conditions under which one is better than the other. In this section, the models for the MTO and CTO strategy are developed. A simulation study is performed to validate the approximations and its result will be present. Last, the managerial insights derived from our analysis are discussed.

51. Supply Chain Models

Consider a firm that supplies a product family consisting of *N* different customized products. To accomplish this, the firm usually operates its production process in some form of make-to-order fashion. We consider two strategies the firm may choose: a CTO strategy and a pure MTO strategy.

In the CTO strategy, production is split into two stages, the generic stage (Stage 1) and the final customization stage (Stage 2). At Stage 1, the generic component manufacturing is triggered by the production orders at Stage 2 and is produced at a centralized facility. Since there is only one component produced at Stage 1, we assume that the line remains set up for the component ensuring that any setup cost is negligible. Under this condition, a base-stock policy is the optimal control mechanism to manage the inventory of the generic component (Zipkin 2000). Final customizations take place at Stage 2 where the products, once completed, are sent directly to the customers. We define the degree of delayed differentiation ($r:\ 0 \le r \le 1$) as the percentage of the total process time consumed at Stage 1. Thus, a value of $r = 0$ implies a pure make-to-order environment while a value of r approaching 1 implies that differentiation occurs at the last possible state. The CTO strategy is illustrated in **Figure 18**.

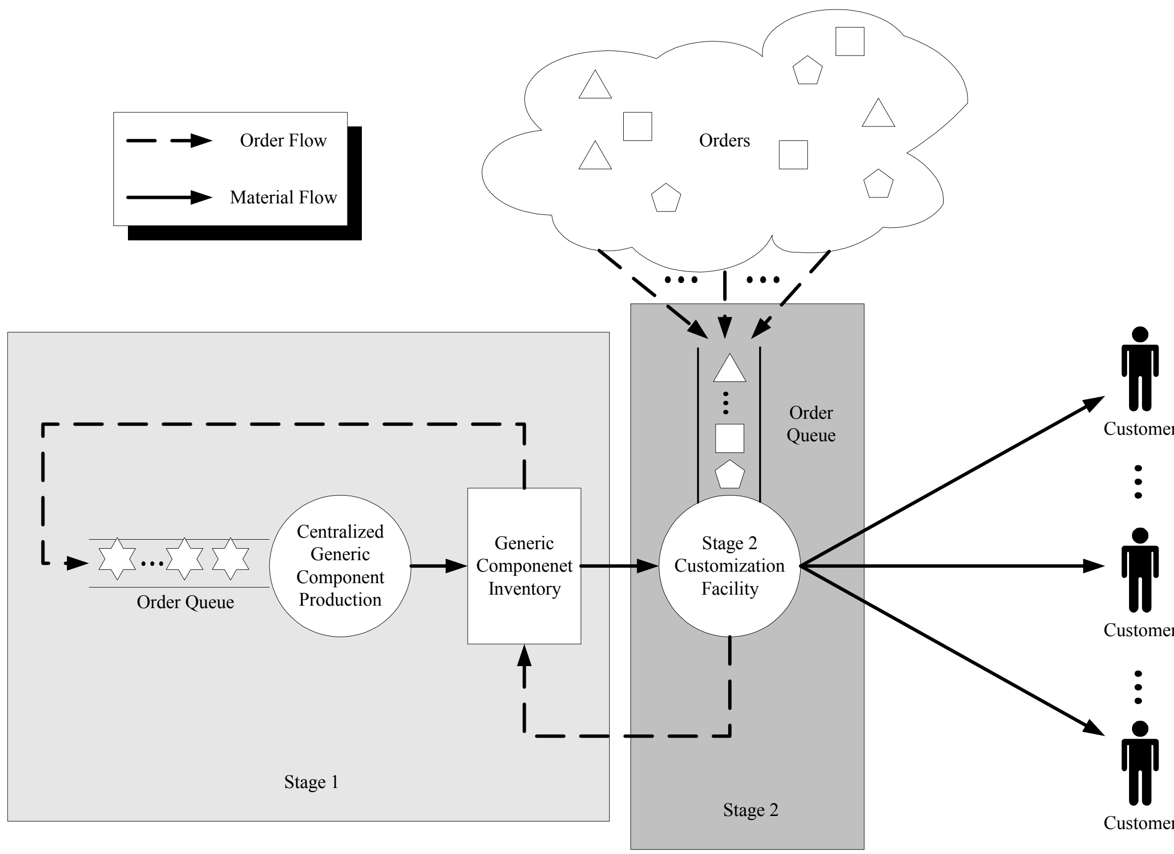

Figure 18: The CTO Strategy.

In the special case of the pure MTO strategy ($r = 0$), we assume all product types are manufactured at the centralized facility and shipped directly to customers following the order receipt. This strategy is illustrated in **Figure 19**.

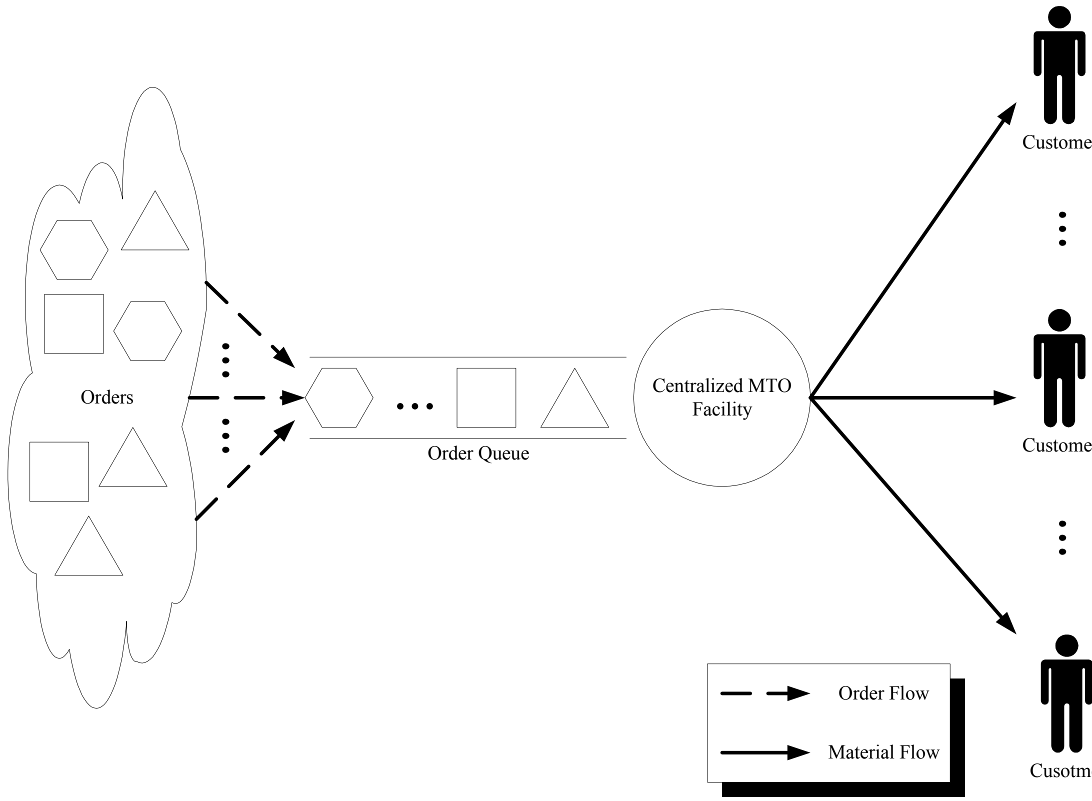

Figure 19: The MTO Strategy.

