Chapter 15
Re-examining Postponement Benefits: An Integrated Production-inventory and Marketing Perspective

Hartanto Wong[1] and Mohamed Naim[2]

Hartanto Wong
Innovative Manufacturing Research Centre
Cardiff University
Cardiff CF10 3 EU, UK
wongh@cardiff.ac.uk

Mohamed Naim
Innovative Manufacturing Research Centre
Cardiff University
Cardiff CF10 3 EU, UK
naimmm@cardiff.ac.uk

Abstract This chapter presents a new perspective to obtain a better understanding of postponement benefits. This new perspective tries to address the important alignment between the production-inventory and marketing functions, under which we are able to obtain a more complete view on how postponement may enhance firms' profitability. We developed stylised models to capture the interactions between several factors including inventory, lead time, price and product

---

[1] Hartanto Wong is an RCUK Academic Fellow at the Cardiff University's Innovative Manufacturing Research Centre (CUIMRC). Prior to joining Cardiff University in October 2005, he was a Postdoctoral Research Fellow at the Centre for Industrial Management, Katholieke Universiteit Leuven-Belgium, from where he also obtained his PhD and MS degrees. His research interests and experience combine elements of supply chain and logistics management and operational research (OR) methodology. He has published in several journals and since recently serves as guest co-editor for a special issue on stochastic models for supply chains in the *International Journal of Production Economics*.

[2] Mohamed Naim is a professor in logistics and operations management at Cardiff Business School. He is a chartered engineer and a member of the Institution of Engineering and Technology, the Chartered Institute of Logistics and Transport. He is a director of the Logistics Systems Dynamics Group and the EPSRC funded Cardiff University Innovative Manufacturing Research Centre. He has published over 80 journal papers and is a former editor-in-chief of the *International Journal of Logistics* and is an advisory committee member for the International Symposium on Logistics. In 1997 he was awarded the Institute of Electrical Engineers' Manufacturing Division Premium for a paper on supply chain modelling and simulation. In 2003 he was granted a Royal Academy of Engineering Global Research Award at Linköping Institute of Technology, Sweden.

variety. Through numerical examples we show how postponement facilitates the attainment of a higher profit as the result of improved capability in compromising product variety and delivery lead time, on top of cost savings associated with reduced inventories.

Abbreviations

DD	Delayed differentiation
HP	Hewlett-Packard
i.i.d.	Independent and identically distributed
MTO	Make-to-order
MTS	Make-to-stock
PC	Personal computer

15.1 Introduction

Postponement or delayed product differentiation is an important concept used to accommodate mass customization, particularly in dealing with uncertainty due to proliferation in product varieties and uncertain demands from customers. This is achieved by properly designing the product structure and the manufacturing and supply chain process so that one can delay or postpone the final customisation of the product as much as possible, pending more accurate product demand information. Postponement offers a compromised solution between the two extreme make-to-stock and make-to-order policies. The unfavourable consequence of a make-to-stock policy characterised by a high level of inventory or lost sales due to forecasting errors is minimised through the customisation of the intermediate goods based on observed demand. At the same time, the long lead time associated to a make-to-order policy is reduced by making intermediate goods to stock.

The concept has received considerable attention from researchers and practitioners in recent years and is perceived as one of the major supply chain management practices having discernible impact on competitive advantage and organisational performance (Li *et al.* 2006). One of the classic examples of successful postponement application is Hewlett-Packard's (HP) for their DeskJet printers (Lee *et al.* 1993, Feitzinger and Lee 1997). The company opted to customise the printers at its local distribution centres rather than at its factories. For example, instead of customising the DeskJet at its factory in Singapore before shipping them to Europe, HP has its European distribution centre near Stuttgart, Germany to perform this job. HP restructured its printer production process by manufacturing a generic Deskjet printer and later localising the generic product by plugging in the localised modules (power supplies, packaging and manuals) at its local distribution centres. This way, HP was able to maintain the same service levels with an 18% reduction in inventory, saving millions of dollars. Another celebrated example is Benetton who used postponement to cope with volatile fashion trends and long production lead-times (Sigronelli and Hesket 1984, Dapiran 1992). By using un-dyed yarn to knit about half of its clothing and delaying the dying process to a later stage, Benetton has a better idea of the popular colours for the season. Other examples include IBM (Swaminathan and Tayur 1998), Whirpool (Waller *et al.* 2000) and Xilinx (Brown *et al.* 2000).

Although there is a long list of publications showcasing postponement as an attractive means to accommodate mass customisation, our literature review suggests that most of the extant academic literature focuses on the evaluation of postponement in the context of production-inventory systems and is strongly based on a cost-minimisation strategy (*e.g.*, Lee and Tang 1997, Garg and Tang 1997, Aviv and Federgruen 2001a, b, Gupta and Benjaafar 2004). The most prevalent finding says that postponement is beneficial due to a significant reduction of the total inventory cost achieved by reducing demand forecast errors and delaying expensive operations, which enable companies to maintain the bulk of its inventories in the cheaper and/or pre-customised form (Lee *et al.* 1993, Swaminathan and Lee 2003). A very common setting used in that type of evaluation is that demand from customers is assumed to be constant, *i.e.*, the impact of factors such as price, lead time and product characteristic on customers' purchase decision is ignored.

With the recent emphasis on integrative, customer-focused decision making, it can be argued that there is a need to explore some cross-functional implications and coordination issues of any particular supply chain strategies. This argument is also in line with one of the authors' experience in studying the implications of various supply chain redesign strategies adopted by a major European manufacturer of personal computers (Berry and Naim 1996). The study, in particular, presents a series of ongoing supply chain redesign strategies in the company including just-in-time, interplant logistics and planning integration, vendor integration, and strategic positioning of decoupling point or postponement. The study reported significant operational improvements achieved from the implementation of such strategies as measured by significant reductions in the total inventory, lead time and order amplification or bullwhip effect. Despite all these benefits, it was recognised that the study was not capable of capturing the cross-functional implications by, for example, linking operational and marketing decisions.

In the context of mass customisation in particular, there are many aspects pertinent to production-inventory and marketing functions. This is especially true when we consider the fact that the customisation is not "free" and needs to be traded-off against lead time, cost and other factors (McCutcheon *et al.* 1994, Squire *et al.* 2006). Techniques such as postponement have proven to be supportive of firms moving from mass production to mass customisation when looked at from the production-inventory perspective, *i.e.*, it is able to minimise the total inventory cost associated with uncertainties due to proliferation in product varieties. However, one may question whether postponement is also beneficial when looked at from a marketing perspective. When firms move from mass production to mass customisation by employing postponement, one might question whether, for example, there would be an effect of longer lead times on customers' willingness to buy, which will consequently influence the total profitability. Similarly, firms' decisions on product price cannot be ignored. The product price must be optimised by taking into account product varieties and inventory cost reductions achieved through postponement balanced against customers' dissatisfaction due to longer lead times.

Therefore, we argue that the currently dominating analysis focusing solely on the production-inventory system is incomplete because such an analysis pays no attention to the presence of marketing factors such as the sensitivity of customers' purchase decisions to product varieties, prices and delivery lead times. Furthermore, that type of analysis is grounded to the traditionally narrow view that overlooks the importance of coordination between different functional areas. This chapter aims to re-examine the role of postponement based on an integrated approach that takes into account both the production-inventory as well as marketing factors.

