Discussion Paper 1222

Postponement And Information In A Supply Chain

Krishnan S. Anand
J.L.Kellogg Graduate School of Management
Northwestern University
Evanston, IL 60208-2001

Haim Mendelson
Graduate School of Business
Stanford University
Stanford, CA 94305-5015

July, 1998

Abstract

We model a supply chain consisting of a production facility, a distribution center and two differentiated markets. Demand information is used to mitigate the effects of uncertainty in the output markets. We study the firm's operational performance under alternative business processes, comparing *early* and *delayed* product differentiation. The comparison yields the *value of postponement*. Our results show that informational considerations have a paramount effect on the effectiveness of postponement strategies. They enable researchers and decisionmakers to perform cost-benefit analyses and quantify the anticipated effects of implementing postponement strategies. They also provide qualitative guidance regarding factors that affect the value of postponement.

1 Introduction

Demand uncertainty is an important feature of contemporary business environments. Proliferation in product varieties [13], accelerated clockspeed [27] and ~~the volatility of~~ the global marketplace increase demand uncertainty, amplifying the tradeoff between inventory-levels and service-levels: inventories can improve firms' responsiveness to customer demand, but this is accomplished at a cost. Some firms have successfully employed "accurate response" strategies to reduce demand uncertainty, usually involving the extensive use of Information Systems (IS) for data collection and demand forecasting (cf. [2]). Nonetheless, uncertainty remains an important factor, especially in fast-clockspeed industries like hi-tech or industries with long production and distribution lead-times like the apparel industry. Firms are thus restructuring the business processes that underlie their supply chains to better cope with demand uncertainty. One strategy that has received considerable attention is design for postponement, or delayed product differentiation.

Consider the process used to make multiple related products, say widgets that are differentiated by color. At the *point of differentiation*, the widgets pass through a painting station and become differentiated (in this case by color). All processes upstream to the point of differentiation are identical for all widgets: a widget gains its particular identity only at that point. Beyond the point of differentiation, the entire order fulfillment process (i.e., both the firm's operations and its information systems) must distinguish between the widgets based on their different colors.

Differentiation can apply in a similar fashion to different product attributes, target market segments and sales regions. *Postponement* is a strategy that delays the point of differentiation, moving it further downstream, often by restructuring the supply chain. The result is increased flexibility in meeting uncertain and changing customer demands. While this concept has its roots in the manufacturing sector, services (for example, fast food) are beginning to apply postponement to their processes as well.

Postponement enables the firm to get the best (or, rather, avoid the worst) of make-to-stock and make-to-order policies. Under a make-to-stock policy, demand needs to be estimated accurately, since forecasting errors are severely punished: in industries like high-tech or apparel, a product can lose 1% of its value per week,[1] and the loss rate is even higher for perishable-goods. Under make-to-

[1] See, for example, [29].

order, long lead times make it difficult for the firm to compete. Postponement is a *via media*: lead times are reduced by making an intermediate good to stock, while inventory markdowns as well as lost sales are also reduced by customizing the intermediate good based on observed demand or order patterns. These improvements can be achieved simultaneously due to better use of information.

Compaq's and IBM's recently-announced initiatives to switch from an exclusively make-to-stock policy to a make-to-order policy for their PCs is expected to facilitate delayed customization, thereby reducing inventories and countering rival Dell Corporation's direct-sale model. Under the make-to-stock policy, when the specific configuration demanded by customers changed, Compaq and IBM had to buy back unsold machines from dealers, compensate them and/or lower prices on unsold products. By reducing these costs through make-to-order, they hope to be able to lower prices by as much as 15 to 29 percent [33, 34, 28]. Benetton provides a celebrated example of using postponement to cope with long production lead-times and fickle fashion trends, using undyed yarn to knit about half of its clothing (cf. [6, 20, 23]). Dyeing is thus postponed to a later stage, when Benetton has a better idea of the popular colors for the season.

Another well-known example of postponement, due to Hau Lee and collaborators, is Hewlett-Packard (HP) [12, 17, 18, 19, 20, 21, 22, 23, 25]. HP manufactured its Deskjet-Plus printers in its Vancouver, Washington Division, and shipped the printers to three Distribution Centers (DCs) in North America, Europe and Asia. The transit time by sea, to the two non-U.S. DCs, was about a month. Depending on the eventual destination country, different power supply modules had to be installed in the printers to accommodate local voltage, frequency and plug conventions. The manuals and labels also had to be localized due to language differences. HP redesigned the printer so that the power module could be added as a simple plug-in, manufactured a generic Deskjet-Plus printer in the U.S. (sans power supply module, manual and labels) and later localized the generic product in Europe, based on observed demand conditions. Restructuring its printer production process in this fashion enabled HP to maintain the same service-levels with an 18% reduction in inventory, saving millions of dollars (Lee, Billington and Carter [22]).

This paper develops an analytical model of postponement under demand uncertainty, extending the extant academic literature on postponement (cf. Aviv and Federgruen[5], Eppen and Schrage [9], Federgruen and Zipkin [10], Gavirneni and Tayur[14], Lee [17], Lee and Tang [23], Lee and Whang [25], Schwarz [30], Swaminathan and Tayur[32]) that focuses on cost-minimization strategies. Lee and Tang [23] minimize the sum of investment, processing and inventory costs subject to service-level constraints for each stage in a multi-stage process. In other models, the total cost to be minimized is the sum of production or ordering costs, holding costs, and shortage costs due to unmet demand [14, 32] or backorder costs due to backlogging [5, 17, 25]. Earlier models [9, 10, 30] show how delayed allocation decisions create "statistical economies of scale" and analyze the associated tradeoffs.

Lee [17] and Lee and Tang [23] consider a stationary, i.i.d Normal demand process. Lee [17] analyzes the inventory reduction due to pooling, enabled by postponement. Lee and Tang [23] address additional factors such as redesign costs and leadtimes. They describe three possible approaches to delaying differentiation - standardization, modular design and process restructuring, and find that *ceteris paribus*, it is better to advance long lead-time, low value-added activities when restructuring the production process for postponement. Gavirneni and Tayur [14] model a supplier with two customers, comparing the relative benefits of customer-inventory information and customer-based delayed product differentiation, as the holding cost, capacity, demand variance and customer heterogeneity are varied. Aviv and Federgruen [5] and Lee and Whang [25] emphasize the role of postponement in improving forecasts. Aviv and Federgruen [5] analyze the value of delaying differentiation for the case of unknown demand distributions, where demand forecasts are updated in a Bayesian fashion, and characterize the structure of near-optimal ordering rules. Lee and Whang [25] model the demand process as a non-stationary random walk. They distinguish between two factors that affect the value of postponement: "uncertainty resolution" up to the point of differentiation (following Eppen and Schrage [9]) and "forecast improvements", i.e., the ability to make more accurate forecasts for a period as it gets closer, due to non-stationary demand. Swaminathan and Tayur [32] compare the benefits of stocking "Vanilla boxes" (semi-finished products) with make-to-stock or assemble-to-order strategies under assembly capacity constraints. They develop an algorithm that simultaneously determines the composition and optimal inventory levels of the vanilla boxes.

We consider the interaction of material flows and information flows in a supply chain, modeling the precision and timing of Information Systems (IS). In our model, the demands for the end-products are stochastic functions of the corresponding prices. Both prices and quantities (and hence, revenues) are functions of the firm's cost parameters; hence the interplay between marketing (pricing, sales quantities and revenues) and operational (production and inventory) decisions can be studied. We compare the inventories, sales and profits under early and delayed product differentiation, quantify the value of postponement, and study their relationship to the firm's IS.

The operational benefits of postponement are related to the firm's ability to use its IS to capture relevant market information for demand forecasting. In general, two key principles govern effective forecasting: *(i)* The shorter the time horizon over which prediction is made, the more accurate is the forecast (cf. Lee and Whang [25]); and *(ii)* Aggregate forecasts are more accurate than disaggregated ones. The first principle reflects the build-up of uncertainty over time; the second is a consequence of the law of large numbers. Postponement exploits *both* of these principles. First, delayed differentiation implies that the demand forecast needed initially is at a more aggregate level. For example, at the time of manufacture, HP needs to predict the demand for its printers over all of Europe (or Asia), and Benetton needs to predict the demand for all colors of a particular style. Second, disaggregated forecasts are needed only following the point of product differentiation, when the forecast horizon is shorter.

To study the value of postponement, it is necessary to open the "black box" known as the supply chain. We model a stylized supply chain consisting of a production facility, a distribution center (DC) and two markets where the ultimate product is sold. We consider two alternative business processes for managing this supply chain. Under the *Early Differentiation* (ED) process, the firm differentiates its products at the production stage. As a result, the two products are already differentiated through all the subsequent stages of the supply chain. Under the *Delayed Differentiation* (DD) process, the firm delays product differentiation until further information about the demand for each product is received. We apply these alternative business processes to the same demand and cost environment. We derive the optimal production, shipping and inventory policies for each process in a dynamic (multiperiod) setting, and compare the resulting production, inventory and sales. We analyze the factors that determine profits under each process and study the *value of postponement* (VOP), defined as the difference between the two expected profits. We also study the effect of changes in the demand variability, demand correlation, information precision and inventory holding costs on the VOP.

We prove that the optimal production, sales and inventory policies under both ED and DD are myopic, and derive the optimal solutions in closed form. While the sales strategies are not identical under ED and DD, expected sales are the same. We also prove that the average inventory is lower under DD, i.e., postponement reduces average inventory. In fact, although both ED and DD profits clearly fall as the holding cost increases, we find that the VOP *increases* in the holding cost.

We find that demand variability is an important factor in determining the VOP. For any IS, the VOP is *increasing* in demand variability, even though both ED and DD profits individually fall with variability. Thus, postponement is particularly effective in highly dynamic environments.

Postponement is further affected by the demand correlations. While greater correlation increases ED profits, its effect on DD profits is ambiguous. However, the net effect on VOP is unambiguous: it falls with the demand correlation.

Our results demonstrate that the effectiveness of postponement strategies is closely linked to the quality of demand information provided by the firm's IS. Two key attributes of information quality are *precision* and detail, or *disaggregation*. Our results show that more precise demand information increases profits under both ED and DD, as might be expected. Further, the VOP also increases with IS precision. Specifically, we show that with greater precision, delayed product differentiation can better exploit the available information about demand. In addition, we show that disaggregation is an important precondition for a positive VOP.

The rest of this paper is organized as follows. In Section *2*, we present our two models of the alternative business processes. We derive the optimal solutions for each in Section *3*. In Section *4*, we compare the optimal policies, sales, inventories and profits under each model, and discuss the effect of information (and hence, forecast) aggregation on the performance of each. We then develop a numerical example to shed further insights. Section *5* concludes with a discussion of possible extensions and future research.

2 Alternative Business Processes

Consider a firm that produces and sells two related products. For concreteness, we assume that the products are sold in two distinct product markets (our model is equally applicable when a single product is sold in geographically separated markets). The firm's supply chain consists of a production facility, a Distribution Center (DC) and two retail outlets, one for each market. The firm's objective is to maximize its discounted expected profits over the long-term (infinite) horizon, where the one-period discount factor is β.

Our specification of demand reflects a tradeoff between the quantity sold in each market and the market price. Unlike traditional models of manufacturing operations, supply chain management models call for an analysis of the impact of the firm's operational decisions on the *entire* supply chain. Thus, "demand" does not represent an exogenous quantity; rather, the firm optimizes the entire supply chain, tying together production, logistics and pricing decisions.