For all strategies, we assume that the needed raw materials are always available. Because all products belong to the same product family, we assume a common distribution for their process times along with negligible changeover times between products. This assumption is suitable for many business applications. For example, Dell Computer promises the same lead-time regardless of the computer configuration that the customer chooses, implying that the process times for different computer configurations are similar. It is reasonable to assume that the times to put in a larger or smaller hard drive will have the same distribution and that the change-over times between assembling different configurations are minimal. Another example is **idtown.com** (McCarthy 2000), who allows their customers to choose their watches in different colors, dials, arms, bodies, and bands. Despite the large number of variations, identical lead-times are quoted for any configuration.

To evaluate the two supply chain strategies, two performance measurements are used: total cost (*TC*) and expected customer waiting time (*ET*). For the CTO strategy, the per-period cost at Stage 1 ($C_{CTO,\,1}$) includes the production cost and holding costs for the WIP and the finished generic component inventory. At Stage 2, the per-period cost ($C_{CTO,\,2}$) includes the production and holding cost for the WIP inventory.

To incorporate delayed differentiation in the CTO strategy, we assume that the firm invests a fixed cost to develop and design the generic component and to set up machinery for both stages. This amortized fixed cost for each period is represented by the parameter F_{CTO}. The total cost of the CTO strategy (TC_{CTO}) is the sum of $C_{CTO,\,1}$, $C_{CTO,\,2}$, and F_{CTO}. For the special case of the pure MTO strategy, the total cost per period (TC_{MTO}) includes the amortized fixed cost (F_{MTO}) and the custom production stage of the CTO model, $C_{CTO,2}$.

5.1.1 The Model for the CTO Strategy

Based on the queuing strategy shown in **Figure 18**, we model the CTO strategy as a queuing system. Since the departure process out of Stage 1 is not a renewal process (Bai, Liu, Serfozo, and Shang 2003), the exact expressions for the expected customer waiting times and expected inventory levels of the generic component are not mathematically traceable. Hence, we offer approximations for these measures by analyzing the two stages of the CTO strategy independently and later verify the approximations by comparing them against the results of a simulation study.

Stage 1 is analyzed as a single class, single server, base-stock system with exponential inter-arrival times and exponential service times, i.e., a base stock M/M/1 system if the demand arrival pattern and the production process both follow Poisson processes. Since Stage 1 produces a single product and there is no setup cost, a base-stock control policy is optimal for managing the inventory (Zipkin 2000). Let the demand of the generic component arrive with a mean rate of λ_g and the processing rate to produce the generic component has a mean of μ_g. The utilization of the server producing the generic component is $\rho_g = {}^{\lambda_g}\!/_{\mu_g}$.

Our analysis of the M/M/1 base-stock system is grounded on the model developed by Buzocott and Shanthikumar (1993). For the generic component, let h_g be the unit holding cost, w_g be the unit cost of the WIP inventory, c_g be the unit production cost, and z be the base-stock level. The expected waiting time at Stage 1 is

$$ET_{CTO,1} = \frac{\rho_g^{\;Z}}{\mu_g - \lambda_g}. \qquad \textbf{(5.1)}$$

The cost at Stage 1 is

$$C_{CTO,1} = h_g E[I] + \rho_g w_g + \lambda_g c_g , \qquad \textbf{(5.2)}$$

where $E[I]$ is the expected inventory level given by:

$$E[I] = z - \frac{\rho_g}{1-\rho_g}(1-\rho_g^{\;z}). \qquad \textbf{(5.3)}$$

The first term of Equation **(5.2)** is the holding cost for the finished generic component and the second term is the holding cost for the WIP inventory. Since there is only one server in the system, the long run average number of WIP units is equal to the time the server is busy; i.e. ρ_g. The last term represent the production cost.

Since all the products are customized at the centralized facility, Stage 2 is modeled as a multiple class M/M/1 system. Orders for a type *k* product arrive at Stage 2 independently with a mean rate of λ_k and are served according to the First Come First Serve rule. The processing rates of all the products follow the same distribution and have a mean rate of μ. The utilization of the server producing product type *k* is then $\rho_k = \lambda_k / \mu$. Finally, let w_k be the unit holding cost of the WIP inventory and c_k be the unit production cost of a type *k* product.

The expected waiting time at Stage 2 for product type *k*, derived using a birth-death process, is

$$ET_{CTO,2,k} = \frac{1}{\left(\mu - \sum_{k=1}^{N} \lambda_k\right)}. \tag{5.4}$$

The total cost at Stage 2 is

$$C_{CTO,2} = \sum_{k=1}^{N} (\rho_k w_k) + \sum_{k=1}^{N} (c_k \lambda_k), \tag{5.5}$$

where the first term is the holding cost of the WIP inventory and the second term is the production cost.

The total expected customer waiting time for product *k* under the CTO strategy is

$$ET_{CTO,k} = ET_{CTO,1} + ET_{CTO,2,k} = \frac{\rho_g^{\ z}}{\mu_g - \lambda_g} + \frac{1}{\left(\mu - \sum_{k=1}^{N} \lambda_k\right)} \quad . \tag{5.6}$$

The total cost for the CTO strategy can now be expressed as

$$\begin{aligned} TC_{CTO} &= F_{CTO} + C_{CTO,1} + C_{CTO,2} \\ &= F_{CTO} + h_g E\left[I\right] + \rho_g w_g + \lambda_g c_g + \sum_{k=1}^{N} (\rho_k w_k) + \sum_{k=1}^{N} (c_k \lambda_k). \end{aligned} \tag{5.7}$$

Since the demand arrival rate for type k product is λ_k, the total demand of all the products is

$$D = \sum_{k=1}^{N} \lambda_k \ . \qquad \textbf{(5.8)}$$

The demand rate of the generic component λ_g is the combination of all the demands of the different products, i.e.

$$\lambda_g = D = \sum_{k=1}^{N} \lambda_k \ . \qquad \textbf{(5.9)}$$

Let S represent the total process time for both Stage 1 and Stage 2. Since the portion of the total process consumed by the generic component is r, the process time at Stage 1 is rS and the process rate is

$$\mu_g = \frac{1}{rS} \ . \qquad \textbf{(5.10)}$$

The process time at Stage 2 is $(1-r)S$ and the process rate is

$$\mu = \frac{1}{(1-r)S} \ . \qquad \textbf{(5.11)}$$

Combining **(5.9)** and **(5.10)** gives the utilization at Stage 1 as

$$\rho_g = \frac{\lambda_g}{\mu_g} = rDS \ . \qquad \textbf{(5.12)}$$

Substituting **(5.9)** through **(5.12)** into **(5.6)** and **(5.7)** gives an expected waiting time of

$$ET_{CTO,k} = \frac{(rDS)^z}{\dfrac{1}{rS} - D} + \frac{1}{\dfrac{1}{(1-r)S} - D} \qquad \textbf{(5.13)}$$

and a total cost of

$$TC_{CTO} = F_{CTO} + h_g E[I] + rS\sum_{k=1}^{N}\lambda_k w_g + \sum_{k=1}^{N}\lambda_k c_g + \sum_{k=1}^{N}\left((1-r)S\lambda_k w_k + \lambda_k c_k\right). \quad \textbf{(5.14)}$$