To the best of our knowledge, this chapter represents the first formal approach to evaluating postponement by considering the marketing-manufacturing interface. Consequently, this new approach would require the use of a profit-maximisation strategy instead of a cost-minimisation strategy. One of the results in this study, in fact, shows that postponement benefits assessed by the cost minimisation strategy are not equivalent to those assessed by the profit maximisation strategy. Furthermore, this new approach is also capable of capturing the interaction between factors such as the lead time, product variety trade-off and its implication on the profitability. To some degree firms can use postponement to mitigate this trade-off, but they cannot eliminate it as Squire *et al.* (2006) show empirically.

We explicitly compare two manufacturing configurations. The first configuration represents a make-to-stock (MTS) system in which multiple product variants are processed through a single-stage production. The second configuration represents a system employing postponement and is modelled as a two-stage production inventory system. Stage 1 produces intermediate goods that are common for all finished products and stage 2 differentiates finished products. Intermediate goods are made-to-stock and then differentiated only after customer demand is achieved. Queuing models are used for the analysis of production and inventory systems. Our marketing model of product differentiation is based on the Hotteling's locational model of customer choice behaviour (Hotelling 1929), which is widely used in the economic and marketing literature (*e.g.*, Lancaster 1990, Syam and Kumar 2006). The model captures the situation in which demand is not only sensitive to price and product characteristic, but also to delivery lead time. An integrated production-inventory and marketing model is then formulated for each of the two configurations and serves as the basis for assessing postponement benefits.

Our main interest in this chapter is to extend the current understanding of postponement particularly in order to better explain how postponement may enhance the total profitability. A better explanation should be able to reveal how postponement facilitates the attainment of higher revenues as a result of improved capability in compromising product variety and delivery lead time, on top of cost savings associated to reduced inventories. Furthermore, examining how postponement benefits are different when evaluated by the cost-minimisation strategy in contrast to the profit-maximisation strategy is also of interest.

The rest of the chapter is organised as follows. In the next section we provide a survey of the relevant literature. Section 15.3 outlines the notation and modelling assumptions and presents all the models. In Section 15.4 we present the numerical experiments and analyse the results and in Section 15.5 we wrap up the chapter with a concluding discussion and some suggestions for future research.

15.2 Literature Background

In this section we discuss the relevant literature with emphasis on the two streams of research that we consider most relevant to this work. The first stream of research is on postponement and the second is deals with the alignment of marketing and production-inventory functions. We summarise our review of these two streams in the following sections.

15.2.1 Postponement to Accommodate Mass Customisation

There is a large body of literature on postponement. We refer the reader to van Hoek (2001), Swaminathan and Lee (2003), Yang *et al.* (2004) and Boone *et al.* (2007), who provide a comprehensive review of research on postponement. The concept of postponement was actually introduced in the literature by Alderson (1950) as a means of reducing marketing costs. He believed that risks related to marketing operations could be reduced by postponing changes in form and identity to the latest possible point in the marketing flow or postponing change in inventory location to the latest possible point in time. Over time, a number of authors have introduced different conceptual categorisations of postponement strategies extending the understanding of where and when postponement is appropriate. In the paper by Zinn and Bowersox (1988), five different types of postponement strategies are identified. Four different strategies of form postponement (labelling, packaging, assembly and manufacturing) which, when combined with time postponement, constitute the five postponement strategies. Bowersox and Closs (1996) made a clear differentiation between logistics postponement and form or manufacturing postponement. Logistics postponement can be seen as a combination of time and place postponement (where place postponement refers to the storage of goods at central locations in the channel until customer orders are received). Pagh and Cooper (1998) provided a classification of postponement applications in the mid- to down-stream stages of the supply chain. Their classification is in fact a reworked version of the classification suggested by Zinn and Bowersox (1988). They identified four generic strategies by combining manufacturing and logistics postponement and speculation. These include: the full speculation strategy, the logistics postponement strategy, the manufacturing postponement strategy and the full postponement strategy. Rabinovich and Evers (2003) studied the effects of time and form postponement on inventory performance. Supported by an empirical survey, their study shows that the joint implementation of time and form postponement is synergistic in nature, giving a positive impact as reflected in a lower proportion of speculative inventory.

Analytical models measuring the costs and benefits of postponement are presented by Feitzinger and Lee (1997), Lee and Tang (1997) and Garg and Tang (1997). They show that the key benefit of postponement is from reductions in safety stock levels due to risk-pooling while the cost of designing the generic component is the main drawback. Aviv and Federgruen (2001a, b) studied postponement by considering a two-stage system in which stage 1 produces undifferentiated items that are later differentiated in stage 2. They introduced the possibility of learning from past realisations of demand as an additional factor that contributes to the value of postponement. They derived the resulting savings in safety stock and show that learning increases the value of postponement. In all these models, the effect of queuing at the production facility is ignored so that lead times are exogenous to the model and assumed to be constant. One of the main limitations of these models is that the existence of the interaction between the production facility utilisation and processing time variability in affecting order delays has been ignored.

Models that endogenise lead times are presented by Gupta and Benjaafar (2004) and Su *et al.* (2005). As in this chapter, those models explicitly take into account the queuing effect as a result of considering a capacitated production facility. Gupta and Benjaafar (2004) considered the capacitated production system employing form postponement as a two-stage system where a common product platform is produced in an MTS fashion in the first stage, which is then differentiated into different products in the second stage in a make-to-order (MTO) fashion. Su *et al.* (2005) compared two specific configurations. In the first configuration products are produced after orders arrive (MTO mode). The second configuration represents the system employing form postponement. Different from Gupta and Benjaafar, Su *et al.* (2005) examined the system where the second stage produces differentiated products in an MTS fashion instead of an MTO fashion. More recently, Wong *et al.* (2009) examined different postponement configurations characterised based on the positioning of the differentiation point and the customer order decoupling point.

There is a wealth of literature analysing postponement, though, to the best of our knowledge, the approach considering an integrated production-inventory and marketing framework has never been developed.

15.2.2 Production-inventory and Marketing Coordination

There exist several papers that study the coordination of marketing and operations decisions similarly to us. De Groote (1994) analysed the joint problem of marketing/manufacturing coordination with the focus on the exploration of some cross-functional implications of the flexibility of manufacturing processes. He formulated two complementary problems: the marketing choice of the breadth of the product line and the manufacturing choice of the flexibility of the production process. One of the key results is that the decentralised solution of the two problems typically yields a suboptimal solution. Dobson and Yano (2002) examined a product line design problem in which a manufacturer faces demand that is influenced by both price and lead time. The firm must decide which products to offer, how to price them, whether each should be made to stock or made to order, and how often to produce them. The offered products are assumed to share a single manufacturing facility where setup times introduce diseconomies of scope and setup costs introduce economies of scale. They assume deterministic demand thatI linearly decreases in price and lead time. Different from De Groote (1994) and this work, the models presented by Dobson and Yano (2002) do not take into account how the number of product variants influences customers' purchase decision. In comparison to the above mentioned two papers, the focus of our analysis is different in that we are particularly interested in comparing different manufacturing configurations in the context of mass customisation.

Jiang *et al.* (2006) compared two configurations; namely mass customisation and mass production. In their model, the mass customistion system consists of two stages: the initial build-to-stock phase and the final customise-to-order phase. The mass production system has a single stage that builds products with predetermined specifications to stock. Our analysis is different from theirs mainly in the following respects. Firstly, their model ignores the effect of congestion at the production facility while we explicitly model the queuing effect as a result of considering a capacitated production facility. Secondly, they do not include delivery lead time as a factor that affects customers' purchase decision as we do. Finally, Alptekinoglu and Corbett (2007) analysed the trade-off between the increased ability to precisely meet customer preferences and the increased lead time from order placement to delivery associated with customised products. The analysis presented in this chapter is different as our main focus is to evaluate the relative merits of postponement compared to the make-to-stock policy while their interest is to determine an optimal product line design in which it is possible to have a combination of made-to-stock and made-to-order products.