The supply chain is held together by two types of "glue": physical (inventory) and virtual (information). With respect to the former, we allow the firm to hold inventory at the DC. We assume that any quantity of intermediate goods may be held at the DC, incurring an inventory holding cost of h per unit per period.[2]

The informational building blocks of the model focus on demand information. We assume that the state of demand in each market i is a binary random variable S_i, with $S_i = 1$ corresponding to a "high" demand state and $S_i = 0$ corresponding to a "low" demand state. Either state may occur with probability $\frac{1}{2}$. The demand curves are linear with a random intercept. The demand intercept in a market with the "high" demand state is a_H, and in the "low" demand state it is a_L, where $a_H > a_L$. The corresponding revenue functions are $R_H(.)$and $R_L(.)$, with

$$R_H(q) = (a_H - bq)q \quad \text{and} \quad R_L(q) = (a_L - bq)q.$$

The demand states in the two markets may be correlated, with correlation coefficient ρ. When the two markets represent geographically-separated retail locations for the same physical product (e.g., laser printers sold in Europe vs. the US), this correlation reflects the effects of common demand patterns on the localized markets. When the markets are for related products (that is, the separation is in product space, e.g., black-and white vs. color laser printers; or apparel with different colors), the correlation coefficient quantifies the degree to which information about the state of demand for one product can help predict the demand for the other. In both situations, the correlation is typically non-negative, and we assume this is the case through the rest of this paper.[3]

[2]The model is altered in a straightforward fashion if the DC has only a finite storage capacity.

[3]Extending the analysis to negatively correlated demands is straightforward. We focus on the more common case of $\rho \geq 0$ to avoid clutter.

The joint distribution of the vector of demand states (S_1, S_2) is given by

$$P[S_1 = S_2 = 1] = P[S_1 = S_2 = 0] = \frac{1+\rho}{4},$$

and

$$P[S_1 = 1,\ S_2 = 0] = P[S_1 = 0,\ S_2 = 1] = \frac{1-\rho}{4}.$$

The firm has a binary *Information System* (IS) $\underline{X} = (X_1, X_2)$ that helps it infer the states of demand in the two markets. The IS is modeled as a two-dimensional vector of binary signals, one from each market, taking on the values 0 or 1. These signals are received at the DC after the production decision has been implemented, but prior to the decision on the quantities to be refined and shipped to each market. The DC uses this information in implementing its refinement, distribution and inventory strategies, updating the probability distribution of the states of the two markets in a Bayesian fashion.

Each signal X_i represents the state of demand in market i with noise: $X_i = S_i$ with probability $(1-\alpha)$ and $(1-S_i)$ with probability α $(i = 1, 2)$. The parameter α is a measure of the imprecision of the firm's IS, with a smaller α corresponding to more informative signals. We assume, without loss of generality, that $0 \le \alpha \le \frac{1}{2}$. This information structure is similar to the one modeled in Anand and Mendelson [2], where the supply chain was a "black box", the model had a single period and demands were assumed independent across markets.

The sequence of events is thus as follows (see Figure 1). At the beginning of each period, intermediate goods are delivered to the DC. Then, the DC obtains the signal $\underline{X}$ and uses this information, and the available quantities, to decide how much of the intermediate goods are to be refined to end-products for each market. The refinement cost for either product is θ per unit; θ also subsumes any transportation costs from the DC to the respective markets. After shipment, the demand curves for each market are realized and sales are made. The firm then makes its production decisions for delivery to the DC next period. The production cost is k per unit of intermediate good. This sequence is repeated period after period, with the (two-dimensional) binary random vectors representing the market states and corresponding signals for the two markets being i.i.d.

We denote by $P(x_1, x_2)$ the probability that $X_1 = x_1$ *and* $X_2 = x_2$, and by $P_{x_1 x_2} = Pr[S_1 = 1 \mid X_1 = x_1,\ X_2 = x_2]$ the probability that a market is in the "high-demand" state, when x_1 is the signal from that same market and x_2 is the signal from the *other* market.[4]

We make the technical assumption that $\frac{a_H+a_L}{2} > \frac{k}{\beta} + \theta$, which guarantees that it is optimal for the firm to produce positive quantities of the products. We also assume that the production cost k is greater than the inventory holding cost h; otherwise, it is never optimal to hold any inventory of either product.

Suppose that the signals received for the two products are $X_1 = x_1$ and $X_2 = x_2$. The expected revenues from shipping quantity q_1 of product 1 are $P_{x_1x_2} \cdot R_H(q_1) + (1 - P_{x_1x_2}) \cdot R_L(q_1)$, and the expected revenues from shipping quantity q_2 of product 2 are $P_{x_2x_1} \cdot R_H(q_2) + (1 - P_{x_2x_1}) \cdot R_L(q_2)$. The shipment quantities q_1 and q_2 are constrained by the quantities of the intermediate good(s) available at the warehouse. For convenience, define

$$f(P_{x_1x_2}) = P_{x_1x_2} \cdot a_H + (1 - P_{x_1x_2})a_L. \tag{1}$$

Thus, for example, $f(P_{10})$ is the expected value of the demand intercept for product 1 when $X_1 = 1$ and $X_2 = 0$.

The firm's decision problem can now be formulated as follows. Let $Q^{1,t}$ and $Q^{2,t}$ denote the production quantities of products 1 and 2 for period t, and let $q^{1,t}$ and $q^{2,t}$ denote the corresponding shipment quantities to the respective markets. Also let $\xi^{1,t}$ and $\xi^{2,t}$ denote the intermediate good inventories carried over from period t to the next period. Since the market states are independent over time, inventories completely specify the firm's state at the start of each period. Let $\underline{X}_t = (X_{1,t}, X_{2,t})$ be the vector of information signals received from the two markets in period t. Correspondingly, let $R^{1,t}(\cdot|\underline{X}_t)$ and $R^{2,t}(\cdot|\underline{X}_t)$ denote the conditional expected revenue functions from sales in markets 1 and 2 in period t. Suppose that the supply chain starts operating at $t = 0$ with zero inventories (i.e., $\xi^{1,0} = \xi^{2,0} = 0$). The firm's problem is to maximize its expected

[4]The derivations of the conditional state distributions $P_{x_1x_2}$ and the proofs of all Theorems are available with the authors.

discounted profits, namely

$$\max_{\substack{Q^{1,0}, Q^{2,0}, \{Q^{1,t}, Q^{2,t}\}_{t=1}^{\infty} \\ \{q^{1,t}, q^{2,t}, \xi^{1,t}, \xi^{2,t}\}_{t=1}^{\infty}}} \left\{ -k\left(Q^{1,0}+Q^{2,0}\right) + E\left[\sum_{t=1}^{\infty} \beta^t \cdot \left(R^{1,t}(q^{1,t}|\underline{X}_t) + R^{2,t}(q^{2,t}|\underline{X}_t)\right)\right] - \left[\sum_{t=1}^{\infty} \beta^t \left(\theta(q^{1,t}+q^{2,t}) + h\cdot(\xi^{1,t}+\xi^{2,t}) + k\cdot\left(Q^{1,t}+Q^{2,t}\right)\right)\right]\right\} \quad (2)$$

subject to (i) non-negativity constraints on all the decision variables, and (ii) the inventory transition law, which states that for any item, the inventory in each period is equal to the sum of the inventory and production from the previous period less the shipment quantity in the current period.

This formulation is common to both the Early Differentiation (ED) and Delayed Differentiation (DD) processes. We next address the differences between the two.

2.1 Early Differentiation (ED) Process

Under ED, the products are already segregated in the production stage, resulting in non-substitutable intermediate goods. Thus, the production quantities of the intermediate goods are separately determined for *each* product. In the second (refinement) stage, the decisions to be made involve the refinement, shipment and inventory quantities for each product. The firm's problem under ED is thus to solve the maximization problem (2) subject to non-negativity constraints on the decision variables $Q^{1,0}, Q^{2,0}$ and $\{Q^{1,t}, Q^{2,t}, q^{1,t}, q^{2,t}, \xi^{1,t}, \xi^{2,t}\}_{t=1}^{\infty}$, and the inventory transition laws given by

$$\xi^{1,t} = \xi^{1,t-1} + Q^{1,t-1} - q^{1,t}, \text{ and} \quad (3)$$

$$\xi^{2,t} = \xi^{2,t-1} + Q^{2,t-1} - q^{2,t}, \quad (4)$$

where $\xi^{1,0} = \xi^{2,0} = 0$. Clearly, this problem is separable across the two products – a direct result of early differentiation.

2.2 Delayed Differentiation (DD, or Postponement) Process

Under the DD process, manufacturing is common to the two products. Thus, at the production stage the firm has to decide only on the production quantity of the common intermediate good. Later, in the refinement stage, each unit of the common intermediate good can be customized at the DC to produce a unit of either good.

The objective function under DD is given by expression (2). However, the additional flexibility offered by the common intermediate good is reflected in the more flexible transition law.

$$\left(\xi^{1,t} + \xi^{2,t}\right) = \left(\xi^{1,t-1} + \xi^{2,t-1}\right) + \left(Q^{1,t-1} + Q^{2,t-1}\right) - \left(q^{1,t} + q^{2,t}\right). \tag{5}$$

Condition (5) replaces conditions (3) and (4) of the ED model; the other (non-negativity) constraints are the same for both ED and DD. Transition law (5) means that the problem under DD is no longer separable across the two products.

To simplify the problem representation under DD, let $\xi^t = \xi^{1,t} + \xi^{2,t}$ and $Q^t = Q^{1,t} + Q^{2,t}$. The firm's problem under DD now simplifies to:

$$\max_{Q^0,\{Q^t,q^{1,t},q^{2,t},\xi^t\}_{t=1}^{\infty}} E\left[-kQ^0 + \sum_{t=1}^{\infty} \beta^t \cdot \left(R^{1,t}(q^{1,t}|\underline{X}_t) + R^{2,t}(q^{2,t}|\underline{X}_t) - \theta(q^{1,t} + q^{2,t}) - h \cdot \xi^t - k \cdot Q^t\right)\right],$$

subject to the non-negativity constraints on the decision variables $Q^t, \xi^t, q^{1,t}$ and $q^{2,t}$, the inventory transition law $\xi^t = \xi^{t-1} + Q^{t-1} - q^{1,t} - q^{2,t}$, and the initial condition $\xi^0 = 0$.

We define the value of postponement (VOP) as the difference in discounted expected profits between the DD and ED processes. Comparing conditions (3) and (4) with (5), it is easy to see that the DD problem is a relaxation of ED, hence the VOP is non-negative. The more flexible inventory transition law reflects the DC's greater flexibility under DD. Under ED, the firm commits early to the production quantities of each product, and the DC cannot alter the sales mix in response to demand information as much as it can under DD. However, the firm might prefer the ED structure due to cost considerations not directly addressed in our model: for example, higher production costs under DD, or the cost of redesigning the products and restructuring the supply chain for postponement. As we demonstrate below, there are circumstances under which postponement does not add enough value to justify its costs.

3 Optimal Solutions

We now derive the firm's optimal policies under both ED and DD. We show that in both cases, the optimal policy is *myopic*, i.e., the decisions taken each period can disregard the consequences for future periods (cf. Heyman and Sobel [15, Chapter 3]). General sufficiency conditions for the optimality of myopic policies have been derived in Kirman and Sobel [16] and Sobel [31]. Myopic optimal policies when Sobel's [31] conditions are not met are a rare occurrence in such dynamic models (a notable exception being Amihud and Mendelson [1]). We show that the optimal policies for our problem are myopic even though our model does not satisfy Sobel's [31] conditions, and we derive them in closed form.

3.1 Early Differentiation

Theorem 1 (Optimal Policies under ED) The following myopic production, shipment and inventory policies (for each product) are optimal:

(A) Shipping/Inventory:

In each period, conditional on the information signals $X_1 = x_1$ and $X_2 = x_2$, the DC refines and ships its entire on-hand quantity of product 1 up to the threshold $q^{ED}_{x_1x_2} = \frac{f(P_{x_1x_2})-(k+\theta-h)}{2\cdot b}$, where $f(\cdot)$ was defined in (1). The threshold for product 2 is $q^{ED}_{x_2x_1} = \frac{f(P_{x_2x_1})-(k+\theta-h)}{2\cdot b}$. Any surplus above these thresholds is held as inventory.

