Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Timing Successive Product Introductions with Demand Diffusion and Stochastic Technology Improvement

R. Mark Krankel, Izak Duenyas, Roman Kapuscinski,

M&SOM MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 8, No. 2, Spring 2006, pp. 119–135
ISSN 1523-4614 | EISSN 1526-5498 | 06 | 0802 | 0119
informs®
DOI 10.1287/msom.1060.0102

Timing Successive Product Introductions with Demand Diffusion and Stochastic Technology Improvement

R. Mark Krankel
Department of Industrial and Operations Engineering, University of Michigan,
Ann Arbor, Michigan 48109-2117, mark_krankel@delmia.com

Izak Duenyas, Roman Kapuscinski
Ross School of Business, University of Michigan, Ann Arbor, Michigan 48109-1234
{duenyas@umich.edu, roman.kapuscinski@umich.edu}

This paper considers a firm's decisions on the introduction timing for successive product generations. We examine the case where a firm introduces multiple generations of a durable product for which demand is characterized by a demand diffusion process. Under fixed introduction costs, we consider the case where available product technology improves stochastically. As such, delaying introduction to a later date may lead to the capture of further technology improvements, potentially at the cost of slowing sales for the existing product (and a decline in market potential for the product to be introduced, given our focus on durable products). We specify a state-based model of demand diffusion and construct a decision model to solve the firm's introduction timing problem. By incorporating technology improvement in our model, we prove the optimality of a state-dependent threshold policy governing the firm's product-introduction decisions. Numerical analysis reveals the influence of key model parameters on the pace of product introduction. Our model helps to explain the product-introduction behavior of firms and provides an alternative to previous explanations of IBM's introduction timing decisions for successive generations of its mainframe computers.

Key words: product innovation; uncertain technology; introduction to market
History: Received: June 15, 2004; accepted: September 8, 2005. This paper was with the authors 4 months for 1 revision.

1. Introduction

Consider an innovative firm that manages the development and production of a single, durable product. Over time, the firm's research and development (R&D) department generates a stochastic stream of new product technology, features, and enhancements for design into successive product generations. These technology discoveries are defined broadly to include new developments by the firm's own R&D department, technological advances in the industry, as well as announcements and discoveries that promote the feasible application of preexisting technology. The firm captures the benefits of such advances by introducing a new product generation. Due to fixed product-introduction costs, it may be unreasonable to immediately release a new product generation after each technology discovery. Rather, the firm may prefer to delay an introduction until sufficient incremental new product technology has accumulated in R&D.

An innovative firm's decisions on introduction timing for successive product generations can dramatically affect its future profits (see Lilien and Yoon 1990). The objective of this paper is to characterize the firm's optimal product-introduction policy. Under a monopolistic setting with exogenous technology improvement process, we investigate the case in which the firm may introduce multiple product generations over an infinite planning horizon. The total number of product generations is not pre-specified; rather, it is determined by the pace of technology improvement along with the firm's dynamic decisions on when to introduce.

Analysis is centered upon two key influences affecting the introduction timing decisions: (1) demand diffusion dynamics, where future product demand is a function of past sales, and (2) technology improvement process, specifically the concept that delaying introduction to a later date may lead to the capture of further improvements in product technology. Previous literature examining incremental technology introduction has focused on either (1) or (2), but none have considered both factors simultaneously. As a result, the present analysis provides new insight into the structure of the optimal introduction timing policy for an innovative firm. Using a proposed decision model that incorporates both key influences, we prove the optimality of a threshold policy: it is optimal for the firm to introduce the next product generation when the technology of the current generation is below a state-dependent threshold, in which the state is defined by the firm's cumulative sales and the technology level in R&D.

Our analysis is related to but differs from those of Wilson and Norton (1989) and Mahajan and Muller (1996). These two papers proceed under a demand diffusion framework, but do not model the progression of product technology. Rather, they assume that the next generation product to be introduced is available at all times starting from Time 0. As a result, they respectively conclude the optimality of "now or never" (the new generation product is introduced immediately or never) and "now or at maturity" (the new generation product is introduced immediately or when the present generation product has reached sufficient sales) rules governing product introductions. By allowing for technology improvement over time, we expand the analysis to reveal how the optimal policy dynamically depends on the system state as characterized by cumulative sales and technology levels.

A firm deciding when to introduce the next product generation faces a trade-off between waiting for further technology gains or capturing the benefits of technology improvements sooner. Our simple models provide a method to capture this trade-off and gain insights not previously offered in the literature. For example, analyzing IBM's introduction timing for new generations of its mainframe computers, Mahajan and Muller (1996) conclude that IBM introduced two of its mainframe generations too late. They reach this conclusion using their model that assumes, implicitly, that the product technology necessary for marketing the new generations at a competitive price was available to IBM at the earlier introduction times they suggest. Our model demonstrates that it is possible to reach a different conclusion by taking into account the stochastic progression of technology.