15.3 The Models

15.3.1 Description of Manufacturing Configurations

We distinguish two manufacturing configurations. In the first configuration, postponement is not employed and a set of finished products is produced in an MTS mode. Items are produced ahead of demand and kept in stock, ready to be shipped upon receipt of orders. This configuration is modelled as a one-stage production-inventory system. The second configuration employing postponement is modelled as a two stage production-inventory system, where stage 1 produces a component that is common for all finished products and stage 2 differentiates finished products. This configuration maintains stocks of a generic component and differentiates the finished products only after demand is realised. Throughout this chapter we term these two configurations MTS and delayed differentiation (DD). Figure 15.1 illustrates the two configurations.

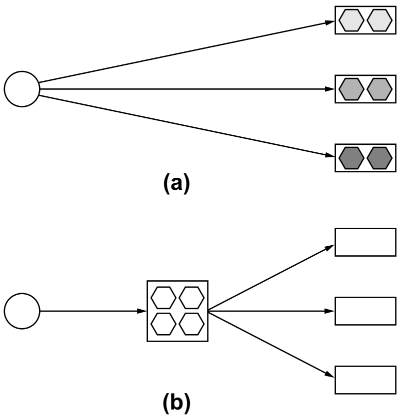

Figure 15.1 One-stage MTS (a) and two-stage DD configurations (b)

15.3.1.1 Configuration I: Make-to-stock System – No Postponement

Consider a manufacturer who offers N finished product variants, indexed by $i = 1$, 2, …, N. We denote d_0 as the aggregate demand rate, where $d_0 = \sum_{i=1}^{N} d_i$. We model this configuration as a single stage production-inventory system. End customer demand of product i arrives in single units according to a Poisson process with rate d_i. Note that, as explained in the previous section, the demand level is a function of price and lead time. We assume that the manufacturing processing times are independent and identically distributed (i.i.d.) random variables and exponentially distributed with a mean processing rate of m (the average manufacturing lead time is equal to $1/m$). These assumptions make the analysis tractable without a significant loss in accuracy, especially as our emphasis is in deriving qualitative patterns and managerial insights. For stability, we require that $d_0 / m = \rho < 1$. For the MTS configuration, demand is satisfied from stock unless the corresponding inventory is empty. All shortages are backlogged. We assume that a base-stock policy is used for the inventory control. Under this assumption, each demand triggers the immediate release of raw material, which is assumed to always be available (see Buzacott and Shathikumar (1993) for a formal definition of a base stock policy). Let S_i denote the base stock level for finished product i. Furthermore, changeover times between products are assumed to be negligible.

15.3.1.2 Configuration II: Mass Customisation with Delayed Differentiation

We model this configuration as a two-stage production-inventory system where stage 1 produces a component that is common for all finished products and stage 2 differentiates finished products. This configuration can be seen as a hybrid strategy in which a generic component is built-to-stock and then differentiated only after customer demand is realised. We assume that the processing times for stages 1 and 2 are i.i.d. random variables and exponentially distributed with mean processing rates of m_1 and m_2, respectively, and that $1/m_1 + 1/m_2 = 1/m$. We define $f (0 < f < 1)$ as the fraction of the mean total processing time consumed by the generic component. Thus, we may write $m_1 = m/f$ and $m_2 = m/(f-1)$. Small f values represent early postponement while large f values represent late postponement. Again, we require that $d_0/m_j = \rho_j < 1$ for $j = 1$ and 2, where ρ_j is the j-stage utilisation rate. The base stock level for generic components is denoted as S_0, while the stock level for finished products is zero for all $i = 1, 2, \ldots, N$. Here we also assume negligible changeover times.

Let c_{MTS} and c_{DD} denote the unit production cost for each of the manufacturing configurations, respectively. We assume that these two costs are identical, so that at the very least this gives us an upper bound on the benefits of postponement. Let h denote the holding cost *per* unit *per* unit time for all finished products, and h_0 denote the holding cost *per* unit *per* unit time for the generic component. For both configurations there is a product proliferation cost k, incurred every time the manufacturer offers a new product variant. This cost could include redesign, *tooling* and setup costs. The linearity assumption of product proliferation cost in the number of products is in line with common observations in the operations literature (Thonemann and Bradley 2002, Benjaafar *et al.* 2004). All the notations used throughout the paper are presented in Table 15.1.

Table 15.1 List of notations

Demand input parameters	
d_0	Total potential demand rate
π_i	First choice probability of product i
Θ	Customers' ideal taste
d_i	Demand rate for product i
R	Customer reservation price
$1-\beta$	Service level
Production input parameters	
m	The production rate for the MTS configuration
m_1	The production rate at stage 1 for the DD configuration
m	The production rate at stage 2 for the DD configuration
f	The fraction of the total lead time required to make the generic component

Table 15.1 Continued

Cost parameters	
h	Unit inventory holding cost for the finished product
h_0	Unit inventory holding cost for the generic component
c_{MTS}	Unit production cost for the MTS configuration
c_{DD}	Unit production cost for the DD configuration
k	Product proliferation cost
c_x	Linear transportation cost
c_t	Linear delay or waiting cost
Decision variables	
S_i	Base stock level for the finished products
S_0	Base stock level for the generic component
$\mathbf{S}$	Vector of base stock levels
N	Number of product lines
x_i	Product i 's characteristic
$\mathbf{x}$	Vector of product characteristics
p	Product price
T	Promised delivery lead time
Performance measures	
I_i	Expected on-hand inventory for the finished products
I_0	Expected on-hand inventory for the generic component
Z	Expected total profit

15.3.2 The Marketing Model

Our marketing model is based on a location model of customer choice behaviour, which is well known in the economics and marketing literature. It is along the lines of the spatial location model of Hotelling (1929) and its extensions (Lancaster 1990). We consider a monopolistic situation where the manufacturing firm serves a market with heterogeneous customers over a single time period. Customers' preferences are uniformly distributed over a closed interval of the product space [0, 1]. The product offerings are horizontally differentiated, each characterised by a single point in that interval quantified by a real number between 0 and 1. We are aware that mass customisation may also include a range of product offerings with vertical differentiation, in which case products offered are different with respect to their qualities. For MP3 players, the horizontal differentiation would be due to different colours or other "taste" attributes, while the vertical quality differentiation would be due to different memory size. However, in order to simplify the analysis we focus on the horizontal product differentiation and leave the inclusion of vertical differentiation as a future research opportunity.

Products are offered with price p, assumed to be identical for all products. The uniform pricing scheme is reasonable when the products are horizontally differentiated with qualities of products at the same level. Each customer buys one unit from the manufacturer and has her own ideal product represented by her location $\theta \in [0, 1]$. Our marketing model captures the situation in which demand is not only sensitive to price and product characteristics, but also to delivery lead time. We assume that the manufacturer commits to satisfying promised lead time t for all products and maintaining a service level of $1-\beta$ (*i.e.*, delivery occurs within t time units with $1-\beta$ probability). The disutility of customers incurred when buying their non-ideal product is represented by a linear transportation cost c_x *per* unit distance between their ideal product and the purchased product. Further, there is also a linear delay cost c_t *per* time unit of delivery or waiting time. Higher values of c_x and c_t mean customers are more sensitive to the deviation from their ideal products and the waiting time, respectively.