(B) Production:

Each period, the firm produces to bring the DC's intermediate good inventories for the next period (for each product) to Q^*_{ED}, where $Q^*_{ED} = \max\Big\{Q^*_{1,ED}, Q^*_{2,ED}, Q^*_{3,ED}, Q^*_{4,ED}\Big\}$, and

$$\begin{aligned}
Q^*_{1,ED} &= \frac{f(P_{11}) - (k+\theta-h) - \frac{\left(\frac{k}{\beta}-k+h\right)}{P(1,1)}}{2\cdot b},\\
Q^*_{2,ED} &= \frac{2\cdot[P(1,1)f(P_{11}) + P(1,0)f(P_{10})] - \left(2\cdot\frac{k}{\beta} - k + h + \theta\right)}{2\cdot b},\\
Q^*_{3,ED} &= \frac{\frac{P(1,1)f(P_{11})+(1-2\cdot P(1,1))f(\frac{1}{2})-\left(\frac{k}{\beta}-k+h\right)}{1-P(1,1)} - (k+\theta-h)}{2\cdot b}, \text{ and}
\end{aligned}$$

$$Q^*_{4,ED} = \frac{f(\frac{1}{2}) - \left(\frac{k}{\beta}+\theta\right)}{2\cdot b}. \quad \blacklozenge$$

To interpret the solution, first consider the firm's optimal shipment/inventory policy for product 1. The expected revenue from shipping the quantity q is $(f(P_{x_1x_2}) - b\cdot q)\cdot q - \theta\cdot q$, hence the expected marginal revenue is $f(P_{x_1x_2}) - 2b\cdot q - \theta$. Moreover, the marginal cost of replenishing the warehouse for the next period is k per unit, and the savings from holding a unit less at the warehouse is h. Thus, given the choice between shipping an additional unit of product 1 versus holding it in inventory, the shipping option is preferable as long as the expected marginal revenue from that unit is greater than the additional cost; i.e., as long as $f(P_{x_1x_2}) - 2b\cdot q - \theta \geq k - h$, or equivalently, $q \leq \frac{f(P_{x_1x_2})-(k+\theta-h)}{2b} = q^{ED}_{x_1x_2}$. Thus, $q^{ED}_{x_1x_2}$ is the maximum quantity of product 1 that the DC would ever ship when the signals are x_1 and x_2; above this threshold, the firm is better off keeping inventory.

The production policy is a stationary threshold policy, independent of demand information by necessity. The threshold value Q^*_{ED} is shaped by the shipment policy. For each product, there are four possible ship-up-to levels based on the four possible combinations of the binary market signals X_1 and X_2. The ship-up-to levels are related as $q^{ED}_{11} \geq q^{ED}_{10} \geq q^{ED}_{01} \geq q^{ED}_{00}$; hence, q^{ED}_{11} is the maximum quantity ever shipped in any period (of either product). It follows that the optimal build-up-to level Q^*_{ED} for either product never exceeds q^{ED}_{11}. Hence, the four possible ranges for Q^*_{ED} are: *(i)* $q^{ED}_{11} > Q^*_{ED} \geq q^{ED}_{10}$; *(ii)* $q^{ED}_{10} > Q^*_{ED} \geq q^{ED}_{01}$; *(iii)* $q^{ED}_{01} > Q^*_{ED} \geq q^{ED}_{00}$; and *(iv)* $q^{ED}_{00} > Q^*_{ED}$. The four possible values of Q^*_{ED} posited in the Theorem correspond to these four cases; which of these cases actually arises depends on the cost-benefit tradeoff between lost sales-revenue and excessive inventory build-up. For example, suppose $q^{ED}_{01} > Q^*_{ED} > q^{ED}_{00}$. When $(X_1, X_2) = (0,0)$ (an event which has probability $P(0,0)$), the surplus above q^{ED}_{00} is stored as inventory; hence, lowering the production build-up saves holding costs and defers production costs. On the other hand, for all other realizations of (X_1, X_2), a lower build-up would have resulted in more lost sales and lower profits.

3.2 Delayed Differentiation

We next derive the optimal solution for the DD process. For notational convenience, define $q^{DD}_{x_1x_2} = \frac{\frac{f(P_{x_1x_2})+f(P_{x_2x_1})}{2}-(k+\theta-h)}{b}$. Thus, $q^{DD}_{11} = \frac{f(P_{11})-(k+\theta-h)}{b}$; $q^{DD}_{00} = \frac{f(P_{00})-(k+\theta-h)}{b}$, and $q^{DD}_{01} = q^{DD}_{10} =$

$\frac{f(\frac{1}{2})-(k+\theta-h)}{b}$. Further, define $d(x_1, x_2) = \frac{f(P_{x_1x_2})-f(P_{x_2x_1})}{2.b}$. Clearly, $d(x_1, x_2) = -d(x_2, x_1)$, and $d(x, x) = 0$ for all x.

Theorem 2 (Optimal Policies under DD) The following myopic production, shipment and inventory policies are optimal:

(A) Shipping/Inventory:

In each period, conditional on the information signals $X_1 = x_1$ and $X_2 = x_2$ and on-hand quantity Q^* of the intermediate good, the DC refines and ships the entire on-hand quantity up to the threshold $q^{DD}_{x_1x_2}$. Any surplus above this threshold is held as inventory. Of the total quantity q^S $\left(= \min\left\{Q^*, q^{DD}_{x_1x_2}\right\}\right)$ refined, the DC ships $\frac{q^S+d(x_1,x_2)}{2}$ units of product 1 and $\frac{q^S-d(x_1,x_2)}{2}$ units of product 2.

(B) Production:

Each period, the firm produces to bring the DC's intermediate good inventories for the next period to Q^*_{DD}, where $Q^*_{DD} = \max\left\{Q^*_{1,DD}, Q^*_{2,DD}, Q^*_{3,DD}\right\}$, and

$$
\begin{aligned}
Q^*_{1,DD} &= \frac{f(P_{11}) - (k+\theta-h) - \frac{\left(\frac{k}{\beta}-k+h\right)}{P(1,1)}}{b}, \\
Q^*_{2,DD} &= \frac{\frac{P(1,1)f(P_{11})+(1-2.P(1,1))f(\frac{1}{2})-\left(\frac{k}{\beta}-k+h\right)}{1-P(1,1)} - (k+\theta-h)}{b}, \text{ and} \\
Q^*_{3,DD} &= \frac{f(\frac{1}{2}) - \left(\frac{k}{\beta}+\theta\right)}{b}. \blacklozenge
\end{aligned}
$$

The optimal policies under DD have an intuitive interpretation similiar to the case of ED, but with an additional level of complexity. Since the intermediate product can be refined into *either* final product, the DC must choose among *three* options: refining (and shipping) more of product 1, of product 2, or holding the intermediate-good inventory. Consider first the optimal allocation problem between products 1 and 2. For a total shipment quantity q of both products, expected revenues are maximized when the allocation equalizes the expected marginal revenues across the two products. When the market signals are $X_1 = x_1$ and $X_2 = x_2$, this results in an allocation of $\frac{q+d(x_1,x_2)}{2}$ and $\frac{q-d(x_1,x_2)}{2}$ to products 1 and 2 respectively – the allocation difference across the two products is exactly $d(x_1, x_2)$, for all $q \geq d(x_1, x_2)$. By arguments similiar to those for ED, $q^{DD}_{x_1x_2}$ is the ship-up-to level; thus, $\min\{Q^*_{DD}, q^{DD}_{x_1x_2}\}$ is the total shipment quantity, and $\left(Q^*_{DD} - q^{DD}_{x_1x_2}\right)^+$ is held as inventory.

The production policy is a stationary threshold policy, with the threshold given by Q^*_{DD}. The three possible ranges for Q^*_{DD} are: *(i)* $q^{DD}_{11} > Q^*_{DD} \geq q^{DD}_{10}(=q^{DD}_{01})$; *(ii)* $q^{DD}_{10} > Q^*_{DD} \geq q^{DD}_{00}$; and *(iii)* $q^{DD}_{00} > Q^*_{DD}$, and the three possible values of Q^*_{DD} posited in the Theorem correspond to these ranges, and reflect a tradeoff between lost sales-revenue and excessive inventory build-up.

4 Alternative Business Processes: Operational Implications

In this Section, we compare the optimal policies and operational performance of the DD and ED processes. In particular, we compare their ship-up-to and build-up-to levels, inventories and sales. We then derive and study the expected profits under each process.

4.1 Optimal Policies

Theorem 3 (Comparison of Optimal Policies) *(i)* The ship-up-to levels under the DD and ED structures are related by $q^{DD}_{x_1x_2} = q^{ED}_{x_1x_2} + q^{ED}_{x_2x_1}$. Thus, $q^{DD}_{00} = 2 \cdot q^{ED}_{00}$, $q^{DD}_{11} = 2 \cdot q^{ED}_{11}$ and $q^{DD}_{10} = q^{DD}_{01} = q^{ED}_{10} + q^{ED}_{01}$.

(ii) The build-up-to levels under the two processes satisfy the relationship $2 \cdot Q^*_{ED} \geq Q^*_{DD}$.

$$\text{Specifically, } 2\cdot Q^*_{ED} - Q^*_{DD} = \begin{cases} \left(\frac{1-2.P(1,1)}{b\cdot P(1,1)}\right)\left[\left(\frac{k}{\beta} - k + h\right) - P(1,1)\cdot(f(P_{11}) - f(P_{10}))\right]^+, \\ \qquad \text{when } \frac{P(1,1)}{2}\left(f(P_{11}) - f(P_{00})\right) \geq \frac{k}{\beta} - k + h; \\ \left(\frac{1-2\cdot P(1,1)}{b\cdot(1-P(1,1))}\right)\left[P(1,1)\cdot f(P_{11}) + P(1,0)\cdot f(P_{10}) - \frac{f(P_{01})}{2} - \left(\frac{k}{\beta} - k + h\right)\right]^+, \\ \qquad \text{otherwise.} \ \blacklozenge \end{cases}$$

To understand the intuition behind Theorem 3, recall that the ship-up-to levels reflect a tradeoff between the expected revenues from selling an additional unit and the value of holding that unit in inventory. Under DD, although the DC can customize its inventories for either market, the thresholds for each market at which additional sales are less profitable than keeping inventory are still operative: these are $q^{ED}_{x_1x_2}$ and $q^{ED}_{x_2x_1}$ respectively. Below the total shipment level $\left(q^{ED}_{x_1x_2} + q^{ED}_{x_2x_1}\right)$, the DC will find it profitable to ship to one or both markets. Above this threshold, both markets are saturated, and inventory is the dominant option. By definition, the ship-up-to level under DD is $q_{x_1x_2}^{DD} = \left(q_{x_1x_2}^{ED} + q_{x_2x_1}^{ED}\right)$.

Since delayed product differentiation enables pooling of the intermediate good, the build-up-to level under DD (Q_{DD}^*) is no more than the total ($2 \cdot Q_{ED}^*$) of the build-up-to levels of the two products under ED, in accord with our intuition. To understand when $2 \cdot Q_{ED}^*$ is *strictly* greater than Q_{DD}^*, consider when substitutability across the two products (i.e., commonality of the intermediate good) would increase total sales and lower inventories compared to the ED optimal policy. When the information signals received are $(0,0)$ or $(1,1)$, substitutability across the two products (and hence pooling) would have no effect on ED sales under the optimal policy. Now suppose the signals are $(1,0)$ or $(0,1)$. When $Q_{ED}^* \leq q_{01}^{ED}$, the entire available quantity is shipped to the markets, and when $Q_{ED}^* > q_{10}^{ED}$, inventories of both products are carried over to the next period: substitutability would not have increased total sales in either case. However, when $Q_{ED}^* \in \left(q_{01}^{ED}, q_{10}^{ED}\right)$, there is a shortage of the higher-demand product and an excess of the other; intermediate good commonality would have boosted sales of the higher-demand product and lowered total inventory. Conversely, a lower build-up-to level is adequate under DD. In fact, it is easily demonstrated that the conditions (given by Theorem 3) for $2 \cdot Q_{ED}^* > Q_{DD}^*$ are identical to the conditions for $Q_{ED}^* \in \left(q_{01}^{ED}, q_{10}^{ED}\right)$.