To verify the accuracy of the approximations, we compare the expected waiting time calculated by **(5.13)** and the expected inventory of generic component calculated by **(5.3)** against the results from a simulation study. We design an experiment by varying three factors (*r*, *z* and the utilization level) by three levels each, giving a total of 27 different settings. For each setting, we run the simulation 30 times and calculate the average customer waiting time and the average inventory after observing 10,000 demand arrivals. The differences between the expected waiting times calculated by the simulation and the times calculated by **(5.13)** are less than 1 percent for most of the settings and are no more than 3.6 percent. The differences between the expected inventory levels of the generic component calculated by the simulation and the levels calculated by **(5.3)** are close to zero for most of the settings and no more than 1.2 percent. The results of the experiment are shown in **Table 4**.

Table 4: Comparison of Analytical Approximation with Simulation.

r	Utilization level	*z*	ET_{CTO} from Simulation	ET_{CTO} from equation (13)	Percent Differences*	E[I] from Simulation	E[I] from equation (3)	Percent Differences**
0.1	0.1	0	0.1093	0.1090	0.2752	0	0	0
0.1	0.5	0	0.8740	0.8708	0.3675	0	0	0
0.1	0.9	0	4.4254	4.3621	1.4511	0	0	0
0.5	0.1	2	0.0528	0.0528	0	1.9474	1.9475	0.0051
0.5	0.5	2	0.3480	0.3542	1.7504	1.6869	1.6875	0.0356
0.5	0.9	2	0.9491	0.9839	3.5369	1.3430	1.3475	0.3340
0.99	0.1	5	0.0010	0.0010	0	4.8900	4.8901	0.0020
0.99	0.5	5	0.0334	0.0342	2.3392	4.0471	4.0489	0.0445
0.99	0.9	5	4.5808	4.5994	0.4044	1.3997	1.4160	1.1511
0.1	0.1	5	0.0987	0.0989	0.2022	4.9899	4.9899	0
0.1	0.5	5	0.8198	0.8182	0.1956	4.9474	4.9474	0
0.1	0.9	5	4.3139	4.2632	1.1892	4.9008	4.9011	0.0061
0.5	0.1	0	0.1050	0.1053	0.2849	0	0	0
0.5	0.5	0	0.6661	0.6667	0.0900	0	0	0
0.5	0.9	0	1.6338	1.6364	0.1589	0	0	0
0.99	0.1	2	0.0021	0.0021	0	1.8911	1.8912	0.0052
0.99	0.5	2	0.2468	0.2452	0.6525	1.2608	1.2600	0.0635

0.99	0.9	2	6.3869	6.4985	1.7173	0.3158	0.3151	0.2222
0.1	0.1	2	0.0989	0.0989	0	1.9899	1.9899	0
0.1	0.5	2	0.8242	0.8183	0.7210	1.9472	1.9475	0.0154
0.1	0.9	2	4.3187	4.2640	1.2828	1.9021	1.9019	0.0105
0.5	0.1	5	0.0525	0.0526	0.1901	4.9474	4.9474	0
0.5	0.5	5	0.3330	0.3337	0.2098	4.6696	4.6670	0.0557
0.5	0.9	5	0.8252	0.8333	0.9756	4.1960	4.1969	0.0214
0.99	0.1	0	0.1103	0.1109	0.5410	0	0	0
0.99	0.5	0	0.9782	0.9852	0.7105	0	0	0
0.99	0.9	0	8.0525	8.1834	1.5996	0	0	0

(*): Percent Differences is calculated by | ET_{CTO} from Simulation - ET_{CTO} from equation (13)| / ET_{CTO} from equation (13) x 100

(**): Percent Differences is calculated by | E[I] from Simulation – E[I] from equation (3)| / E[I] from equation (3) x 100

5.1.2 The Model for the MTO strategy

We now study the special case of the pure MTO strategy. Since the MTO strategy only involves Stage 2 of the CTO strategy, we derive our expressions for its time and cost by setting $r = 0$ in **(5.13)** and **(5.14)**. Doing so in **(5.13)** gives and expected waiting time of

$$ET_{MTO,k} = \frac{1}{\frac{1}{S} - D}. \qquad \textbf{(5.15)}$$

Setting $r = 0$, substituting F_{CTO} for F_{MTO}, and removing the cost associated with producing and storing the generic component in **(5.14)** gives a total cost of

$$TC_{MTO} = F_{MTO} + \sum_{k=1}^{N} \left(S\lambda_k v_k + \lambda_k b_k \right), \qquad \textbf{(5.16)}$$

where v_k and b_k are the unit WIP cost and the unit production cost for the k^{th} product in the MTO strategy respectively.

5.2 Analysis of Time and Cost

In this section, we compare the cost and time of the MTO and CTO strategies to determine the conditions under which each strategy is preferred. To make a meaningful comparison between the MTO and CTO strategies, we let both strategies have the same demand and process time distributions for all product types. We begin by comparing the expected waiting times.

> ***Lemma 5.1*: *Assuming the same demand and process time distributions for any product,* $ET_{MTO,k} \geq ET_{CTO,k}$.**

All proofs are given in the Appendix.

Lemma 5.1 shows that the CTO strategy has a shorter customer waiting time than that of the MTO strategy. While this result may seem straight forward, **(5.13)** shows that the waiting time of the CTO strategy is dependent on the choice of the base-stock level, z. One might conjecture that a poorly chosen base-stock level may result in longer waits for the CTO strategy versus the MTO strategy because of excessive backorders for the generic component. Nevertheless, ***Lemma 5.1*** proves the result for any choice of the base-stock level.

Next, we compare the costs between the CTO and MTO strategies. To do so, we first make the following four assumptions:

1. $F_{CTO} \geq F_{MTO}$. To set up the CTO strategy, a firm often needs to make additional investments such as designing the generic component and purchasing additional equipment (such as the paint mixers installed in the hardware stores).
2. The production cost is proportional to the production time and

$$c_g = rb_{\max}, \qquad \textbf{(5.17)}$$

where $b_{\max} = \max(b_k, \forall k)$, b_k is the unit production cost for a type k product in the MTO strategy, and c_g is the unit production cost of the generic component. This assumption implies that the r portion of the total product production time is spent at the generic component stage (Stage 1) and the generic component must be built to meet the highest standard in the product family.

From (5.**17**),

$$c_g \geq rb_k, \forall k. \qquad \textbf{(5.18)}$$

In addition, since *(1-r)* portion of the production time is spent at Stage 2,

$$c_k = (1-r)b_k. \qquad \textbf{(5.19)}$$

combining **(5.18)** and **(5.19)** gives

$$c_g + c_k \geq b_k, \qquad \textbf{(5.20)}$$

ensuring that the total unit production cost of the CTO strategy is no less than that of the MTO strategy. We will revisit this assumption in ***Lemma 5.3***.