The utility of customer whose ideal taste is θ from buying product i with characteristic x_i, price p and delivery time guarantee t is given by

$$U(\theta, x_i, p, t) = r - p - c_t t - c_x \,|\, \theta - x_i \,|, \tag{15.1}$$

where r is a reservation price, defined as the maximum amount of money customers are willing to spend to buy the products. All customers are assumed to have a common reservation price. A customer buys the product that maximises her utility provided that it is non-negative, otherwise she does not make a purchase. Product i is said to be the first choice of a particular customer if it gives a non-negative utility and its utility is the maximum among all products offered by the manufacturer. Denoting π_i as the *first choice probability* of product i, the demand rate for product i can be defined as $d_i = \pi_i d_0$, $i = 1, 2, \ldots, N$. We assume that r is large enough so that the net utility is always greater than zero and so all customers will buy a product. Consistent with this, we also assume that complete market coverage is optimal. This assumption is common in the marketing and economics literature (Alptekinoglu and Corbett 2007).

To determine an optimal design of the product line, we use the well known optimality condition for Hotteling's location model, which ia also identified in de Groote (1994) and Gaur and Honhon (2006). That is, for a given N, the optimal product line has a simple structure: the market should be partitioned in segments of equal lengths, the characteristics of the products should correspond to the taste of the customers located in the middle of the segments and the manufacturer should set prices to make customers located at the extreme of the segments indifferent between buying and not buying.

Consider the example shown in Figure 15.2. In this particular example, the manufacturer offers one product ($N = 1$). The guaranteed delivery time t is assumed to be known (as the consequence of the inventory decision). Following the optimality condition stated above, the optimal product design is obtained by setting the product's characteristic at $x^* = 0.5$. As there is only one segment for $N = 1$, that characteristic corresponds to the taste of the customers located in the middle of the segment. Furthermore, Figure 15.2 also shows two disutility functions that correspond to two different prices. The price p_B leads to full market coverage, *i.e.*, $\pi_B = 1$ while a higher price, p_A, leads to a lower market coverage ($\pi_A < \pi_B$). In this case, p_B is the maximum price that gives full market coverage and makes the customers located at the extreme of the segment indifferent to buying or not buying, as indicated by the disutility value being equal to the reservation price. As we assume that full market coverage is optimal, given that *N* and *t* are fixed, it is straightforward to see that p_B is the optimal price. All prices less than p_B are suboptimal because they result in lower revenues.

Consider now the other example shown in Figure 15.3, in which the manufacturer offers two products ($N=2$). Given that there are two segments, the optimal design of the product line is obtained by setting the two products' characteristics at $x_1^* = 0.25$ and $x_2^* = 0.75$, and price at *p*. The two characteristics partition the market into two segments of equal lengths and correspond to the taste of the customers located in the middle of the two segments. Moving away from these characteristics, as illustrated by setting the second product characteristic at x_1 instead of x_1^*, will lead to a suboptimal situation as a result of lower total market coverage.

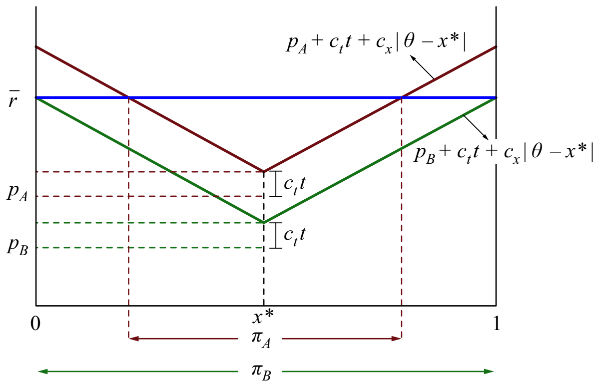

Figure 15.2 An example of the disutility function for one product

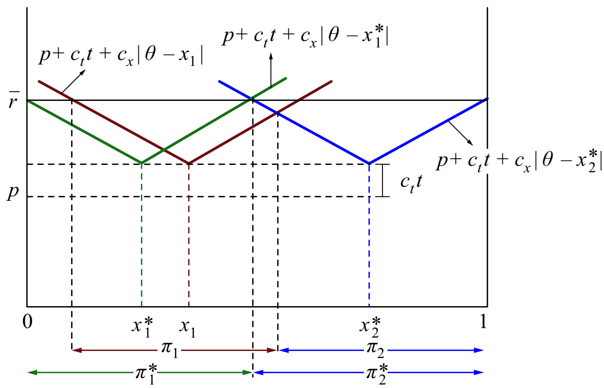

Figure 15.3 An example of the disutility function for two products

To put it more formally, given that N, t, c_t and c_x are fixed, we obtain full market coverage with the maximum revenue by setting:

$$x_1^* = \frac{2i-1}{2N}, \; i = 1, 2, ..., N \tag{15.2}$$

$$p^* = r - c_t t - \frac{c_x}{2N}. \tag{15.3}$$

From (15.2) and our assumption of the optimality of complete market coverage, it is easy to show that $\pi_i = 1/N$ and $d_i = d_0/N$ for all $i = 1, 2, \ldots, N$.

15.3.3 The Production-inventory Model

In this section we present the models used to evaluate the production-inventory systems for each of the two configurations.

15.3.3.1 The MTS Configuration

Following Buzacott and Shanthikumar (1993), for a given base stock level, the expected inventory for finished product i is given by

$$I_i(S_i) = S_i - \left(\frac{d_i}{m-d_0}\right)\left(1-\hat{\rho}_i^{S_i}\right), \tag{15.4}$$

where $\hat{\rho}_i = d_i/(m-d_{-i})$ and $d_{-i} = \sum_{j \neq i} d_j$. The probability that the order-fulfilment time will not exceed a quoted lead time t ($t \geq 0$) is given by

$$P_R\left[T_i(S_i) \leq t\right] = 1 - \hat{\rho}_i^{S_i} \times e^{-(m-d_0)t} \tag{15.5}$$

The manufacturer sets a service level $1-\beta$, where $0 < \beta < 1$, guaranteeing that the actual lead time will not exceed the promised lead time, *i.e.*, $P_R\left[T_i(S_i) \leq t\right] \geq 1-\beta$. It is very straightforward to find that for a given base stock level S_i the manufacturer will be reasonably interested in setting the promised lead time such that the service constraint is binding. We can state

$$t = \max\left(0, \frac{1}{m-d_0}\left(S^i \ln \hat{\rho}_i - \ln \beta\right)\right). \tag{15.6}$$

15.3.3.2 The DD Configuration

We use the evaluation models derived in Gupta and Benjaafar (2004). Suppose f, the proportion of the total lead time used to manufacture the generic component, is known. For a given base stock level of generic component, the expected inventory level is given by

$$I_0(S_0, f) = S_0 - \left(\frac{\rho_1 (1-\rho_1^{S_0})}{1-\rho_1} \right), \tag{15.7}$$

where $\rho_1 = fd_0 / m_1$. The probability that the order-fulfilment time exceeds a quoted lead time t ($t \geq 0$) is given by

$$P_r(T(S_0) \geq t) \approx \begin{cases} (1+\rho^{S_0}(1-\rho)\mu t)e^{-\mu(1-\rho)t} & \text{if } \rho_1 = \rho_2 = \rho_3 \\ e^{-\mu_2(1-\rho_2)t} + \left(\dfrac{(1-\rho_2)\rho_1^{S_0+1}}{\rho_2 - \rho_1} \right) & \\ e^{-\mu_2(1-\rho_2)t} - e^{-\mu_1(1-\rho_1)t} & \text{otherwise.} \end{cases} \tag{15.8}$$

15.3.4 *The Integrated Model*

In this section we present the integrated model for each of the configurations. First, a formal expression of the optimisation problem is introduced. After that we present the solution procedure for determining the optimal solution for each configuration.