4.2 Sales and Inventories

We now compare the sales and inventories under the two business processes. Clearly, the average total (2-product) inventory under ED is given by

$$\begin{aligned} I_{ED} &= 2 \cdot \Big[P(1,1) \cdot \left(Q_{ED}^* - q_{11}^{ED}\right)^+ + P(1,0) \cdot \left(Q_{ED}^* - q_{10}^{ED}\right)^+ + P(0,1) \cdot \left(Q_{ED}^* - q_{01}^{ED}\right)^+ \\ &\quad + P(0,0) \cdot \left(Q_{ED}^* - q_{00}^{ED}\right)^+\Big]. \end{aligned} \tag{6}$$

The average inventories under DD are

$$\begin{aligned} I_{DD} &= P(1,1) \cdot \left(Q_{DD}^* - q_{11}^{DD}\right)^+ + P(1,0) \cdot \left(Q_{DD}^* - q_{10}^{DD}\right)^+ + P(0,1) \cdot \left(Q_{DD}^* - q_{01}^{DD}\right)^+ \\ &\quad + P(0,0) \cdot \left(Q_{DD}^* - q_{00}^{DD}\right)^+. \end{aligned} \tag{7}$$

Since the build-up-to levels in each period are $2 \cdot Q^*_{ED}$ and Q^*_{DD} under ED and DD respectively, the expected (average) sales under each process are $S_{ED} = 2 \cdot Q^*_{ED} - I_{ED}$ and $S_{DD} = Q^*_{DD} - I_{DD}$. Thus, the difference in expected sales is

$$S_{ED} - S_{DD} = (2 \cdot Q^*_{ED} - Q^*_{DD}) - (I_{ED} - I_{DD}) . \tag{8}$$

The next Theorem compares sales and inventories under the two business processes.

Theorem 4 *(i)* The average inventory under ED, I_{ED}, is not less than the average inventory under DD, I_{DD}. Specifically, $I_{ED} - I_{DD} = 2 \cdot Q^*_{ED} - Q^*_{DD}$, i.e., the difference in inventories is equal to the difference in build-up-to levels given by Theorem 3.

(ii) The expected sales under ED and DD are identical. ♦

Theorem 4 implies that postponement cuts down inventory costs; these savings are entirely attributable to DD's lower build-up-to levels. To understand what drives the second part of the Theorem, observe that, since $Q^*_{DD} \leq 2 \cdot Q^*_{ED}$, the DC has a larger quantity available for sales under ED than under DD, in each period. We also know that $Q^*_{DD} \leq q^{DD}_{11}$ and $2 \cdot Q^*_{ED} \leq 2 \cdot q^{ED}_{11}$. Thus, when the signals received are $X_1 = X_2 = 1$, the total sales under ED are always at least as large as under DD. When $X_1 = X_2 = 0$, the ED and DD shipments are identical. Yet, the expected sales are the same due to the greater flexibility under DD when the signals are mixed ($(1, 0)$ or $(0, 1)$): the DC can choose among shipping *either* product or holding inventory, constrained only by the *total* quantity of the intermediate good available. In particular, the firm can ship a large quantity of the product facing higher expected demand at the expense of the other product. In contrast, under ED there is less opportunity for exploitation of differential shipments across the two markets, hence inventory becomes a more attractive option. It follows that sales are greater under DD when the signals are mixed, compensating for the reverse phenomenon when the signals are $(1, 1)$. While the expected sales are the same for ED and DD, the sales strategies are not, reflecting DD's greater responsiveness to demand information, which is enabled by the greater flexibility in distribution under DD.

To sharpen our intuition on the drivers of profit and the role of information, we next compare the optimal production, sales and inventory policies under ED and DD when the demand forecasts aggregate the two products.

4.3 Aggregate vs. Disaggregate Forecasts

We distinguish between two types of forecast: *aggregate* and *disaggregate*. In our model, the information generated by the IS (hence, the demand forecast) are disaggregate, i.e., they are specific to each product.

Now suppose that the firm's IS provides only an estimate of the *aggregate* demand across the two markets. Specifically, we assume that the signals $(1,0)$ and $(0,1)$ are garbled so the firm does not know which market is in the low state and which in the high state. In this case, the demand forecast can be *high* (when $(X_1, X_2) = (1,1)$, which occurs with probability $\frac{1}{4}$), *medium* (when the signals are 1 and 0, which has probability $\frac{1}{2}$) or *low* (when the signals are both 0, which has probability $\frac{1}{4}$). Thus, disaggregated information can be inferred perfectly from the aggregate when the forecast is high or low, but not when the forecast is medium. Theorem 5 provides the optimal production, shipment and inventory policies for both DD and ED, under aggregate forecasts.

Theorem 5 (Optimal Policies for Aggregate Forecasts)

(i) The optimal production, shipment and inventory policies under DD are identical to those for disaggregated forecasts, and given by Theorem 2. Only the allocation rule across products is different; in this case, the DC allocates exactly *half* of the total shipment to each product.

(ii) The optimal policies for ED are, *in effect*, identical to those for DD under aggregate forecasts, and obtained by setting $Q^*{}_{ED} = \frac{Q^*_{DD}}{2}$. ♦

An important implication of Theorem 5 is that when forecasts are made only in the aggregate, the actions taken each period under ED and DD, and hence sales, inventories, revenues and profits, are *identical*. Thus, the availability of disaggregated forecasts (and implicitly, an IS that provides information disaggregated on a product/market basis) is a *necessary* condition for a positive VOP. As will be shown later, however, the availability of disaggregated information is not sufficient for a positive VOP.

A second implication of Theorem 5 is that the total costs under DD are identical for both aggregate and disaggregate forecasts; however, forecast disaggregation may increase revenues through a better-informed allocation of the intermediate good across the two products. Further, by Theorem

4, the total production and inventory costs for ED are always greater than those for DD. ED may in fact generate greater sales revenues than DD, yet DD outperforms ED by its superior management of the marketing-operations interface. We look more closely at the profits under ED and DD in the next Section.

4.4 Profits under ED and DD

Under both ED and DD, three factors determine the net profit: sales revenues, production and refinement costs, and inventory holding costs. To interpret the profit expressions, consider first the DD process. Comparing Theorems 2 and 5, it is clear that under DD, production and inventory holding costs are the same for both aggregate and disaggregate forecasts. However, disaggregate forecasts increase revenues (and hence profits, since costs are the same as for aggregate forecasts). In fact, DD revenues may be decomposed into two components, based on the level of forecast aggregation: *(i) Revenues using aggregate forecasts:* When only aggregate forecasts are available, the firm estimates the average demand curve across the two markets, and ships half of the total quantity to each market (recall Theorem 5). The result is equivalent to selling the average of the total shipment quantity to two 'virtual' markets, in each of which the demand curve is the average of the demand curves of the two real markets; and *(ii) Additional Revenues from forecast disaggregation:* In addition to the averaging process, the firm can exploit the commonality of the intermediate good to make differential shipments of the two products based on product-specific demand forecasts. (The optimal shipment strategy attempts to equalize the expected marginal revenues from the two markets.) This component of revenue captures the additional value of disaggregated forecasts, and is the difference in revenues from following the DD policies specified in Theorems 2 and 5.

As shown in Theorem 5, ED and DD perform identically under aggregate forecasts; hence their profits are identical. Under forecast disaggregation, however, ED production, sales and inventory policies are markedly different (as given by Theorem 1). The additional value of forecast disaggregation under ED can be computed from the ED profits given by equation (9) of Theorem 6 below and our foregoing results. The following Theorem derives the expected disaggregated-forecast profits for ED and DD under their respective optimal policies.

Theorem 6 The infinite horizon discounted expected profits for the ED and DD processes are given by:

$$\Pi_{ED} = -2 \cdot k \cdot Q^*_{ED} - \frac{2 \cdot \beta}{1-\beta} \left[\sum_{\{x_1,x_2\}} P(x_1,x_2) \cdot \left(k \cdot \min\left\{Q^*_{ED}, q^{ED}_{x_1x_2}\right\} + h \cdot \left(Q^*_{ED} - \min\left\{Q^*_{ED}, q^{ED}_{x_1x_2}\right\}\right)\right) \right]$$
$$+ \frac{2 \cdot \beta}{1-\beta} \left[\sum_{\{x_1,x_2\}} P(x_1,x_2) \cdot \left(f(P_{x_1x_2}) - \theta - b \cdot \min\left\{Q^*_{ED}, q^{ED}_{x_1x_2}\right\}\right) \cdot \min\left\{Q^*_{ED}, q^{ED}_{x_1x_2}\right\} \right]. \tag{9}$$

and

$$\Pi_{DD} = -k \cdot Q^*_{DD} - \frac{\beta}{1-\beta} \left[\sum_{\{x_1,x_2\}} P(x_1,x_2) \cdot \left\{ k \cdot \min\left\{Q^*_{DD}, q^{DD}_{x_1x_2}\right\} + h \cdot \left(Q^*_{DD} - \min\left\{Q^*_{DD}, q^{DD}_{x_1x_2}\right\}\right)\right\} \right]$$
$$+ \frac{\beta}{1-\beta} \left[\sum_{\{x_1,x_2\}} \left\{ 2 \cdot P(x_1,x_2) \cdot \left(\frac{f(P_{x_1x_2}) + f(P_{x_2x_1})}{2} - \theta - b \cdot \frac{\min\left\{Q^*_{DD}, q^{DD}_{x_1x_2}\right\}}{2} \right) \cdot \frac{\min\left\{Q^*_{DD}, q^{DD}_{x_1x_2}\right\}}{2} \right\} \right]$$
$$+ \frac{\beta}{1-\beta} \left[\sum_{\{x_1,x_2\}} P(x_1,x_2) \cdot \left\{ b \cdot \frac{[d(x_1,x_2)]^2}{2} \right\} \right]. \tag{10}$$

The profit (net present value) under ED, Π_{ED}, can be interpreted as follows. Since the profit contribution under ED is separable across the two products, Π_{ED} is twice the single-product profit. Since the supply chain starts period 0 with zero inventory, the production cost in period 0 is the cost of producing the build-up-to level Q^*_{ED} for each product, hence the first term in (9). The factor $\frac{\beta}{1-\beta}$ is the familiar infinite-horizon net present value multiplier. Next, for a given signal pair (x_1, x_2), the shipment quantity of product 1 (say) under the optimal policy is $\min\left\{Q^*_{ED}, q^{ED}_{x_1x_2}\right\}$, and the quantity held as inventory is $\left(Q^*_{ED} - \min\left\{Q^*_{ED}, q^{ED}_{x_1x_2}\right\}\right)$. Under the optimal production policy, the production quantity of product 1 for the next period is $\min\left\{Q^*_{ED}, q^{ED}_{x_1x_2}\right\}$. Thus, the next term in (9) captures the production costs, $k \cdot \min\left\{Q^*_{ED}, q^{ED}_{x_1x_2}\right\}$, and the inventory holding cost, $h \cdot \left(Q^*_{ED} - \min\left\{Q^*_{ED}, q^{ED}_{x_1x_2}\right\}\right)$, weighted by the probability $P(x_1, x_2)$ of observing the signal pair(x_1, x_2). Finally, the revenues from shipping the quantity $\min\left\{Q^*_{ED}, q^{ED}_{x_1x_2}\right\}$when the demand signals are (x_1, x_2) are $\left(f(P_{x_1x_2}) - \theta - b \cdot \min\left\{Q^*_{ED}, q^{ED}_{x_1x_2}\right\}\right) \cdot \min\left\{Q^*_{ED}, q^{ED}_{x_1x_2}\right\}$; the last term in (9) thus reflects the expected revenues.