3. The demand and the process time distributions for both strategies are the same.
4. The WIP has the average added value. That is, for the MTO strategy the unit WIP holding cost is $v_k = \dfrac{b_k i}{2}$, where i is the firm's interest rate. Substituting the holding cost into **(5.16)** gives

$$TC_{MTO} = F_{MTO} + \sum_{k=1}^{N}\left(S\lambda_k \frac{b_k i}{2} + \lambda_k b_k \right). \qquad \textbf{(5.21)}$$

For the CTO strategy, the unit WIP holding cost at Stage 1 is $w_g = \dfrac{c_g i}{2}$ and the unit WIP holding cost at Stage 2 is $w_k = \dfrac{2c_g + c_k}{2} i$. Substituting these costs into **(5.14)** gives

$$TC_{CTO} = F_{CTO} + h_g E\left[I\right] + \sum_{k=1}^{N} \lambda_k \left(rS \frac{c_g i}{2} + c_g + (1-r)S \frac{2c_g + c_k}{2} i + c_k \right). \qquad \textbf{(5.22)}$$

Based on the four assumptions stated above, we now offer the following lemma to compare the cost between the CTO and MTO strategies.

Lemma 5.2: Given assumptions 1 through 4, $TC_{CTO} \geq TC_{MTO}$.

Lemmas 5.1 and ***5.2*** show that when both strategies have the same demand and process time distributions and the conditions of assumptions 1-4 are met, the expected customer waiting time of the CTO strategy is always no worse than that of the MTO strategy while the total cost of the MTO strategy is always no worse than that of the CTO strategy. These results help clarify the choice for a company implementing a delayed differentiation strategy. For companies focusing on customer waiting times, CTO is a better strategy while companies focusing on cost savings may prefer a pure MTO strategy. Unlike in the make-to-stock environment where reducing inventory holding cost is the main benefit of implementing delayed differentiation, reductions in the expected customer waiting times is the main benefit in the make-to-order environment.

We now take a critical look at the assumptions behind ***Lemma 5.2*** to see if there are any conditions where a CTO strategy may dominate a pure MTO strategy in both time and cost. In assumption 2, we state that $c_g = rb_{\max}$ which implies that the unit cost of the generic component is proportional to the cost of the most expensive product offered. This assumption is reasonable if the generic component has to meet the quality standard of the most expensive product and quality is an increasing function of the production cost. However, replacing the individual parts with a single generic component could lead to economies of scale in the production of the generic component. The example can be found in the automobile industry. Despite the extra production processes needed to make the common platform that fits various models, automobile manufacturers such as Toyota, Honda, Nissan, GM, and Ford are able to enjoy the economies of scale by adopting the common platform (Robinet, 2001). In ***Lemma 5.3***, we give conditions on the production cost of the generic component for which $TC_{MTO} \geq TC_{CTO}$. Under these conditions, the CTO strategy dominates the MTO strategy in both performance measurements.

Lemma 5.3: *Given assumptions 1, 3, 4, and* $c_k = (1-r)b_k$, $TC_{MTO} \geq TC_{CTO}$ *if*

$$c_g \leq \frac{1}{D}\frac{1}{(1+Si-0.5rSi)}\left[(F_{MTO}-F_{CTO})-h_g E[I]+\sum_{k=1}^{N}\left\{\lambda_k b_k\left(0.5Si\left(1-(1-r)^2\right)+r\right)\right\}\right]$$

Numerical Example*:*

We now give a numerical example to demonstrate the result stated in ***Lemma 5.3***. First, we assume there are 3 distinct products in the product family being produced. Let the demand rates for each product be $\lambda_1 = 2000$, $\lambda_2 = 3000$, $\lambda_3 = 4000$ and the unit production cost be $b_1 = 1$, $b_2 = 2$, $b_3 = 4$ respectively. Also, choose $z = 2$, $r = 0.7$, $S = 1/1000$, $i = 0.1$, $F_{MTO} = 1000$, and $F_{CTO} = 1500$. Under the assumptions of ***Lemma 5.2***, $c_g = rb_{\max} = 2.1$. From **(5.21)** and **(5.22)**, we get $TC_{MTO} = 21000$ and $TC_{CTO} = 26400$. Thus, as shown in ***Lemma 5.2***, $TC_{CTO} > TC_{MTO}$. Now assume that economies of scale exist for the production of the generic component. If c_g can be reduced by 28.57% to less than 1.78 (calculated by the condition given by ***Lemma 5.3***), TC_{MTO} will be greater than TC_{CTO} and the MTO strategy will be dominated by the CTO strategy in both time and cost.

Several scenarios make the condition of the ***Lemma 5.3*** easier to meet and hence the CTO strategy more beneficial. We describe two of them below.

1. If the difference of the fixed cost between the CTO and the MTO strategy is small, the condition of ***Lemma 5.3*** is easier to meet. This finding can be easily verified by observing ***Lemma 5.3***. This result is consistent with Lee and Tang (1997).
2. Smaller base-stock levels (z) make the condition of ***Lemma 5.3*** easier to meet. The right hand side of the condition is decreasing in $E[I]$. From **(5.3)**,

   $E[I] = z - \frac{\rho_g}{1-\rho_g}(1-\rho_g{}^z) = z - \rho_g\left(1+\rho_g+\rho_g{}^2+\cdots+\rho_g{}^{z-1}\right)$. It is easy to verify

   that $E[I]$ is an increasing function in z. Hence, smaller z will result in smaller $E[I]$, and the condition of the ***Lemma 5.3*** will be easier to meet.

5.3 Analysis of Degree of Delayed Differentiation (*r*)

We now study how the time and cost of the CTO strategy behave with respect to the degree of delayed differentiation (r). We begin with the expected waiting time.

Lemma 5.4: *$ET_{CTO,k}$ is strictly convex in r.*

Since $ET_{CTO,k}$ is strictly convex in r, the optimal r w.r.t. the expected waiting time can be obtained by solving its first order conditions. Let r^t represent this global minimizer. Although there is no simple closed form solution for the first order condition, r^t can be obtained by using a golden search type of algorithm on **(5.13)**.

To further study the behavior of r^t, we numerically solve **(5.13)** using a golden search method for 10 different utilization rates (SD) and 5 different base-stock-levels (z). We normalize the process times (S) to 1 and vary the demand rates (D) from 0.1 to 0.99 so that the utilization rate ranges from 0.1 to 0.99. At each utilization rate, r^t is calculated for 5 different base-stock levels, ranging from 1 to 5. The results of this experiment are shown in **Figure 20**.

Two observations can be drawn from **Figure 20**. First, r^t is close to 1 when the utilization rate is very low, and is close to 0.5 when the utilization is very high. This implies that the benefit of spreading the workload between the two stages is higher when utilizations are high. Second, r^t increases as the base-stock level increases. Equation **(5.1)** shows that a higher base-stock level reduces the waiting time of Stage 1. Hence, as the base-stock level increases, it is more desirable to move a higher portion of the total work to Stage 1 to reduce the expected waiting time.