15.3.4.1 The MTS Configuration

Define $\mathbf{x} = [x_1\ x_2\ ...\ x_N]$ as a vector of product characteristics and $\mathbf{S} = [S_1\ S_2\ ...\ S_N]$ as a vector of base stock levels. We formalise the manufacturer's optimisation problem as follows.
Problem $\mathrm{P}^{\mathrm{MTS}}$

$$\text{Max } Z(N, \mathbf{x}, \mathbf{S}, p) = \sum\nolimits_{i=1}^{N} (p - c_{\mathrm{MTS}}) \cdot d_i(x_i, S_i, p) - h \cdot I_i(S_i) - k \cdot N \tag{15.9}$$

Recall that the optimal product line has a structure in which the market should be partitioned in segments of equal lengths. This means that for a given N, $\pi_1 = \pi_2 = ... = \pi_N$ and so $d_1 = d_2 = ... = d_N$. Because of this symmetry it is reasonable to have identical optimal base stock levels for all products, *i.e.*, $S_1^* = S_2^* = ... = S_N^*$. As already discussed in Section 15.3.2, for a given N we are able to determine the optimal $\mathbf{x}$ using (15.2) and this optimal $\mathbf{x}$ will not be affected by decisions made for the base stock level $\mathbf{S}$ and price p. We also know that when the base stock level is given, we are able to calculate the promised lead time, which in turn allows us to determine the optimal price using (15.3). This leads us to develop a two-stage based solution procedure. In the first stage, we fix N and optimise $\mathbf{S}$, and in the second stage we optimise N.

From (15.6) it can be seen that the promised lead time t linearly decreases with S before reaching a zero level. It can also be proven that $I_i(S_i)$ increases and is a convex function of S_i. This means that the expected total profit Z is a concave function of S, which helps us to determine the optimal base stock level. We are now able to determine the optimal solution for a given N. The next step is to optimise N, which can be done by gradually increasing N starting from $N = 1$. For each value of N we optimise $\mathbf{x}$, p, and $\mathbf{S}$. The search can be terminated when the following condition is met $(r - c_{MTS})d_0 \le kN$. The left term in this condition is a constant and represents the maximum profit that can be gained by setting the price equal to the reservation price. The right term represents the proliferation cost,which linearly increases with N. So the condition ensures that no better improvement is possible by increasing N.

15.3.4.2 The DD Configuration

Problem P^{DD}

$$Z[f, N, \mathbf{x}, S_0, p] = \sum_{i=1}^{N} (p - c_{\mathrm{DD}}) \cdot d_i[x_i, S_0, p] - h_0(f) \cdot I_0[S_0, f] - k \cdot N \tag{15.10}$$

The optimal solution for the above defined problem can be obtained using a technique similar to that used in solving the problem for MTS. For a given N we need to optimise $\mathbf{S}$ and f, and then we need to optimise N. Different from the problem P^{MTS}, however, it is not easy to prove whether or not the profit function is concave in S_0. Given that f is fixed, we know that $I_0(S_0)$ is increasing in S_0 and that when S_0 is relatively large we will reach a situation where $t(S_0) \approx t(S_0 + 1)$. If we reach such a situation, no improvements can be made by increasing S_0 further. To optimise f, we use a simple search technique, which is also used in Wong *et al.* (2009). Then the next step is to optimise N, which can be done using the same technique as for solving the problem P^{MTS}.

15.4 Analyses

In this section we present the numerical analysis that focuses on two lines of enquiries as outlined in the Introduction. Firstly, we aim to get some insights from the comparison of the cost minimisation and profit maximisation strategies used for the evaluation of postponement benefits. Secondly, we are interested in assessing how postponement may actually enhance the manufacturer profitability taking into account marketing as well as production-inventory factors.

15.4.1 Cost Minimisation Versus Profit Maximisation

In this section we present numerical examples to demonstrate differences on postponement benefits when the profit maximisation strategy is used rather than the cost minimisation strategy. Consider the following system parameters:

1. aggregate demand rate $d_0 = 5$ / time unit;
2. reservation price $r = 500$;
3. production rate $m = 6$ / time unit;
4. product proliferation cost $k = 10$;
5. linear cost associated to waiting $c_t = 45$;
6. linear transportation cost $c_x = 120$;
6. unit production cost $c_{MTS} = c_{DD} = 100$;
8. unit holding cost $h_i = 20$ / time unit; and
9. service level = 98%.

Under these parameters and the integrative model, we obtain the optimal solutions for the two configurations summarised in Table 15.2.

Table 15.2 Comparison of optimal parameters (MTS *vs*. DD) – an example

Optimal parameter	MTS	DD
Expected profit	1453.50	1747.70
Price	473.16	466.89
Number of product variants	3	5
Base stock level	8	4
Promised lead time	.1520	.4691

Under the profit maximisation strategy, the benefit of postponement can be determined by calculating the relative difference of profits earned by the MTS and DD configurations (profit gain of DD over MTS). The measure we use is

$$\%PROFIT\ GAIN = \frac{Z_{DD}(f, N, x, S_0, p) - Z_{MTS}(N, x, S, p)}{Z_{MTS}(N, x, S, p)} \times 100\% \qquad (15.11)$$

For this particular example it can be shown that the profit generated by the DD configuration is 20.24% higher than the MTS configuration. It is shown in Table 15.2 that the optimal stock level for the DD configuration (four units for the generic component) is significantly less than that of the MTS configuration (eight units for each product). This inventory reduction obviously contributes to the total increased profitability. The result also shows the advantage of employing postponement in offering more product variety, thereby enhancing the customisation level. However, the example also shows the downside of postponement in terms of responsiveness. It is observed that under the same service level (98%), the promised lead time that can be offered by the DD configuration is longer than the MTS configuration. To compensate the negative effect of the longer lead time on customers' demand, the system responds by lowering the price of the products. This may sound counter-intuitive if one does not take into account the lead time *versus* product variety trade-off and ignore the fact that this trade-off has an effect on the pricing decision.

If the lead time is assumed to be the same or customers are not sensitive to delivery lead times, one would expect that customers can be charged a higher price for having more options. Likewise, while it is obvious that the greater product variety afforded by the DD configuration provides customers with some incentives to compensate the longer lead time, the optimal price of the customised products will be dependent on whether or not the incentives are sufficient. If not, then it is reasonable to set a lower price for the customised products than the standard products, as illustrated in this example. Nevertheless, it is worth noting that the result shown in this particular example should not be generalised too far by stating that, for example, the optimal price for the DD configuration (mass customisation) is always lower than the optimal price for the MTS configuration (mass production). The results will ultimately be dependent on the value of the parameters.

Now consider the evaluation of postponement based on the cost minimisation strategy. The optimisation problem based on the cost minimisation strategy can be formulated straightforwardly by removing all the marketing-related parameters from the set of decision variables. We refer the reader to Wong *et al.* (2009) for details of the model description. Under the cost minimisation strategy, the benefit of postponement is determined by calculating the relative difference in the total inventory costs

$$\%COST\ SAVING = \frac{COST_{MTS} - COST_{DD}}{COST_{MTS}} \times 100\% \tag{15.12}$$

Suppose the number of product variants for the two configurations is exogenously determined, $N=3$. For the MTS configuration the optimal stock level is $S_i = 8$ for $i = 1$, 2, and 3, resulting in promised lead time of $t = 0.152$ with a service level of 98% and the expected inventory cost 382.33. If we apply the same service level and promised lead time to the DD configuration, the following optimal solution is obtained: the stock level of generic component $S_0 = 13$ and the expected inventory cost is 201.42. The postponement benefit for this particular example is as high as 47.32%.