Unlike ED, the net profit under DD (expression (10)) is not separable across the products. The first two terms of Π_{DD} capture the production and holding costs, which are lower than ED's, since both the build-up-to level and the average inventory are lower. The third term in

Π_{DD} is the revenue from aggregate forecasts. To interpret this term, observe that when the signals received are (x_1, x_2), the total shipment quantity to the two markets is $\min\{Q^*_{DD}, q^{DD}_{x_1x_2}\}$; on average, each market receives $\frac{\min\{Q^*_{DD}, q^{DD}_{x_1x_2}\}}{2}$. If just this average quantity were shipped to each market, the revenues from the two markets would be $\left(f(P_{x_1x_2}) - \theta - b \cdot \frac{\min\{Q^*_{DD}, q^{DD}_{x_1x_2}\}}{2}\right) \cdot \frac{\min\{Q^*_{DD}, q^{DD}_{x_1x_2}\}}{2}$ and $\left(f(P_{x_2x_1}) - \theta - b.\frac{\min\{Q^*_{DD}, q^{DD}_{x_1x_2}\}}{2}\right).\frac{\min\{Q^*_{DD}, q^{DD}_{x_1x_2}\}}{2}$. The total revenue for the firm would then be $2 \cdot \left(\frac{f(P_{x_1x_2})+f(P_{x_2x_1})}{2} - \theta - b \cdot \frac{\min\{Q^*_{DD}, q^{DD}_{x_1x_2}\}}{2}\right).\frac{\min\{Q^*_{DD}, q^{DD}_{x_1x_2}\}}{2}$, which is twice the revenue obtained by shipping the "average" quantity $\frac{\min\{Q^*_{DD}, q^{DD}_{x_1x_2}\}}{2}$ to a market with "average" demand, given by the revenue function $R_{av}(q) = \left(\frac{f(P_{x_1x_2})+f(P_{x_2x_1})}{2} - \theta - b \cdot q\right) q$. The final term in Π_{DD}, $\sum_{\{(x_1,x_2)\}} P(x_1, x_2) \cdot \frac{b}{2} \cdot [d(x_1, x_2)]^2$, reflects the additional expected revenues from exploiting disaggregated forecasts by making differential shipments. To illustrate, suppose the market signals are x_1 and x_2, and the total quantity shipped is q. If the average quantity $\frac{q}{2}$ were shipped to each market, the total expected revenues would be $\left(f(P_{x_1x_2}) - \theta - b \cdot \frac{q}{2}\right) \cdot \frac{q}{2} + \left(f(P_{x_2x_1}) - \theta - b \cdot \frac{q}{2}\right) \cdot \frac{q}{2}$, which simplifies to $\left(\frac{f(P_{x_1x_2})+f(P_{x_2x_1})}{2} - \theta\right) q - b\frac{q^2}{2}$. Now suppose the firm allocates the quantity q optimally, namely $\frac{q+d(x_1,x_2)}{2}$ to market 1 and $\frac{q-d(x_1,x_2)}{2}$ to market 2, thereby equalizing the respective marginal revenues. Total revenues are then $\left(\frac{f(P_{x_1x_2})+f(P_{x_2x_1})}{2} - \theta\right) q - b\frac{q^2}{2} + b\frac{(d(x_1,x_2))^2}{2}$. Thus, the difference in revenues between the optimal allocation and an allocation of the same quantity to each market is exactly $b \cdot \frac{(d(x_1,x_2))^2}{2}$.

To summarize, the additional value of forecast disaggregation under DD is $\sum_{\{(x_1,x_2)\}} P(x_1, x_2) \cdot \frac{b}{2} \cdot [d(x_1, x_2)]^2$. Since VOP= $\Pi_{DD} - \Pi_{ED}$, the additional value of disaggregate forecasts under ED is $\sum_{\{(x_1,x_2)\}} P(x_1, x_2) \cdot \frac{b}{2} \cdot [d(x_1, x_2)]^2 - VOP$. Thus, DD benefits more from forecast disaggregation than ED; the difference in benefits is equal to the value of postponement.

It should also be clear from the preceding discussion that information disaggregation does not guarantee a positive VOP unless it translates operationally into differential shipments to the two markets. For example, in the extreme case where the two markets are perfectly positively correlated ($\rho = +1$), the posterior expected demand curves (given the signals) are always identical for the two markets — in the absence of differential shipments, the VOP is identically zero. Mathematically, when $\rho = +1$, $P_{10} = P_{01}$, and so $d(x_1, x_2) = 0$ for all x_1 and x_2. Hence, forecast disaggregation does not improve ED or DD profits, and the VOP is zero. More generally, postponement adds little value when the demands for the end-products are highly positively correlated ($P_{10} \approx P_{01}$), since (operationally) shipment quantities are not highly differentiated across products.

We develop an illustrative numerical example in the next Section.

4.5 Illustrative Example

We consider the demand intercepts $a_H = \$70$ (high) and $a_L = \$40$ (low) with a slope of $b = 1$ for the demand curve. Also, the unit production cost is $k = \$18$, the refinement (or customization and shipping) cost is $\theta = \$10$ and the discount factor is $\beta = 0.9$. We initially set the inventory holding cost h at \$6. We varied the information precision parameter α from 0 to 0.5 in increments of 0.1. The coefficient of correlation ρ between the two markets was varied from 0 to 1, also in increments of 0.1.

Figure 2 shows the expected profits[5] under ED as α and ρ are varied. Predictably, for a fixed ρ, the profits increase as α falls, i.e., as information precision increases. However, for a fixed α, since the optimal solution and profits for the two products are separable, one might naively expect that changing the demand correlation ρ will not affect profits. In fact, we find that the profits *increase* in ρ, particularly for intermediate levels of information precision ($\alpha = 0.1$ or 0.2) and moderate to high levels of correlation ($\rho \geq 0.4$). The reason is that the signals from both markets are used in the Bayesian demand prediction for each, so the forecast accuracy for either market improves with the demand correlation. The incremental value of the correlation in making better demand predictions is higher for moderate values of the information precision parameter α: when $\alpha = 0$, the demand curves for each product are perfectly predictable using the signals from their respective markets, and the correlation does not add any value; on the other hand, when $\alpha = 0.5$ (pure noise), the information signal from the *other* market is also useless, hence the correlation does not improve prediction. In both of these cases, the profits plotted against ρ are flat lines (see the top and bottom curves in Figure 2).

Figure 3 shows that the profits under DD also increase, *ceteris paribus*, in information precision. The effect of the demand correlation on profits under DD is less obvious, as two counteracting effects come into play. The first is the Bayesian effect observed under ED, which tends to increase profits with correlation. On the other hand, a *pooling* effect operates in the opposite direction: the value

[5] Of course, the dollar profits under ED and DD and the dollar value of the VOP depend on the quantity units (e.g. thousands or millions).

of pooling enabled by delayed differentiation is a decreasing function of the correlation between the two markets. For most values of α and ρ, the pooling effect dominates, and so the profits under DD fall with increasing ρ. However, for moderate levels of information precision ($\alpha = 0.1, 0.2$ and 0.3) and high levels of correlation ($\rho \geq 0.8$), the Bayesian effect prevails, and profits increase with ρ in this region. When $\alpha = 0.4$, the pooling effect is stronger for all ρ. When information is completely precise ($\alpha = 0$), Bayesian analysis provides no additional information, and only the pooling effect is observed. Hence, unlike under ED, the profits fall with ρ even for $\alpha = 0$. When $\alpha = 0.5$, the signals are too noisy to provide any useful information, both effects disappear and profits are flat in ρ.

Figure 4 plots the Value of Postponement (VOP) for the same parameter values. The VOP is non-negative in our setting, because delayed product differentiation always outperforms early differentiation due to the DC's greater flexibility under DD in its customization and shipment choices.[6] Clearly, DD is better suited to exploit information than ED. Indeed, Figure 4 demonstrates that the VOP increases in information precision; in the extreme case when information is so noisy ($\alpha = 0.5$) that it precludes a useful revision of forecasts, the VOP is zero. It should also be clear from the preceding discussion (Figures 2 and 3) that the VOP falls as the market correlation increases, and this is also confirmed by Figure 4. Thus, design for postponement is generally not cost-effective when the demands for the products sold are highly correlated or demand information is very noisy.

Figure 5 shows the effect of demand variability, measured by the intercept spread $a_H - a_L$, on the VOP. The average demand intercept, $\frac{a_H + a_L}{2}$, was kept constant at \$55, and the spread was varied from \$24 to \$34 (ρ was set at 0). For all levels of α, the VOP is *increasing* in demand variability, even though both ED and DD profits individually fall with variability (except for $\alpha = 0.5$, corresponding to useless information, where the VOP is identically zero). Thus, demand variability is clearly a key factor in determining the VOP: the more dynamic the demand environment, the higher the VOP (and the higher the payoff from redesigning the products and processes for postponement).

[6] In practice, this additional flexibility usually comes at a price: an additional fixed cost of redesigning the supply chain for postponement, or higher unit costs for the DD process. (Figure 7 illustrates these tradeoffs.)

Figure 6 illustrates the effect of the inventory holding cost h on the VOP.[7] We find that although both ED and DD profits fall with h, the VOP *increases* in h. For sufficiently high levels of the holding cost, holding any inventory is too costly to be viable under either process, and the VOP flattens to a constant.

Finally, Figure 7 illustrates the tradeoff between the benefits of building postponement into the firm's production process and its costs. As discussed previously, restructuring the production process of the supply chain for postponement would incur additional fixed costs.[8] Further, the DD production process might incur higher unit production costs. In the example, we set $\rho = \alpha = 0$, fix the unit production cost k at \$18 for ED, and let k vary for DD. Figure 7 is the indifference curve (which is nearly linear) showing the combinations of additional fixed costs and incremental unit production costs for which the firm would do equally well under ED or DD. Firms whose added DD costs are below this curve are better off implementing postponement; firms above this curve should opt for early differentiation. Thus, when the incremental marginal cost is 0 (i.e., $k = \$18$ under DD as well), the incremental fixed cost up to which the firm will prefer DD is 359.25, which is the value of postponement at these parameter values. As k under DD increases, the break-even incremental fixed costs between ED and DD falls. We find that when $k = \$19.47$, i.e., the marginal production cost under DD is 8.18% higher than under ED, DD and ED do identically (assuming no additional fixed costs under DD). In our example, DD is never preferred when the unit production cost increases to more than 8.18% of ED's costs.

The choice of business process should be informed by demand conditions (including demand information) as well as cost data. As illustrated by the example, the VOP increases with demand variability, end-product demand correlation, information precision, and inventory holding costs.

5 Concluding Remarks

In this paper, we modeled a supply chain's use of delayed product differentiation, or postponement, to cope with demand uncertainty. We compared the performance of two alternative business

[7]We plot the inventory holding cost as a percentage of the costs of production and distribution, $\frac{h}{k+\theta}$.

[8]For the effects of those costs in the case of Compaq, see [28, 29].

processes – early and delayed differentiation – to quantify the value of postponement. Our analysis demonstrates that both the degree of demand uncertainty and the correlation across markets or products are important determinants of the value of postponement. Postponement is closely linked to demand forecasting, whose accuracy is increasing in the level of aggregation and decreasing in the forecast lead time. Postponement exploits both of these properties by structuring the supply chain so that (i) the production process requires only an aggregate demand forecast; and (ii) the individual-product forecasts are only needed at a later stage, when the lead time is shorter. A firm's ability to forecast demand is intimately connected to the data gathered by its IS; its ability to respond depends on the nature and timing of the activities in its supply chain. Therefore, our model incorporates both material and information flows, including IS precision (accuracy), IS timing, and the activities (production and distribution) internal to the supply chain.

We show that the value of postponement largely derives from the ability to make better use of information about demand. Three important attributes of information are *(i)* its timing within the supply chain (relative to the material flows), *(ii)* its level of aggregation, and *(iii)* its precision. All three attributes play an important role in determining the value of postponement.

The role of information timing, which is implicit in the model structure, illustrates the close link between operations and information systems in determining firms' choice of business processes and in affecting their profits. If demand information is received early enough to affect the firm's production decisions, delayed product differentiation would not substantively alter the firm's operational decisions, and would probably increase the firm's costs. If demand information is received very late (after the products are shipped to their respective markets), postponement again becomes irrelevant. Postponement is only meaningful when information is received after the initial production decision and prior to customization/distribution, as in our model. The synchronization of material and information flows is thus an important enabler of postponement as a strategic lever.

Our results showed that the second attribute of information, viz., level of aggregation, is also an important factor influencing the choice of business process. Under aggregate forecasts, both early and delayed differentiation perform identically. We showed that disaggregate forecasts, stemming from product- (or market-) specific information, which in turn leads operationally to differential product shipments, is necessary for postponement to be of value to the firm. Finally, our analysis showed that the value of postponement was strongly dependent on the precision of the firm's IS.