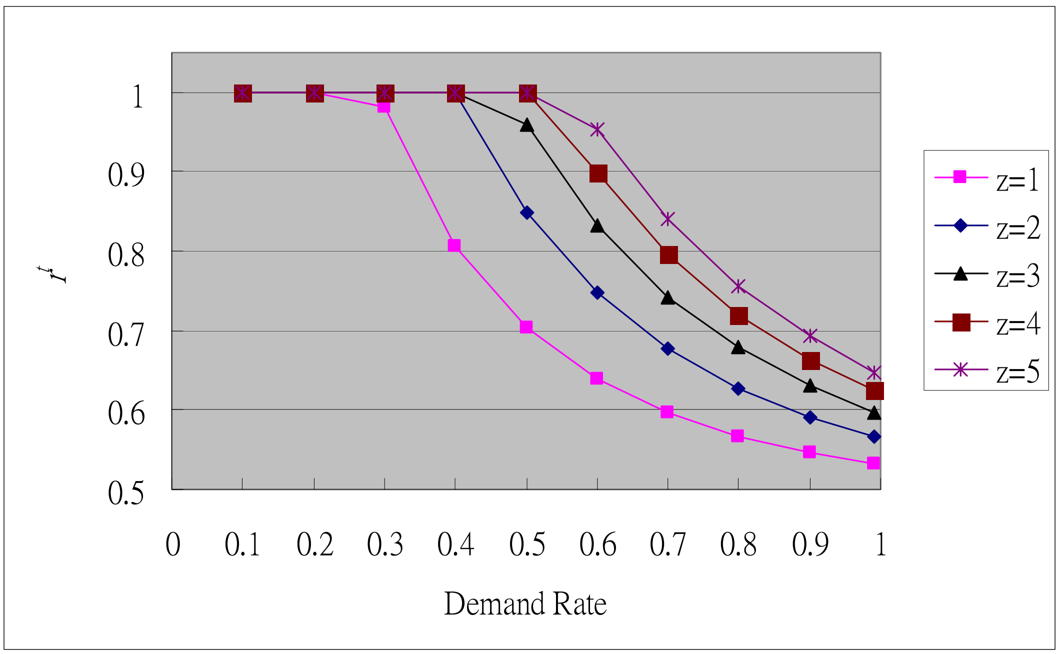

Figure 20: The r^t with Respect to Demand Rate and Base Stock Level.

Lemma 5.4 presents a surprising result since intuitively one would expect that the larger the portion of the product (r) is made-to-stock, the shorter the waiting time is. However, our results shows that r^t can be between 0 and 1. The convexity of the expected waiting time is due to the queuing effect. If r is very close to 1 or very close to 0, the production workload is concentrated on just Stage 1 or Stage 2, and the utilization rate of that stage increases. Since the expected waiting time increases exponentially with the utilization rate, it is better to spread the workload between two stages rather than to concentrate most of it in just one of the stages.

To study how the cost of the CTO strategy behaves with respect to r, we assume the fixed cost (F_{CTO}) is a linear increasing function of r, i.e.

$$F_{CTO} = rF \tag{5.23}$$

where F is a constant. In other words, we assume that it becomes more expensive to delay the product differentiation further down the production process. One of the reasons behind this assumption is that products are often over designed so that some functionality may be turned off if a customer is not willing to pay for it. Building the product with the minimum amount of functionality needed for each product should require a smaller fixed cost. Another reason is that supply chains tend to be tree like in strategy. Thus, the further down the production process the differentiation takes place, the more the number of individual sites that will require equipment to perform the differentiation process.

We also assume the holding cost for the generic component (h_g) is equal to the production cost (c_g) times the capital cost (i), i.e.

$$h_g = c_g i \, . \tag{5.24}$$

Lemma 5.5: *If $F_{CTO} = rF$, $h_g = c_g i$, and given assumptions 1 to 4, TC_{CTO} is concave in r.*

For firms seeking to minimize cost, *Lemma 5* shows that the best point to implement delayed differentiation is either $r = 0$, or $r = 1$. Combining this result with the result of ***Lemma 5.2***, if minimizing cost is the main objective, a firm should always choose the MTO strategy ($r = 0$). On the other hand, ***Lemma 5.3*** suggests that the cost of the CTO strategy may be less costly than that of the MTO strategy when the production cost of the generic component can be reduced below a certain level due to economies of scale.

CHAPTER 6

CONCLUSION AND FUTURE RESEARCH

Component commonality is a good option for firms facing the increasing challenges due to product proliferation. In Section 3, we evaluate three different component commonality strategies, i.e., DP, PCC, and MCC, and discuss the conditions under which one is preferred. We develop two models to analyze both the constant and the stochastic demand scenarios. The solution to minimizing the total inventory cost is presented and the managerial insights are derived from our analysis. We find that when the demand is constant, the MCC strategy is never beneficial. In addition, the PCC strategy is preferred when the price of the common component is low, the ordering cost is high, or the interest rate is high. On the other hand, when the demand is stochastic, we find that the MCC strategy can be used to reduce inventory cost if the demand variation is high. The PCC strategy is preferred when the price of the common component is low, the demand variation is high, the ordering cost is high, the lead time is long, or the interest rate is high. Furthermore, we conclude that when demand variation is moderate, unit shortage cost is not a significant factor in the choice of component commonality strategies. In the case of high demand variation, the PCC strategy is preferred when shortage cost is high, and the MCC strategy can be adopted for a range of moderate shortage cost.

There are several opportunities to enrich our study in the future. First, we assume all parts and modules are outsourced and all product design and development costs are excluded. In future research, we will include these costs in our models and study the issues regarding the coordination between product development and component commonality strategy. Second, we assume that all components are outsourced from the same or similar suppliers and that the ordering costs are very similar. Future research may extend our models to the multiple suppliers' problem, where each supplier has different capacity; ordering cost; and quantity discount scheme, to examine the impact of vendor selection.

In Section 4, we develop models representing two possible mass customization postponement strategies, TP and FP, and study their performance in terms of total supply chain cost and the expected customer waiting times. We find that once the number of products increases above some threshold level, the TP strategy is preferred under both performance metrics. We prove this analytically for the case of exponential arrival and process times and show it numerically for the general distribution case.

For general arrival and process time situations, we use G/G/1 approximations and design a numerical experiment to investigate how different factors affect the performance and attractiveness of the TP and FP strategies. Our experiment shows that higher arrival time and process time variation makes the FP strategy more favorable while increases in the number of products and higher interest rates make the TP strategy more favorable.

For managers needing guidance on the allocation of resources for process improvement, we find that increasing the coverage of the generic component and reducing the number of products provide a larger impact on improving the customer waiting times of the FP strategy than do reductions in the variability of the arrival and process times.

There are several opportunities to extend our study. First, we use expected cost and expected waiting time as performance measures. However, the variance of the cost and waiting time could also be important in some business applications. Our models can be extended to study these new measures. Second, the optimal coverage of the generic component and the optimal number of products may be interesting to some managers. This will involve more detail regarding the relationship between the fixed cost and the amount of redesign effort required. Third, we assume there is only one product family. Including multiple product families could lead to some interesting extensions including partial demand substitution and savings from product platforming. Fourth, we assume a constant unit production cost. In practice, high arrival and process time variation could increase production cost by increasing the possibility of rush orders and over time requirements to meet demand. This could be an interesting extension to be addressed.