The above calculation example clearly indicates there could be significant differences of postponement benefits when evaluated under the two different strategies. Under the cost minimisation strategy, the value of postponement can be as high as 47.32%. However, one should recognise the fact that this saving is obtained without considering the effect of product line offerings on revenues as N is exogenously determined. The value of postponement calculated using the profit maximisation strategy, which is 20.24%, can be seen as a more reasonable assessment of the actual postponement benefits. Furthermore, the existing trade-off between product variety and lead time is also neglected. Under the integrated approach, the parameters c_x and c_t representing how sensitive customers are to the deviation between their ideal taste and the feature offered and to the delivery lead time would obviously determine the profitability (see Section 15.4.2 for more details). Under the cost minimisation strategy, however, the importance of these parameters is invisible.

In summary, through these numerical examples we show that the integrated model we develop allows us to better explain how postponement results in increased supply chain profitability by having giving us clearer visibility regarding the interaction among all the factors attributed to the production-inventory as well as the marketing functions.

15.4.2 The Impact of Postponement on Profitability

We now present numerical results in assessing how postponement enhances the manufacturer profitability. It is also our aim to examine how the profitability level is affected by different parameters. A numerical experiment was conducted to achieve this purpose and the list of parameter values used in the experiment is presented in Table 15.3.

Table 15.3 presents all the parameter values used in the experiment. The aggregate demand rate is fixed at $d_0 = 6$ in this experiment. The reservation price is also fixed at $r = 600$, which we consider large enough to ensure that the net utility is always greater than zero and so all customers will buy a product. We fix the unit production cost for MTS and DD at $c_{\mathrm{MTS}} = c_{\mathrm{DD}} = 100$. As stated earlier, setting the same unit production cost for the two configurations will constitute an upper bound of the value of postponement. Five different values of m are used for the production rate. To study the effect of the sensitivity of customers to the delivery lead time and the deviation from their ideal preference, four different values of c_t and c_x are tested in this experiment. Further, four values are also used as the product proliferation cost. The combination of all these parameter values makes in total 1280 problem instances. We summarise the main findings as follows.

Table 15.3 The parameter values used in the numerical experiment

Parameter	Unit	Number of values	Values
d_0	/ time unit	1	5
r	£	1	600
c_{MTS} , c_{DD}	£	1	200
m	/ time unit	5	6, 7, 8, 9, 10
h_i	£/unit/time unit	4	5, 10, 15, 20
c_t	£	4	20, 40, 60, 80
c_x	£	4	50, 100, 150, 200
k	£	4	5, 10, 15, 20

15.4.2.1 Aggregate Comparison

Table 15.4 summarises the overall average values for different measures that could be of interest when comparing the two configurations. The results show that postponement leads to increased profitability in general. The value of postponement, measured by *%PROFIT GAIN* has an average of 8.4% and can be as high as 37.8%. However, it is also notable to mention that the dominance of DD over MTS is not observed in the whole problem instances. There are six particular instances in which MTS brings greater profits than DD, which reflects the detrimental effect of postponement. Parameters of these instances are characterised by the smallest c_x (= 50), the highest c_t (= 80), the smallest hi (= 10) and the highest k (= 20). It is also shown that the average optimal number of products DD can offer is 6.44, while MTS can only offer 4.41 products. While the product proliferation cost is the same for the two configurations, increased flexibility offered by postponement allows the manufacturer to enhance the customisation level by offering more product variants. Postponement also leads to a higher average optimum price that can be charged to customers. Although the difference between the average prices for the two configurations appears to be insignificant, for some instances we may find that the difference is much larger. We shall discuss this later in more detail.

Table 15.4 Aggregate comparison between MTS and DD

Output measures	MTS	DD
Maximum profit	1851.05	1895.77
Minimum profit	1192.96	1555.01
Average profit	1634.54	1764.91
Average number of products	4.41	6.44
Average price	567.91	575.50

15.4.2.2 The Effect of Production Rate (*m*)

The effect of production rate on the benefit of postponement is illustrated in Figure 15.4. As the demand is held constant, this also represents the effect of the capacity utilisation level. It is shown that while the average profits for the two configurations increase in the production rate, the benefit of postponement appears to be diminishing. This finding is in line with what is reported in Wong *et al.* (2009). For a very congested system in which the production rate is low, the relatively high benefits of postponement come from significant differences in total stocks held by the MTS and DD configurations. For MTS, we observe that the reduction in total stock across all product variants caused by increased production rates has more profound effects in comparison to the reduction of the total stock of generic component in the DD configuration.

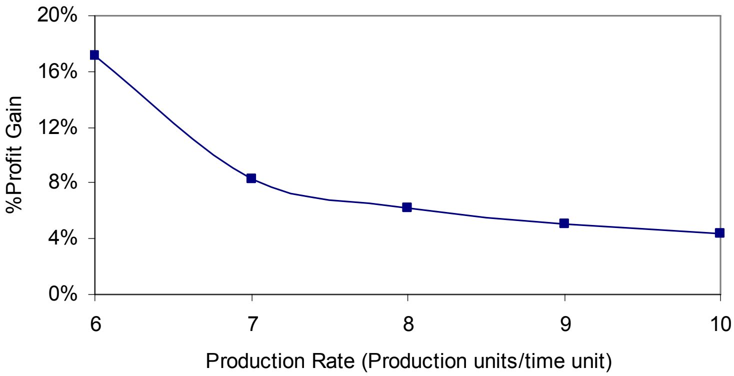

Figure 15.4 The effect of production rate on % profit gain

15.4.2.3 The Effect of Unit Inventory Holding Cost (h_i)

Our experiment shows that the effect of unit inventory holding cost on the value of postponement is in accordance with what is reported in most of studies on postponement (*e.g.*, Gupta and Benjafaar 2004, Wong *et al.* 2009). As illustrated in Figure 15.5 the value of postponement is increasing with the unit inventory holding cost.

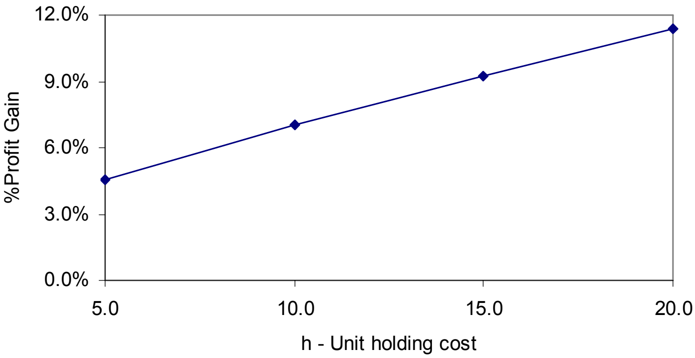

Figure 15.5 The effect of unit holding cost on % profit gain

15.4.2.4 The Effect of Customers' Disutility on Waiting (c_t)

Figure 15.6 shows how the relative profit differences between MTS and DD as a function of the customers' disutility cost of waiting. It is shown that the average profit gain first increases and then decreases. Figure 15.7 is also provided to depict the average profits for the two configurations. For the DD configuration, the average profit steadily decreases in the range of cost values used in the experiment. But for the MTS configuration, the average profit first decreases before reaching a plateau. When c_t increases, the MTS system will reduce the lead time by holding a higher inventory level. There is, however, a point where the inventory level is sufficiently high to allow a zero lead time. From this point on, the profit of MTS will not change while the profit of DD still decreases. This observation suggests that postponement benefits would vanish when customers are more sensitive to delivery lead times. In the extreme case where customers really want to get their product instantly (and customised attributes are less important), postponement is obviously not a recommendable strategy.