Our model suggests a number of open avenues for future research. While our dynamic model assumes that only the DC can hold inventory, the model could be extended to the case where retail outlets also hold inventory. The optimal solution to this problem may not possess a simple structure, and may not be derivable in closed form. Nevertheless, numerical analysis of this solution may yield further insights into the costs and benefits of postponement.

Component commonality, i.e., the use of standardized, common components for multiple products, can be effective not just to pare inventory costs [17] but also as a enabler of postponement [23]. Examples include IBM Europe's PC strategy of building subassemblies of "shells" to stock and final assemblies to order [11] and Honda's flexible design platform for its best-selling Accord to cater to diverse markets [7]. Honda recently brought to market three distinct Accords at savings of hundreds of millions of dollars following this approach [7]. Future research could build on our mathematical model to shed further light on this issue.

Another important area relates to the interaction of multiple players within the supply chain, linking inventory, information and incentives issues (cf. [3, 4, 8, 24]). Our model assumes that the production facility, the DC and the retail outlets are all under the control of a single decision-making authority (or equivalently, they work together as a *team* [26]). When the different parts of a complex supply chain are controlled by different firms, as is often the case, coordinating their actions through proper contracting mechanisms becomes an important issue. The study of postponement in such a setting, relating the effects of contract structure to the value and impact of postponement, is another promising area for future research.

References

[1] Amihud, Y. and H. Mendelson, "Multiperiod Sales-Production Decisions under Uncertainty", *Journal of Economic Dynamics and Control*, 1983, v5, pp. 249-265.

[2] Anand, Krishnan S., and H. Mendelson, "Information and Organization for Horizontal Multi-market coordination", *Management Science*, December 1997, v43n12, pp. 1609-1627.

[3] Anupindi, Ravi and Y. Bassok, "Centralization of Stocks: Retailers vs. Manufacturer", Working Paper, Kellogg Graduate School of Management, Northwestern University, 1998.

[4] Anupindi, Ravi, Y. Bassok, and E. Zemel, "Inventory Pooling with Claims: A Duopoly Model", INFORMS Fall Meeting, Dallas, 1997.

[5] Aviv, Yossi, and A. Federgruen, "Design for Postponement: A Comprehensive Characterization of its Benefits Under Unknown Demand Distributions", Working Paper, Columbia Graduate School of Business, 1997.

[6] *Benetton (A)*, Harvard Business School Case, 9-685-014, 1984.

[7] *Business Week*, "Can Honda build a world car ?", September 8, 1997, v40n14, pp. 100-108.

[8] Cachon, Gerard P., "Stock Wars: Inventory Competition in a Two Echelon Supply Chain with Multiple Retailers", Working Paper, Fuqua School of Business, Duke University, 1997.

[9] Eppen, G. and L. Schrage, "Centralized Ordering Policies in a Multiwarehouse System with Leadtimes and Random Demand", in *Multi-Level Production/Inventory Control Systems: Theory and Practice* (Schwarz, L.B., Ed.), North Holland, New York, 1981, pp. 51-67.

[10] Federgruen, A. and P. Zipkin, "Approximations of Dynamic, Multilocation Production and Inventory Problems", *Management Science*, v30n1, January 1984, pp. 69-84.

[11] Feigen, G., C. An, D. Connors and I. Crawford, "Shape Up, Ship Out", *OR/MS Today - Online Edition*, April 1996, v23n2.

[12] Feitzinger, Edward and H.L.Lee, "Mass Customization at Hewlett-Packard: The Power of Postponement", *Harvard Business Review*, January-February, 1997, pp. 116-121.

[13] Fisher, M.L., J.H. Hammond, W.R. Obermeyer and A. Raman, "Making Supply Meet Demand in an Uncertain World", Harvard Business Review, May-June 1994, pp. 83-93.

[14] Gavirneni, S. and S. Tayur, "Delayed Product Differentiation and Information Sharing", Working Paper, Graduate School of Industrial Administration, Carnegie-Mellon University, 1997.

[15] Heyman, D.P. and M.J. Sobel, *Stochastic Models in Operations Research, Volume II: Stochastic Optimization*, McGraw-Hill Book Company, New York, 1984.

[16] Kirman, A.P. and M.J. Sobel, "Dynamic Oligopoly with Inventories", *Econometrica*, 1974, v42n2, pp. 279-287.

[17] Lee, Hau L., " Effective Inventory and Service Management through Product and Process Redesign", *Operations Research*, 1996, v44n1, pp. 151-159.

[18] Lee, Hau L. and C. Billington, "Managing Supply Chain Inventory : Pitfalls and Opportunities", *Sloan Management Review*, 1992, v33n3, pp. 65-73.

[19] Lee, Hau L. and C. Billington, "Material Management in Decentralized Supply Chains", *Operations Research*, 1993, v41n5, pp. 835-847.

[20] Lee, Hau L. and C. Billington, "Designing Products and Processes for Postponement", in *Management of Design: Engineering and Management Perspectives*, edited by S. Dasu and C. Eastman, 1994, Kluwer Academic Publishers, Boston, pp. 105-122.

[21] Lee, Hau L. and C. Billington, "The Evolution of Supply-Chain-Management Models and Practice at Hewlett-Packard", *Interfaces*, 1995, v25n5, pp. 42-63.

[22] Lee, Hau L., C. Billington and B. Carter, "Hewlett-Packard gains Control of Inventory and Service through Design for Localization", *Interfaces*, 1993, v23n4, pp. 1-11.

[23] Lee, Hau L. and C. S. Tang, "Modelling the Costs and Benefits of Delayed Product Differentiation", *Management Science*, 1999, v43n1, pp. 40-53.

[24] Lee, Hau L. and S. Whang, "Decentralized Multi-Echelon Supply Chains: Incentives and Information", Working paper, Stanford University, Stanford, CA, 1997.

[25] Lee, Hau L. and S. Whang, "Value of Postponement from Variability Reduction Through Uncertainty Resolution and Forecast Improvements", Working paper, Stanford University, Stanford, CA.

[26] Marschak, J. and R. Radner, *Economic Theory of Teams*, 1972, New Haven, Yale University Press.

[27] Mendelson, H., and R. Pillai, "Industry Clockspeed: Measurement and Operational Implications", *Manufacturing and Service Operations Management*, 1998, forthcoming.

[28] Ramstad, Evan, "Compaq Stumbles as PCs Weather New Blow", *The Wall Street Journal*, March 9, 1998, pp. B1.

[29] Ramstad, Evan, "Compaq Seeks to Cut Inventories with Free Monitors for Businesses", *The Wall Street Journal*, March 16, 1998, pp. B10.

[30] Schwarz, L.B., "A Model for Assessing the Value of Warehouse Risk-Pooling: Risk-Pooling over Outside-Supplier Leadtimes", *Management Science*, v35n7, July 1989, pp. 828-842.

[31] Sobel, M.J., "Myopic Solutions of Markov Decision Processes and Stochastic Games", *Operations Research*, 1981, v29, pp. 995-1009.

[32] Swaminathan, J.M. and S. Tayur, "Managing Broader Product Lines through Delayed Differentiation using Vanilla Boxes", Working Paper, Graduate School of Industrial Administration, Carnegie-Mellon University, 1996.

[33] Zuckerman, Laurence, "Compaq to Adopt Made-to-Order Sales Model", *Cybertimes*, Web Edition of *The New York Times*, May 5, 1997.

[34] Zuckerman, Laurence, "Compaq: We're No. 1, but we can change", *Cybertimes*, Web Edition of *The New York Times*, June 16, 1997.