In Section 5, we study the impact of delayed differentiation in make-to-order environments. Previous research shows that in a make-to-stock environment, the main benefit of adopting delayed differentiation comes from savings in inventory holding cost due to risk pooling. In contrast, we find that in a make-to-order environment, shorter expected customer waiting times provide the main benefit for implementing delayed differentiation. We show that under common assumptions, the introduction of delayed differentiation results in shorter customer waiting times and higher cost over a pure make-to-order strategy. However, we give reasonable conditions where the introduction of delayed differentiation results in shorter customer waiting times *and* lower cost, thus dominating a pure make-to-order strategy on both dimensions.

We also address the choice of where in the production process a company should differentiate its product, i.e. in the beginning, middle or end of the process. For companies seeking to minimize customer waits, we find that the expected waiting time is a convex function of the point of the differentiation. Surprisingly, the optimal (in terms of minimizing customer waiting times) point for differentiation in the process does not always occur at the end of process. Through a numerical study, we show that the optimal point occurs earlier in the process with increases in the utilization rate and decreases in the stocking level of the generic component.

There are several opportunities to extend our study. First, we assume *our* cost parameters are linear increasing in the percentage of the process that differentiation occurs. In some cases, however, these costs may increase nonlinearly. For example, there may be diminishing returns on the investment required for designing the generic component such that differentiation may occur further down the production process. Second, we separate time and cost as two performance measurements. A firm looking to globally optimize the point of differentiation in their product may do so by assigning a cost to the customer waiting time, thus consolidating the two performance measurements.

APPENDIX

Lemma 5.1*: *Assuming the same demand and process time distributions for any product, $ET_{MTO,k} \geq ET_{CTO,k}$.

Proof:

From Equations **(5.13)** and **(5.15)**, let

$$\Delta T = ET_{MTO,k} - ET_{CTO,k} = \frac{1}{\frac{1}{S} - D} - \frac{(rDS)^z}{\frac{1}{rS} - D} - \frac{1}{\frac{1}{qS} - D}$$

where, $q = (1\text{-}r)$. This can be expanded to

$$\Delta T = \frac{S(1-rSD)(1-qSD) - rS(rDS)^z(1-SD)(1-qSD) - qS(1-SD)(1-rSD)}{(1-SD)(1-rSD)(1-qSD)}$$

let $\Delta T = \dfrac{f}{g}$ **(5.25)**

and let us focus on the denominator, g first. Since $SD < 1$ (total demand rate is less than the total process rate), $(1\text{-}SD) > 0$. Furthermore, since both r and q are less than 1, both $(1\text{-}rSD)$ and $(1\text{-}qSD)$ are greater than 0. Hence;

$g > 0$ **(5.26)**

For the numerator f,

$$f = S(1-rSD)(1-qSD) - rS(rDS)^z(1-SD)(1-qSD) - qS(1-SD)(1-rSD)$$

Because $(rDS)^z \leq 1$,

$$f \geq S(1-rSD)(1-qSD) - rS(1-SD)(1-qSD) - qS(1-SD)(1-rSD)$$

if $r \geq q$

$$f \geq S(1-rSD)(1-qSD) - rS(1-SD)(1-qSD) - qS(1-SD)(1-rSD)$$

$$\geq S(1-rSD)(1-qSD) - rS(1-SD)(1-qSD) - qS(1-SD)(1-qSD)$$

$$= S(1-rSD)(1-qSD)-(1-SD)(1-qSD)(rS+qS)$$

$$= S(1-rSD)(1-qSD)-(1-SD)(1-qSD)S$$

$$= S(1-qSD)(1-rSD-1+SD)$$

$$= S(1-qSD)SD(1-r) \geq 0 \qquad \textbf{(5.27)}$$

Therefore, $f \geq 0$ when $r \geq q$

If $q \geq r$,

$$f \geq S(1-rSD)(1-qSD)-rS(1-SD)(1-qSD)-qS(1-SD)(1-rSD)$$

$$\geq S(1-rSD)(1-qSD)-rS(1-SD)(1-rSD)-qS(1-SD)(1-rSD)$$

$$= S(1-rSD)(1-qSD)-(1-SD)(1-rSD)(rS+qS)$$

$$= S(1-rSD)(1-qSD)-(1-SD)(1-rSD)S$$

$$= S(1-rSD)(1-qSD-1+SD)$$

$$= S(1-rSD)SD(1-q) \geq 0 \qquad \textbf{(5.28)}$$

Therefore, $f \geq 0$ when $q > r$

From Equations **(5.25)** to **(5.28)**, we get $\Delta T \geq 0$; hence $ET_{MTO,k} \geq ET_{CTO,k}$. ■

Lemma 5.2: *Given assumptions 1through 4,* $TC_{CTO} \geq TC_{MTO}$.

<u>Proof:</u>

Let $\Delta C = TC_{CTO} - TC_{MTO}$. From Equations **(5.21)** and **(5.22)**,

$$\Delta C = \left(F_{CTO} - F_{MTO}\right) + h_g E\left[I\right] + \sum_{k=1}^{N} \lambda_k \mathrm{H} \qquad \textbf{(5.29)}$$

where

$$\mathrm{H} = rS\frac{c_g i}{2} + c_g + (1-r)S\frac{2c_g + c_k}{2}i + c_k - \frac{Sb_k i}{2} - b_k$$

$$= \left(c_g + c_k - b_k\right) + \frac{iS}{2}\left(rc_g + (1-r)\left(2c_g + c_k\right) - b_k\right)$$

From Equation **(5.20)**, we get $\left(c_g + c_k - b_k\right) \geq 0$, and

$$\mathrm{H} \geq \frac{iS}{2}\left(rc_g + (1-r)\left(2c_g + c_k\right) - b_k\right)$$

$$= \frac{iS}{2}\left(rc_g + 2c_g + c_k - 2rc_g - rc_k - b_k\right)$$

$$= \frac{iS}{2}\left(2c_g + c_k - rc_g - rc_k - b_k\right)$$

$$= \frac{iS}{2}\left((1-r)\left(c_g + c_k\right) + c_g - b_k\right)$$

From Equation **(5.18)**, $c_g \geq rb_k, \forall k$, we get

$$\mathrm{H} \geq \frac{iS}{2}\left((1-r)\left(c_g + c_k\right) + rb_k - b_k\right)$$

$$= \frac{iS}{2}\left((1-r)\left(c_g + c_k\right) + (r-1)b_k\right)$$

$$= \frac{iS}{2}\left((1-r)\left(c_g + c_k - b_k\right)\right) \geq 0$$

since both $(1-r)$ and $\left(c_g + c_k - b_k\right)$ are greater than or equal to zero.