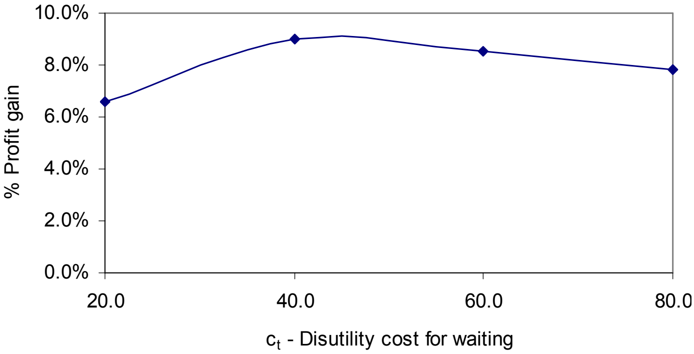

Figure 15.6 The effect of waiting cost on % profit gain

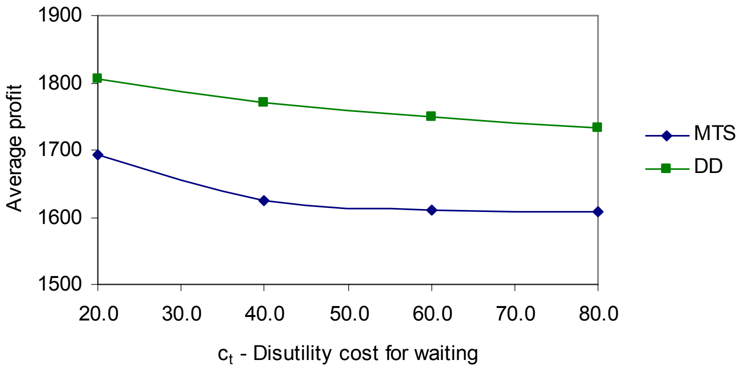

Figure 15.7 The effect of waiting cost on average profit

15.4.2.5 The Effect of Transportation Cost (c_x)

Figures 15.8 and 15.9 are presented to explain the effect of transportation cost on postponement benefits. Recall that the transportation cost represents customers' dissatisfaction in not getting their ideal preferences. As depicted in Figure 15.7 the profit gain is higher when the transportation cost increases. By holding the common intermediate goods and executing the final customisation later, the DD configuration would enable minimising the customers' transportation cost by offering more product variants in the market than the MTS configuration, as depicted in Figure 15.8. The proven success of Dell suggests that to a certain extent PC customers seem to have good appreciation of the introduced customisation feature, reflecting the possible existence of high transportation costs. But there are also applications in mass customisation, *e.g.*, customised shoes (Berger 2003) where we conjecture that these customised products serve only a niche market and the total market is still dominated by the mainstream products. It is not well understood whether the difference in the adoption level of the mass customisation concept is due to the difference in the transportation cost. Empirical research that attempts to assess and compare the transportation cost for different products would certainly be worthwhile.

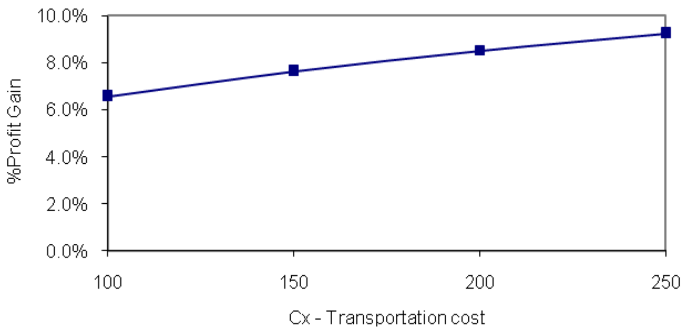

Figure 15.8 The effect of transportation cost on % profit gain

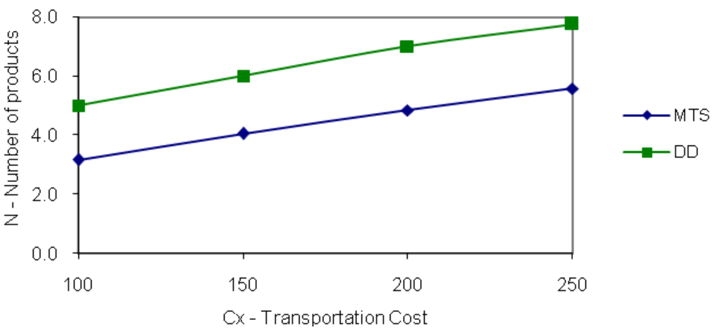

Figure 15.9 The effect of transportation cost on the number of products

15.5 Conclusions

Postponement has been recognised as an important technique that has great potential in creating supply chain improvement through reduction in uncertainty and cost while satisfying customer needs. By restructuring the production and distribution of products in such a way that the customisation of these products is made as close as possible to the point when the demand is known, postponement can greatly improve the flexibility capabilities of the firms that employ it. Today, where more and more industries move towards creating markets of one, such significant flexibility improvement is important in accommodating mass customisation strategies.

This chapter is an attempt to obtain a better understanding of postponement benefits by developing a new perspective that considers cross-functional implications and coordination issues. In contrast to the vast majority of existing studies on postponement where the focus has been on the production-inventory system, the new perspective we developed tries to address the important alignment between the production-inventory and marketing functions. Under this integrated perspective, we are able to get a more complete view on how postponement may enhance firms' profitability by capturing interactions among many factors including inventory, lead time, price and product variation.

The stylised models presented in this chapter allow us to find out how postponement benefits could be different when assessed by the proposed integrated approach as opposed to the traditional production-inventory focused approach. The evaluation based on the cost minimisation strategy used under the production-inventory focused approach would reveal how much inventory cost savings can be gained by employing postponement. However, such an evaluation is helpful only when the main intent is to evaluate postponement benefits under exogenously predetermined demand and product variety. As such, this traditional approach is not capable of reflecting the more realistic and complex problem that may involve marketing-related aspects in the sense that customers' demands are actually influenced by price, lead time and product variation. The proposed integrated approach is able to overcome such limitations. Through numerical examples we show that the inventory cost minimisation problem is not equivalent to the profit maximisation problem. In extreme situations such as in markets where customers are sensitive to delivery lead time, the benefit of postponement in terms of inventory cost savings may need to be offset by some costs to compensate longer lead times, which can be reflected in lower prices. This kind of observation is only possible if we use the integrated approach.

Through the numerical experiment we demonstrate how different system parameters may have an impact on the benefit of postponement measured by the relative difference of profits for the MTS and DD configurations. It is shown that the production rate increase has a positive effect on the average profit of the two configurations. However, the benefit of postponement appears to diminish as production rate increases. Our research also confirms what has been reported in most of the postponement studies: the benefit of postponement increases in line with unit inventory holding cost. This suggests that postponement is more beneficial when products have high inventory costs.

The effects of the marketing factors are also examined. The results show that the average postponement benefit first increases and then decreases with the customers' disutility cost for waiting. This observation suggests that postponement would not be appropriate in a market where customers are highly sensitive to delivery lead times.