Figure 1

Model Setup

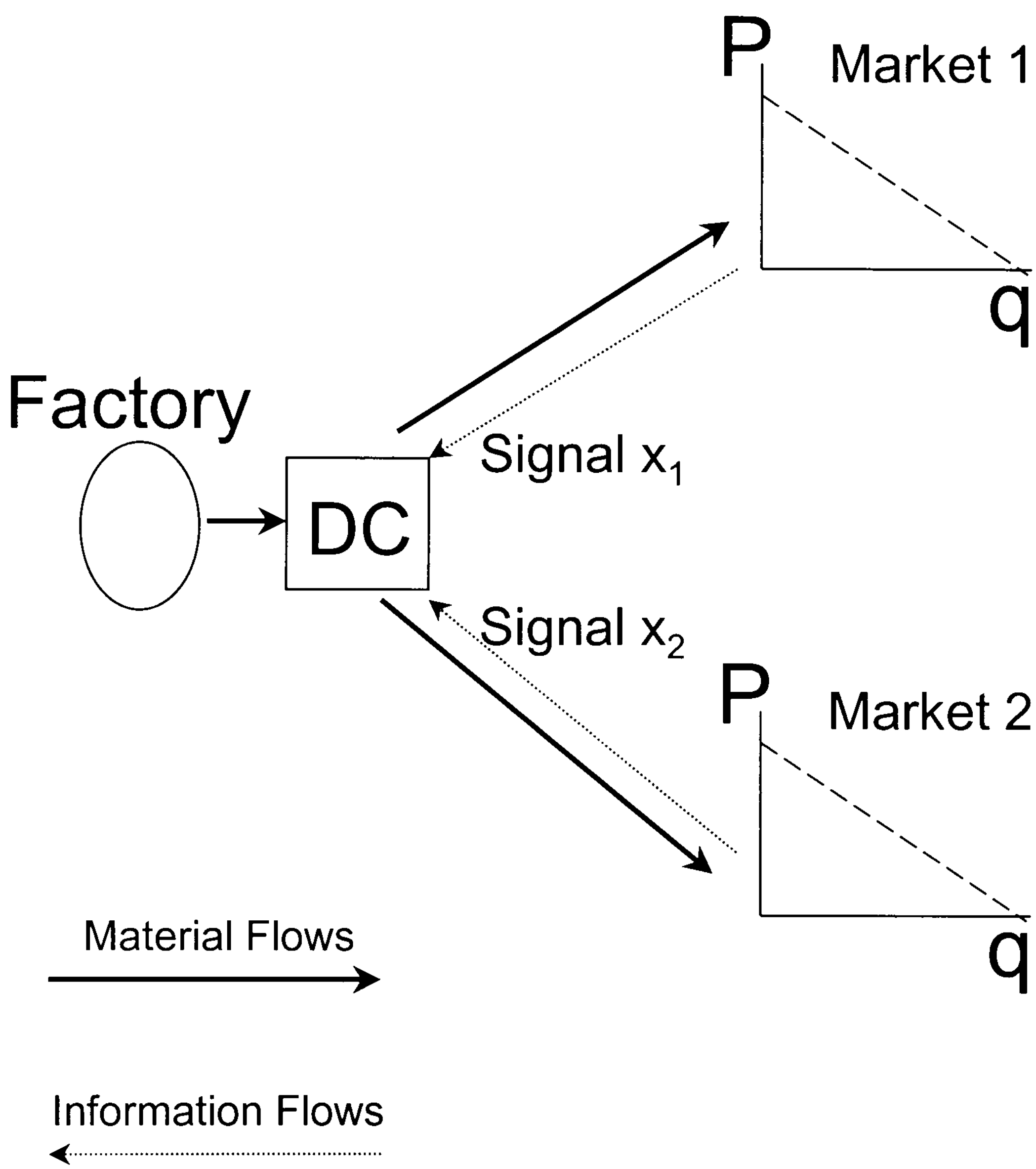


Figure 2

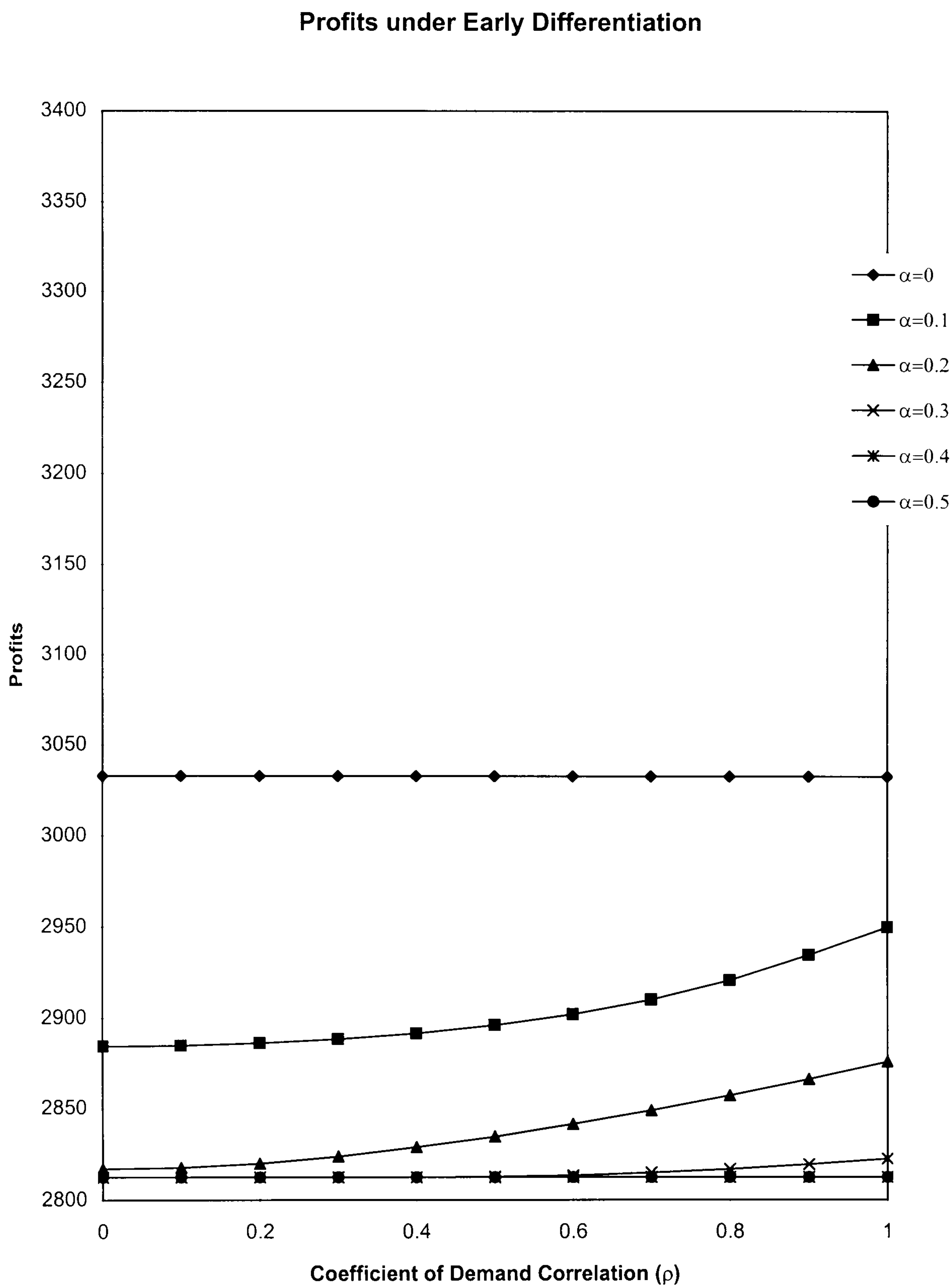

Profits under Early Differentiation
3400
3350
3300
3250
3200
3150
3100
3050
3000
2950
2900
2850
2800
Profits
0
0.2
0.4
0.6
0.8
1
Coefficient of Demand Correlation (ρ)
α=0
α=0.1
α=0.2
α=0.3
α=0.4
α=0.5

Figure 3

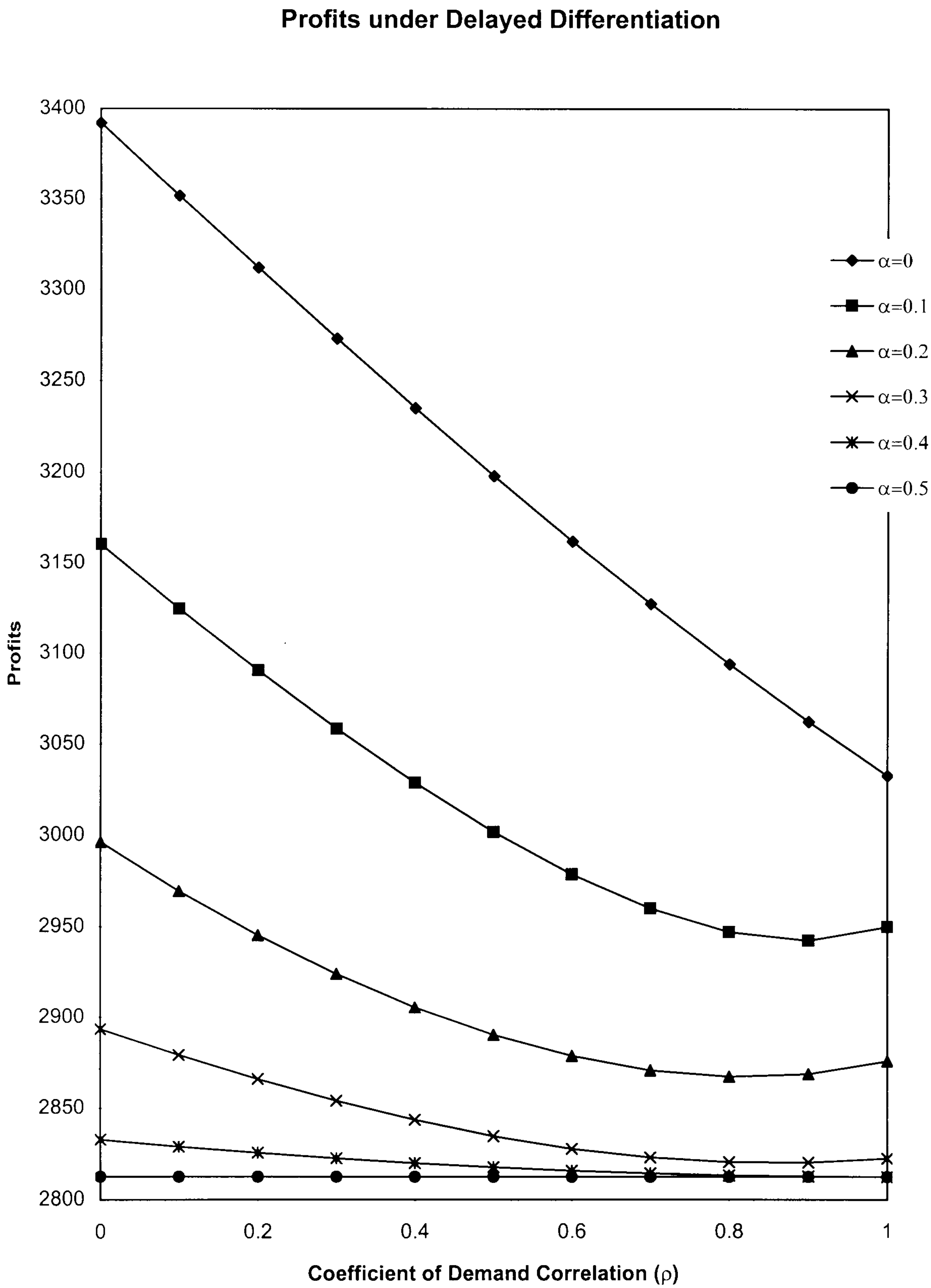

Profits under Delayed Differentiation
Profits
3400
3350
3300
3250
3200
3150
3100
3050
3000
2950
2900
2850
2800
0
0.2
0.4
0.6
0.8
1
Coefficient of Demand Correlation (ρ)
α=0
α=0.1
α=0.2
α=0.3
α=0.4
α=0.5

Figure 4

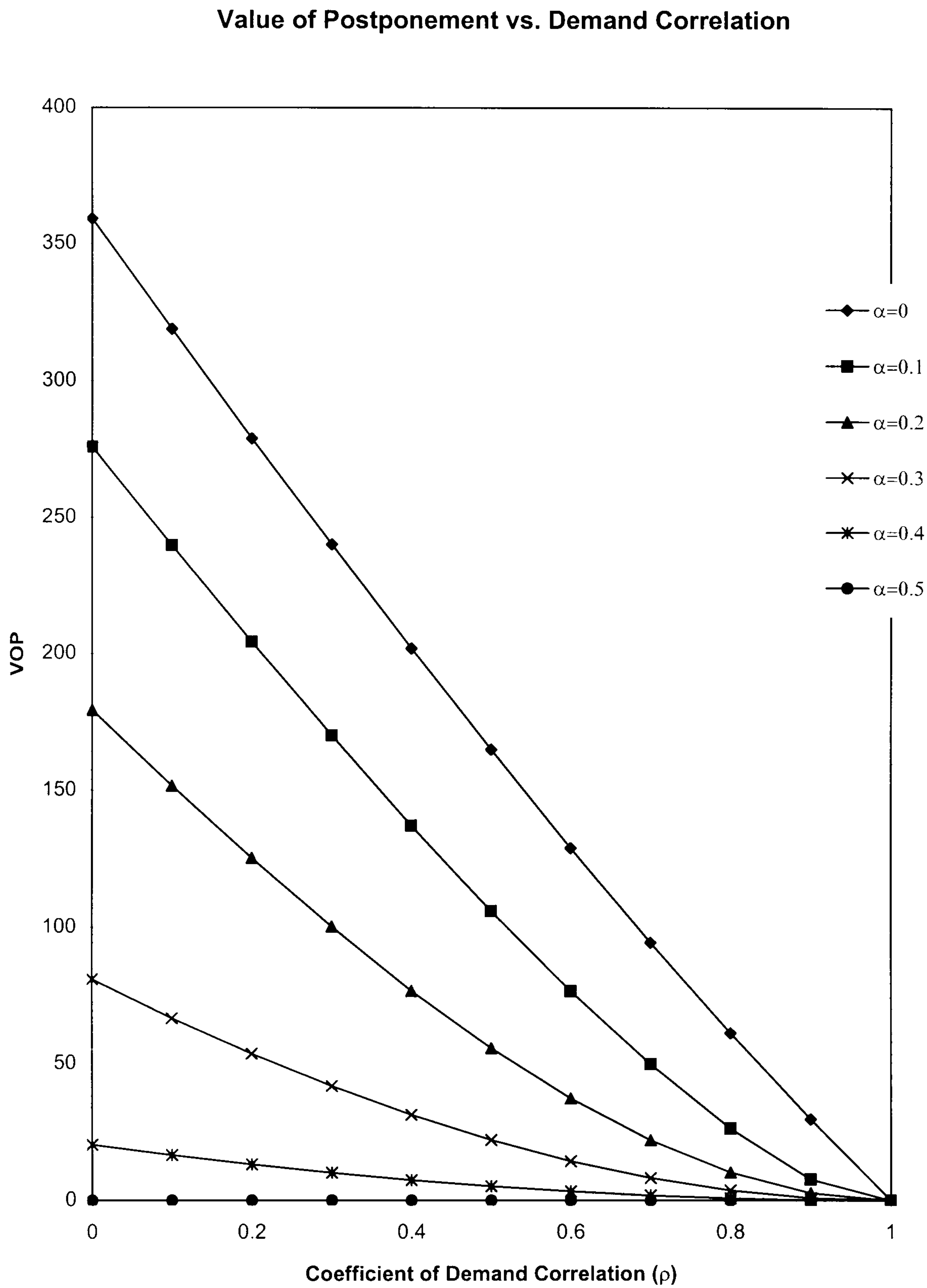

Value of Postponement vs. Demand Correlation
VOP
Coefficient of Demand Correlation (ρ)
α=0
α=0.1
α=0.2
α=0.3
α=0.4
α=0.5
400
350
300
250
200
150
100
50
0
0
0.2
0.4
0.6
0.8
1