The remainder of this paper is organized as follows: §2 presents a review of relevant literature. The model formulation and analytical results on policy structure are presented in §§3 and 4, respectively. Section 5 details the results from a computational study and then examines IBM's product-introduction time decisions for successive generations of its mainframe computers. Section 6 concludes with a summary of key insights and recommendations for future research.

2. Literature

Two main research areas are directly relevant to the current work. The first centers on models of demand. Papers in this area describe the patterns of demand exhibited by single or multiple product generations, specifically in relation to new innovations. These papers concentrate on system dynamics and/or model fit with empirical data. The second research area examines decision models for technology adoption timing. A subset of this group includes papers that model the introduction of new products subject to demand diffusion.

Research focusing on the development of demand models to describe new product adoption over time includes a large body of literature. Bass (1969) initiates the stream that examines demand diffusion models by formulating a model for a single (innovative) product. The Bass model specifies a potential adopter population of fixed size and identifies two types of consumers within that population: innovators and imitators. Innovators act independently, whereas the rate of adoption due to imitators depends on the number of those who have already adopted. The resulting differential equation for sales rate as a function of time describes the empirically observed s-shaped pattern of cumulative sales: exponential growth to a peak followed by exponential decay.

Other related demand diffusion models are presented by Norton and Bass (1987), Wilson and Norton (1989), and Jun and Park (1999). Norton and Bass (1987) extend the original Bass model by incorporating substitution effects to describe the growth and decline of sales for successive generations of a frequently purchased product. Jun and Park (1999) examine multiple-generation demand diffusion characteristics by combining diffusion theory with elements of choice theory. Wilson and Norton (1989) propose a multiple-generation demand diffusion model based on information flow. In a recent work, Kumar and Swaminathan (2003) modify the Bass model for the case in which a firm's capacity constraints may limit the firm's ability to meet all demand. Using their revised demand diffusion model, they determine conditions under which a capacitated firm's optimal production/sales plan is a "build-up policy," in which the firm builds up an initial inventory level before the start of product sales and all demand is met thereafter.

A model of particular relevance to our work is that of Mahajan and Muller (1996). They introduce a diffusion model framework for a durable good with multiple product generations. We discuss their framework in more detail in §3. Mahajan et al. (1990) contains an excellent summary evaluation of the various models and progress in the first area of research.

The present work is related to a second research area that examines the adoption timing decisions of firms. Gjerde et al. (2002) model a firm's decisions on the level of innovation to incorporate into successive product generations. The article examines the conditions under which a firm should innovate to the technology frontier as opposed to pursuing a policy of incremental innovation. The paper does not consider the diffusion dynamics of the existing products in the market (product sales rates do not depend on cumulative sales). In our paper, when the firm introduces a new product, it will innovate to the technology frontier available in R&D. We are interested in whether the firm will introduce the product available in R&D and how that decision is affected by the diffusion rate of the existing product and market potential of the new product, which differentiates this paper from Gjerde et al. (2002).

Our paper is thus also related to Cohen et al. (1996), which considers a firm deciding when to introduce a new product to market. Cohen et al. assume that delaying the product-introduction time can lead to a more improved product, hence resulting in a higher sales rate. They assume that this product can only be sold during a fixed window of time. Therefore, delaying the product introduction for further development will lead to a better product and higher revenues but over a shorter time. Cohen et al. further assume that the product currently in the market or the newly introduced product both have sales at a constant rate (though the constant is different for the two generations). Thus, they do not consider the diffusion dynamics that we take into account here. They also do not consider the stochastic nature of the R&D process, and we do not assume a set fixed window of time in which the new product has to be introduced, which differentiates our model from theirs.

Balcer and Lippman (1984) study the optimal timing of new process technology adoption by a firm facing an exogenous stochastic innovation process. They conclude that a firm will adopt the current best technology if its lag in process technology exceeds a certain threshold. The threshold is either nonincreasing or nondecreasing in time, dependent on expectations with respect to potential for technology discovery. Farzin et al. (1998) considers a similar problem under a dynamic programming framework. The paper explicitly addresses the option value of delaying adoption and compares results to those using traditional net present value methods, in which technology adoption takes place if the resulting discounted net cash flows are positive. Jensen (1982) and Chatterjee and Eliashberg (1990) examine the determinants of adoption at the individual level through a decision analytic framework and then expand the dynamics to explain the observed aggregate demand diffusion patterns. In each of these works, the technology adoption decision does not explicitly consider the effects of adoption timing on product-demand dynamics. Our model differs in that future product demand is influenced by both the timing and technology level of new introductions. As such, our model is especially related to the next subset of papers that explicitly consider demand diffusion in the decision process.

Wilson and Norton (1989) consider the one-time introduction decision for a new product generation. In their model, product introduction has fixed positive effects on market potential along with negative effects due to cannibalization. Under the assumption that the new generation has a lower profit margin than the current one, they conclude that the optimal policy for the firm is given by a "now or never" rule. That is, it will either be optimal to introduce the improved product as soon as it is available or never at all. Mahajan and Muller (1996) extend the work of Wilson and Norton to develop a model that allows them to drop the assumption on decreased profit margins. In their analysis, Mahajan and Muller consider discounting of future profits and subsequently arrive at a "now or at maturity" rule governing the introduction time decision, i.e., they conclude that it will be optimal to either introduce the improved product as soon as it is available or when enough sales have been accumulated for the previous product generation.