Finally, the effect of the transportation cost on the value of postponement appears to be obvious. The value of postponement is higher when the transportation cost increases. As the transportation cost represents the customers' dissatisfaction with not getting their ideal preferences, postponement would allow the DD configuration to minimise customers' transportation costs by offering more product variants in the market than the MTS configuration. All in all, the results obtained from our study are still in line with mainstream findings suggesting that postponement may lead to significant benefits. However, the integrated approach allows us to have a better view, especially when the prevalent lead time *versus* variety trade-off comes into play.

Like all models, ours has limitations. First, our results rely on the assumption that customers are heterogeneous only in terms of their ideal preference. Customers are assumed to be homogeneous in terms of their reservation price, transportation and waiting costs. In situations where customer heterogeneity is not only limited to the ideal preference, some key insights may change. For example, if consumers are allowed to have uncommon reservation price for their ideal products, the full market coverage may no longer be optimal. Second, our model ignores competition. In particular, our model is concerned with postponement evaluation in that it ignores the existence of competition between standard and customised products in the market. We believe that incorporating an extended customer heterogeneity and product competition may prove to be fruitful in future.

Some other opportunities for future research arise from this work. As we stated earlier, empirical research to assess different parameters in real settings is a challenge. In particular, research to estimate customers' disutility associated to lead time as well as to the deviation between their ideal taste and what is offered would be very valuable. This research would not only be useful in identifying postponement benefits on a more realistic scale, but also in getting a better explanation of why some mass customisation practices are successful while others are not. Another opportunity is to extend the concept of product differentiation. While this chapter focuses only on horizontal product differentiation, research that also considers vertical differentiation certainly warrants attention. This is true as for many products such as electronic gadgets, PCs or bicycles; customisation would involve both vertical and horizontal differentiation.

Last but not least, it may be worth highlighting that this chapter represents one of the very few studies addressing issues that lie within the interface of the operations management and marketing disciplines. We believe that, particularly in the context of mass customisation, much work still needs to be done and most of it would require multi-disciplinary efforts involving expertise from these two domains.

References

Alderson W (1950) Marketing efficiency and the principle of postponement. Cost and Profit Outlook, September

Alptekinoglu A, Corbett C (2007) Leadtime-variety tradeoff in product differentiation. Proceedings of World Conference on Mass Customization and Personalization, MIT

Aviv Y, Federgruen A (2001a) Design for postponement: a comprehensive characterisation of its benefits under unknown demand distributions. Operations Research, 49: 578–598

Aviv Y, Federgruen A (2001b) Capacitated multi-item inventory systems with random and seasonally demand fluctuating demands: implications for postponement strategies. Management Science 47: 512–531

Benjaafar S, Kim JS, Vishwanadham N (2004) On the effect of product variety in production-inventory systems. Annals of Operations Research 126: 71–101

Berger C, ed. (2003) Mass customisation – an Adidas perspective. IEE Seminar on Mass Customisation: Turning Customer Differences into Business Advantages (Digest No. 2003/10031)

Berry D, Naim M (1996) Quantifying the relative improvements of the redesign strategies in a PC supply chain. International Journal of Production Economics 46–47: 181–196

Boone CA, Craighead CW, Hanna JB (2007) Postponement: an evolving supply chain concept. International Journal of Physical Distribution and Logistics Management 37: 594–611

Bowersox DJ, Closs DJ (1996) Logistical Management: The Integrated Supply Chain Process. McGraw–Hill, New York

Brown AO, Lee HL, Petrakian, R (2000) Xilinx improve its semiconductor supply chain using product and process postponement. Interfaces 30: 65–80

Buzacott J, Shanthikumar JG (1993) Stochastic Models of Manufacturing Systems. Prentice-Hall, Upper Saddle River, NJ

Dapiran P (1992) Benetton – Global logistics in action. International Journal of Physical Distribution and Logistics Management 22: 7–11

De Groote X (1994) Flexibility and marketing/manufacturing coordination. International Journal of Production Economics 36: 153–167

Dobson G, Yano CA (2002) Product offering, pricing, and make-to-stock/make-to-order decisions with shared capacity. Production and Operations Management 11: 293–312

Feitzinger E, Lee HL (1997) Mass customisation at Hewlett Packard: the power of postponement. Harvard Business Review 75: 116–121

Garg A, Tang CS (1997) On postponement strategies for product families with multiple points of differentiation. IIE Transactions 29: 641–650

Gaur V, Honhon D (2006) Assortment planning and inventory decisions under a locational choice model. Management Science 52: 1528–1543

Gupta D, Benjaafar S (2004) Make-to-order, make-to-stock, or delayed product differentiation? A common framework for modelling and analysis. IIE Transactions 36: 529–546

Hotelling H (1929) Stability in competition. The Economic Journal 39: 41–57

Jiang K, Lee HL, Seifert, RW (2006) Satisfying customer preferences via mass customization and mass production. IIE Transactions 38: 25–38

Lancaster K (1990) The economics of product variety: A survey. Marketing Science 9: 189–206

Lee HL, Billington C, Carter B (1993) Hewlett Packard gains control of inventory and service through design for localisation. Interfaces 23: 1–11

Lee HL, Tang CS (1997) Modelling the costs and benefits of delayed product differentiation. Management Science 43: 40–53

Li S, Ragu-Nathan B, Ragu-Nathan TS, Subba Rao S (2006) The impact of supply chain management practices on competitive advantage and organizational performance. Omega 34: 107–124

McCutcheon, DM, Raturi, AS, Meredith, JR (1994) The customization-responsiveness squeeze. Sloan Management Review 35:89–99

Pagh JD, Cooper MC (1998) Supply chain postponement and speculation structures: how to choose the right structure? Journal of Business Logistics 19: 13–34

Rabinovich E, Evers PT (2003) Postponement effects on inventory performance and the impact of information systems. International Journal of Logistics Management 14: 33–48

Signorelli S, Hesket JL, Benetton A (1984) Harvard Business School Case, 9–685-014

Squire B, Brown S, Readman J, Bessant J (2006) The impact of mass customization on manufacturing trade-offs. Production and Operations Management 15: 10–21

Su JCP, Chang Y, Ferguson M (2005) Evaluation of postponement structures to accommodate mass customization. Journal of Operations Management 23: 305–318

Swaminathan JM, Lee H (2003) Design for postponement, in Handbooks in Operations Research and Management Science: Supply Chain Management: Design, Coordination and Operation, Eds: Graves S, de Kok, E. Elsevier 199–228

Swaminathan JM, Tayur S (1998) Managing broader product lines through form postponement using vanilla boxes. Management Science 44: S161–S172

Syam NB, Kumar N (2006) On customized goods, standard goods, and competition. Marketing Science 25: 525–537

Thonemann UW, Bradley JR (2002) The effect of product variety on supply-chain performance. European Journal of Operational Research 143: 548–569

Van Hoek RI (2001) The rediscovery of postponement: a literature review and directions for research. Journal of Operations Management 19: 161–184

Waller MA, Dabholkar PA, Gentry JJ (2000) Postponement, product customization, and market-oriented supply chain management. Journal of Business Logistics 21: 133–159

Wong H, Wikner J, Naim M. (2009) Analysis of form postponement based on optimal positioning of the differentiation point and stocking decisions. International Journal of Production Research 47: 1201–1224

Yang B, Burns ND, Backhouse CJ (2004) Postponement: a review and an integrated framework. International Journal of Operations and Production Management 24: 468–487

Zinn W, Bowersox DJ (1988) Planning physical distribution with the principle of postponement. Journal of Business Logistics 9: 117–136