Figure 5

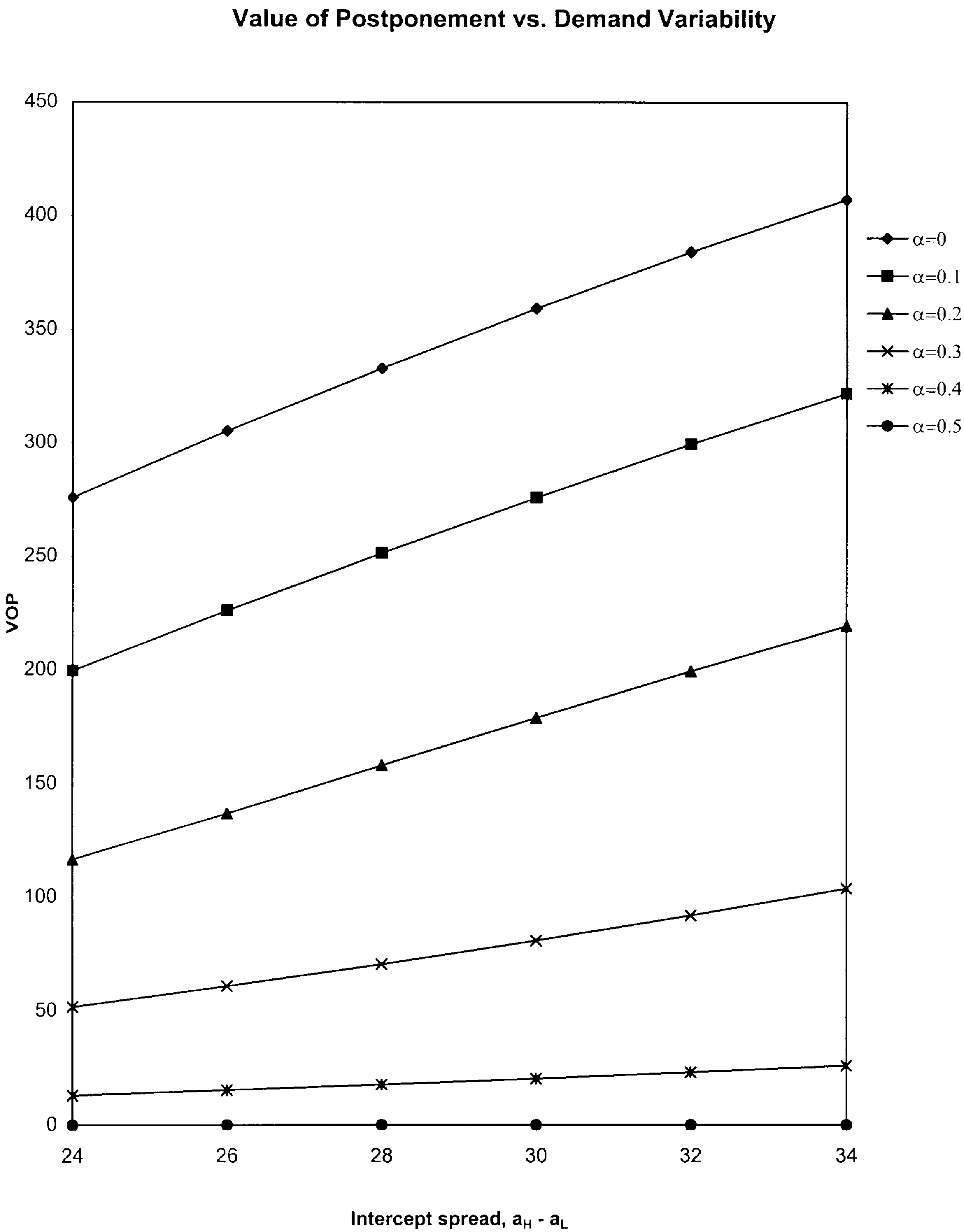

Value of Postponement vs. Demand Variability
450
400
350
300
250
200
150
100
50
0
VOP
24
26
28
30
32
34
Intercept spread, a_H - a_L
α=0
α=0.1
α=0.2
α=0.3
α=0.4
α=0.5

Figure 7

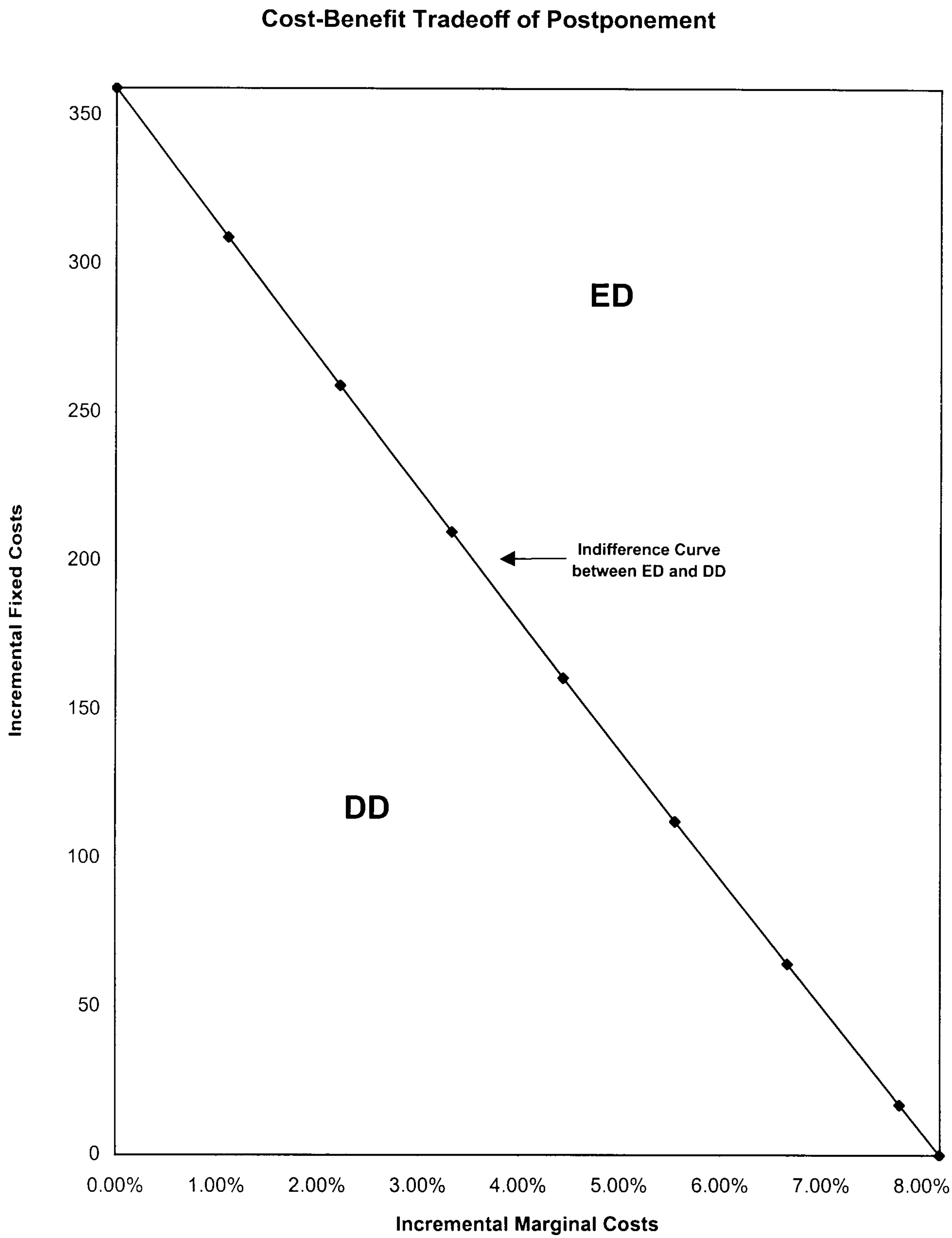
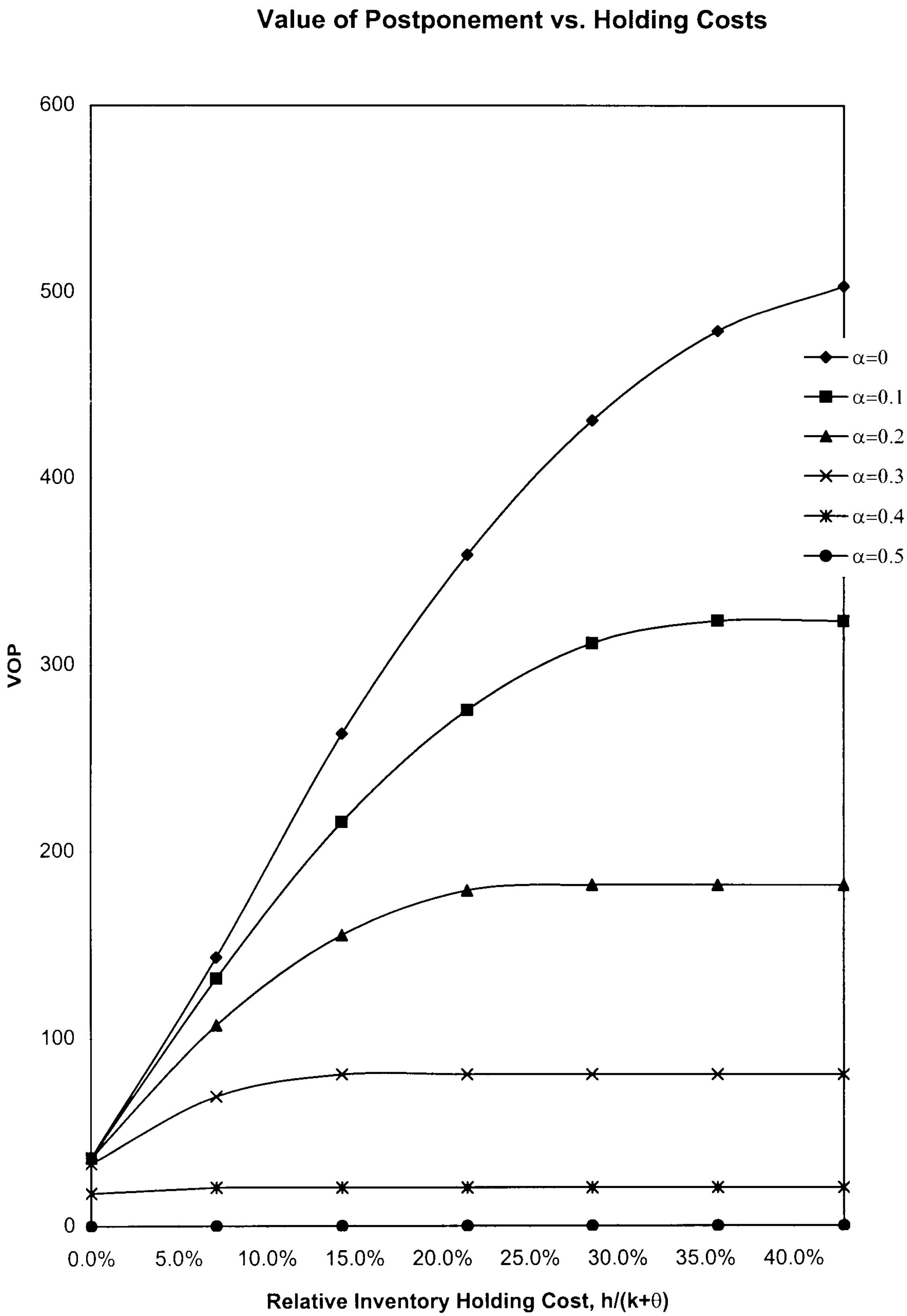

Cost-Benefit Tradeoff of Postponement
350
300
250
200
150
100
50
0
Incremental Fixed Costs
ED
DD
Indifference Curve
between ED and DD
0.00%
1.00%
2.00%
3.00%
4.00%
5.00%
6.00%
7.00%
8.00%
Incremental Marginal Costs

Figure 6



Value of Postponement vs. Holding Costs
600
500
400
300
200
100
0
VOP
0.0%
5.0%
10.0%
15.0%
20.0%
25.0%
30.0%
35.0%
40.0%
Relative Inventory Holding Cost, h/(k+θ)
α=0
α=0.1
α=0.2
α=0.3
α=0.4
α=0.5