Since $\mathrm{H} \geq 0$, and $F_{CTO} \geq F_{MTO}$, Equation **(5.29)** shows that $\Delta C \geq 0$. Therefore $TC_{CTO} \geq TC_{MTO}$ and the cost of the CTO strategy is no less than that of the MTO strategy.■

Lemma 5.3: *Given assumptions 1,3, 4, and $c_k = (1-r)b_k$, $TC_{MTO} \geq TC_{CTO}$ if*

$$c_g \leq \frac{1}{D}\frac{1}{(1+Si-0.5rSi)}\left[(F_{MTO}-F_{CTO})-h_g E[I]+\sum_{k=1}^{N}\left\{\lambda_k b_k\left(0.5Si\left(1-(1-r)^2\right)+r\right)\right\}\right]$$

Proof:

From Equation **(5.21)**

$$TC_{MTO} = F_{MTO} + \sum_{k=1}^{N}\left(S\lambda_k \frac{b_k i}{2} + \lambda_k b_k\right)$$

From Equation **(5.22)**

$$\begin{aligned}
TC_{CTO} &= F_{CTO} + h_g E[I] \\
&+\sum_{k=1}^{N}\lambda_k\left(rS\frac{c_g i}{2} + c_g + (1-r)S\frac{2c_g + c_k}{2}i + c_k\right). \\
&= F_{CTO} + h_g E[I] + \sum_{k=1}^{N}\left\{\lambda_k c_g\left(\frac{rSi}{2} + 1 + (1-r)Si\right)\right\} + \sum_{k=1}^{N}\left\{\lambda_k c_k\left(\frac{(1-r)Si}{2} + 1\right)\right\} \\
&= F_{CTO} + h_g E[I] + Dc_g(0.5rSi + 1 + Si - rSi) + \sum_{k=1}^{N}\{\lambda_k(1-r)b_k(0.5(1-r)Si + 1)\} \\
&= F_{CTO} + h_g E[I] + Dc_g(1 + Si - 0.5rSi) + \sum_{k=1}^{N}\left\{\lambda_k b_k\left(0.5(1-r)^2 Si + (1-r)\right)\right\}
\end{aligned}$$

Let

$$\begin{aligned}
\Delta TC &= TC_{MTO} - TC_{CTO} \\
&= (F_{MTO} - F_{CTO}) - h_g E[I] - Dc_g(1 + Si - 0.5rSi) + \sum_{k=1}^{N}\left\{\lambda_k b_k\left(0.5Si + 1 - 0.5(1-r)^2 Si - (1-r)\right)\right\} \\
&= (F_{MTO} - F_{CTO}) - h_g E[I] - Dc_g(1 + Si - 0.5rSi) + \sum_{k=1}^{N}\left\{\lambda_k b_k\left(0.5Si\left(1-(1-r)^2\right) + r\right)\right\}
\end{aligned}$$

We get

$$TC_{MTO} \geq TC_{CTO}$$

$$c_g \leq \frac{1}{D}\frac{1}{(1+Si-0.5rSi)}\left[(F_{MTO}-F_{CTO})-h_g E[I]+\sum_{k=1}^{N}\left\{\lambda_k b_k\left(0.5Si\left(1-(1-r)^2\right)+r\right)\right\}\right]$$

Lemma 5.3 is proved.■

Lemma 5.4: *$ET_{CTO,k}$ is strictly convex in r.*

Proof:

From Equation **(5.13)**,

$$ET_{CTO,k}(r) = \frac{(rDS)^z}{\frac{1}{rS} - D} + \frac{1}{\frac{1}{(1-r)S} - D}$$

$$= \left(r^{z+1} S^{z+1} D^z\right)(1 - rSD)^{-1} + (1-r)S(1-(1-r)SD)^{-1}$$

we get

$$\begin{aligned} \frac{\partial ET_{CTO,k}(r)}{\partial r} &= (z+1)r^z S^{z+1} D^z (1-rSD)^{-1} + r^{z+1} S^{z+2} D^{z+1} (1-rSD)^{-2} \\ &- S(1-(1-r)SD)^{-1} - (1-r)S^2 D(1-(1-r)SD)^{-2} \end{aligned} \quad \textbf{(5.30)}$$

and

$$\begin{aligned} \frac{\partial ET_{CTO,k}(r)^2}{\partial^2 r} &= z(z+1)r^{z-1} S^{z+1} D^z (1-rSD)^{-1} + (z+1)r^z S^{z+2} D^{z+1} (1-rSD)^{-2} \\ &+ (z+1)r^z S^{z+2} D^{z+1} (1-rSD)^{-2} + r^{z+1} S^{z+2} D^{z+1} (2SD)(1-rSD)^{-3} \\ &+ S(1-(1-r)SD)^{-2}(SD) + S^2 D(1-(1-r)SD)^{-2} \\ &+ 2(1-r)S^3 D^2 (1-(1-r)SD)^{-3} > 0 \end{aligned} \quad \textbf{(5.31)}$$

Hence, $ET_{CTO,k}$ is strictly convex in r. ■

Lemma 5.5: *If $F_{CTO} = rF$, $h_g = c_g i$, and given assumptions 1 to 4, TC_{CTO} is concave in r.*

Proof:

From Equation **(5.22)**

$$TC_{CTO}(r) = F_{CTO} + h_g E[I] + \sum_{k=1}^{N} \lambda_k \left(rS\frac{c_g i}{2} + c_g + (1-r)S\frac{2c_g + c_k}{2} i + c_k \right)$$

Let $f_1 = F_{CTO}$, $f_2 = h_g E[I]$, and

$$f_3 = \sum_{k=1}^{N} \lambda_k \left(rS\frac{c_g i}{2} + c_g + (1-r)S\frac{2c_g + c_k}{2} i + c_k \right). \qquad \textbf{(5.32)}$$

From Equation **(5.23)**, f_1 is a linear (and concave) function in r.

Substituting Equation **(5.24)** into Equation **(5.3)** gives

$$f_2 = h_g E[I] = c_g i \left(z - \frac{\rho_g}{1-\rho_g}(1-\rho_g{}^z) \right). \qquad \textbf{(5.33)}$$

Substituting Equation **(5.17)** into Equation **(5.33)** gives

$$f_2 = rb_{\max} i \left(z - \frac{\rho_g}{1-\rho_g}(1-\rho_g{}^z) \right)$$

$$= rb_{\max} i \left(z - \rho_g \left(1 + \rho_g + \rho_g{}^2 + \rho_g{}^3 + \cdots + \rho_g{}^{z-1} \right) \right).$$

From Equation **(5.12)**, $\rho_g = rDS$. Substituting it into the above equation gives

$$f_2 = rb_{\max} i \left(z - rDS\left(1 + rDS + (rDS)^2 + (rDS)^3 + \cdots + (rDS)^z \right) \right)$$

$$= b_{\max} i \left(rz - r^2 DS\left(1 + rDS + (rDS)^2 + (rDS)^3 + \cdots + (rDS)^z \right) \right).$$

$$\frac{\partial^2 f_2}{\partial r^2} = b_{\max} i \Gamma(r) \qquad \textbf{(5.34)}$$

where

$$\begin{aligned} \Gamma(r) = &-2rDS\left(1 + rDS + (rDS)^2 + (rDS)^3 + \cdots + (rDS)^z\right) \\ &-4rDS\left(DS + 2r(DS)^2 + 3r^2(DS)^3 + \cdots + zr^{z-1}(DS)^z\right) \\ &-r^2 DS\left(2(DS)^2 + 6r(DS)^3 + \cdots + z(z-1)r^{z-2}(DS)^z\right) \le 0. \end{aligned}$$

Hence $\frac{\partial^2 f_2}{\partial r^2} \le 0$, and f_2 is concave in r.

For f_3, substituting Equation **(5.17)** and **(5.19)** into Equation **(5.32)** gives

$$f_3 = \sum_{k=1}^{N} \lambda_k \left(rb_{\max} + (1-r)b_k + Si\left(\frac{r^2 b_{\max}}{2} + (1-r)\frac{2rb_{\max} + (1-r)b_k}{2} \right) \right).$$

$$\frac{\partial^2 f_3}{\partial r^2} = \sum_{k=1}^{N} \lambda_k Si\left(-b_{\max} + b_k\right) \leq 0 \qquad \textbf{(5.35)}$$

Therefore, f_3 is concave in r.

Because f_1, f_1, and f_3 are concave, $TC_{CTO} = f_1+f_2+f_3$ is concave in r. ■

REFERENCE

1. Bai, L., Liu, L., Serfozo, R., Shang, W., 2003. Inter-departure Times in Base-stock Inventory-queues, Working Paper, Hong Kong University of Science and Technology.

2. Baker, K., Magazine, M., Nuttle, H., 1986. The effect of commonality on safety stock in a simple inventory model, Management Science, vol. 32, pp. 982-988.

3. Bucklin, L. P., 1965. Postponement, speculation and the structure of distribution channels, Journal of Marketing Research, vol. 2, no. 1, February, pp. 26-31.

4. Buzacott, J., and Shanthikumar J. G., 1993. Stochastic models of manufacturing systems. Prentice Hall, Upper Saddle River, New Jersey pp. 103-106.

5. Christopher, M., 1992. Logistics and supply chain management. Pitman Publishing, London.

6. Collier D., 1981. The measurement and operating benefit of component part commonality, Decision Sciences, vol. 12, pp. 85-96.

7. Cooper, J.C., 1993. Logistics structures for global business, International Journal of Physical Distribution and Logistics Management, vol. 23, no. 4, pp. 12-23.

8. Eynan, A., Rosenblatt, M., 1996. Component commonality effects on inventory costs, IIE Transactions, vol. 28, pp. 93-104.

9. Feitzinger, E., and Lee H. L., 1997. Mass customization at Hewlett Packard: the power of postponement. Harvard Business Review, vol. 75, no. 1, January-February, pp. 116-120.

10. Fisher, M., Ittner, C., 1999. The impact of product variety on automobile assembly operations: Empirical evidence and simulation analysis, Management Science, vol 45, pp. 771-786.

11. Garg, A., Tang, C. S., 1997. On Postponement Strategies for Product Families with Multiple Points of Differentiation. IIE Transactions, vol. 29, no. 8, pp. 641-650.

12. Gerchank, Y., Magazine, M.J., and Gamble, A.B., 1988. Component commonality with service level requirements, Management Science, vol. 34, pp. 753-760.

13. Gilmore, H. and Pine, B. J., 1997. The four faces of mass customization, Harvard Business Review, vol. 75, no. 1, January-February, pp. 91-101.

14. Gross, D., and C. Harris, 1985. Fundamentals of Queueing Theory. John Wiley & Sons, New York.

15. Gupta, D., and Srinivasan, M. M., 1998. Note: how does product proliferation affect responsiveness?, Management Science, vol. 44, no. 7, July, pp. 1017-1020.

16. Hayes, R. H., Wheelwright, S. C., 1979. Link Manufacturing Process and Product Life Cycles, Harvard Business Review, Vol. 57, no. 1, January-Febuary pp. 133 - 140.

17. Hill, T., Chambers, S., 1991. Flexibility - A Manufacturing Conundrum. International Journal of Operations & Production Management, vol. 11 no. 2, pp. 5-13.

18. Hillier, M.S., 1999. Component commonality in a multiple-period inventory model with service level constraints, International Journal of Production Research, vol. 37, pp. 2665-2683.

19. Johnson, L.A., Montgomery, D.C., 1974. Operations Research in Production Planning, Scheduling, and Inventory Control, John Wiley & Sons, New York.

20. Lee, H. L. 1996. Effective inventory and service management through product and process redesign, Operations Research, vol. 44, no. 1, pp. 151-158.

21. Lee, H. L., Billington, C., and Carter, B., 1993. Hewlett Packard gains control of inventory and service through design for localization, Interfaces, vol. 23, no. 4, July-August, pp. 1-11.

22. Lee, H. L., Tang, C. S. 1997. Modeling the Cost and Benefits of Delayed Product Differentiation, Management Science, vol. 43, no. 1, pp. 40-53.

23. Marchal, W. C., 1976. An approximation formula for waiting time in single-server queues, AIIE Transition, vol. 8, pp. 473-474.

24. Matthias, H., Frits, K. P., 2001. Successful build-to-order strategies start with the customer. MIT Sloan Management Review, vol. 43 no. 1, pp. 74-83.

25. McCarthy, F. T., 2000. All Yours. The Economist, vol. 355, no. 8164, pp. 57-58.

26. McCutcheon, D., Raturi, A., Meredith, J., 1994. The Customization-Responsiveness Squeeze, Sloan Management Review, vol.35 no. 2, pp. 89-99.

27. Pagh, J. D., and Cooper, M. C., 1998. Supply chain postponement and speculation structures: how to choose the right structure?, Journal of Business Logistics, vol. 19, no. 2, pp 13-34.

28. Phadke, M. S., 1989. Quality engineering using robust design. Prentice Hall, Englewood Cliffs, New Jersey.

29. Pine B. J., Victor B., and Boynton A. C., 1993. Making mass customization work, Harvard Business Review, vol. 71, no. 5, September-October, pp.108-116.

30. Pine, J., 1993. Mass customizing products and services, Strategy & Leadership, vol. 21, pp. 6-14.

31. Robinet, M., 2001. The Perfect Platform. Automotive Industries, vol. 181, no. 11, pp. 40.

32. Sanderson, S., Uzumeri, M., 1995. Managing product families: The case of the Sony Walkman, Research Policy, vol. 24, pp. 761-782.

33. Shore, H., 1988. Simple Approximations for the GI/G/c Queue -- II: The Moments, the Inverse Distribution Function and the Loss Function of the Number in the System and of the Queue Delay, The Journal of the Operational Research Society, vol. 39, no. 4, April, pp. 381 -391.

34. Swaminathan, J., Tayur, S. 1998. Managing Broader Product Lines through Delayed Differentiation Using Vanilla Boxes, Management Science, vol. 44, no. 12, pp. 161-172.

35. van Hoek, R. I., 2001. The rediscovery of postponement: A literature review and directions for research, Journal of Operations Management, vol. 19, no. 2, February, pp. 161-184.

36. Wolff, R. W., 1989. Stochastic modeling and the theory of queues. Prentice Hall, Englewood Cliffs, New Jersey.

37. Zinn, W. and Bowersox, D. J., 1988. Planning physical distribution with the principle of postponement, Journal of Business Logistics, vol. 9, no. 2, pp. 117-136.

38. Zipkin, P., 2000. Foundations of Inventory Management. McGraw-Hill, New York.