A major difference between our work and the previous two lies in consideration of the dynamic nature of technological improvement. Both Wilson and Norton (1989) and Mahajan and Muller (1996) implicitly assume that the next product generation is available and remains unchanged regardless of when it is introduced. In contrast, we assume that a firm that delays introduction of the next product generation expects to capture greater technological advances at a later date. This translates to the expectation of a higher market potential (or sales rate curve) associated with a later introduction time. Additionally, we recognize that the technology improvement process is inherently stochastic and incorporate this into the model.

3. Model

Under a discrete-time, infinite-horizon scenario, consider a single base product that progresses through a series of product generations over time. The firm's R&D department works continuously to develop and assimilate new product technology. The benefits of improved technology are realized only through introduction of a new product generation that incorporates the latest technology available in R&D. An improvement in the incumbent product technology leads to a higher sales potential for the new product generation. However, each new generation requires a fixed introduction cost. The firm seeks an introduction policy that maximizes net profits. In each period, the firm has the option to either introduce the latest technology or continue selling at the current incumbent technology level (wait). We model the level of technology in R&D using a single index, and assume that this level improves stochastically during each period. Our objective is to characterize the firm's optimal introduction policy given this stochastic R&D process.

3.1. Notation and Assumptions

We begin with the following definitions under a dynamic-programming framework:

π = Gross unit profit margin, constant across product generations, $\pi < \infty$

K = Fixed cost of introducing a new product generation, $0 \leq K < \infty$

δ = One-period discount factor, $0 \leq \delta < 1$

z_r = Technology level available in R&D, $z_r \in \{0, 1, 2, \ldots\}$

z_m = Current technology level of incumbent product, $z_m \in \{0, 1, 2, \ldots\}$ (initial technology level is 0), $z_m \leq z_r$

ξ_i = Technology gain in period $i \in \{0, 1, 2, \ldots\}$; ξ_i i.i.d., random variables with distribution F on the sample space of nonnegative integers

$N(z_m)$ = Market potential (max cumulative sales) as a function of incumbent technology level

$\mathcal{S}(z_m) = \{0, 1, \ldots, N(z_m)\}$, the state space for cumulative sales given incumbent technology level z_m

s = Cumulative sales during all previous periods, $s \in \mathcal{S}(z_m)$

$g(s, z_m)$ = Number of units sold in the current period given that cumulative sales (from proceeding periods) is s and incumbent technology level is z_m

$V(s, z_m, z_r)$ = Optimal profit-to-go from the state (s, z_m, z_r)

$I(s, z_m, z_r)$ = Profit-to-go given the firm introduces a new product generation at the state (s, z_m, z_r) and follows optimal policy afterward

$W(s, z_m, z_r)$ = Profit-to-go given the firm delays introduction decision (waits) at the state (s, z_m, z_r) and follows optimal policy afterward

As in Cohen et al. (1996), we consider a durable base product for which product technology is additive and introduction of a new product generation results in complete obsolescence of the previous generation; i.e., once a new generation is introduced, sales of the previous generation immediately drop to and remain at zero. This property is referred to later as the "complete replacement" condition.

It is assumed that (1) available product technology improves in each period according to a stochastic process, and (2) sales for any given generation follow a demand diffusion process, detailed in the next subsection.

The firm incurs a fixed cost K upon introduction of a new product generation. This cost is constant regardless of the introduction time or technology level of the product generation introduced. It is also assumed that the firm discounts future cash flows at a constant discount rate δ.

Both the technology level and the price of a new product are expected to influence the product's market potential and associated demand diffusion dynamics (e.g., Robinson and Lakhani 1975, Kalish 1985, Krishnan et al. 1999). To understand the effects of progressing technology independent of other compounding factors, we assume a very specific but realistic pricing strategy that maintains constant unit profit margins. As mentioned above, sales potential is assumed to be an increasing function of product technology level. Moreover, we do not model capacity constraints and assume that all demand can be met so that sales equals demand (see Kumar and Swaminathan 2003 for a detailed consideration of demand diffusion dynamics under finite capacity).

3.2. Model Formulation

The product technology level available in R&D at period i is given by the technology improvement process $\{z_r(i),\ i \geq 0\}$. The improvement process takes on only positive integer values and is modeled as a stochastic function of the period i. Given $z_r(i) = z_r$, the technology available in the next period is obtained through $z_r(i+1) = z_r + \xi_i$, where the ξ_i are i.i.d. with distribution F and expectation $E\xi < \infty$.

For each level of incumbent product technology z_m, the sales rate curve $g(s, z_m)$ is taken as a pre-specified function of cumulative sales s. Hence, the complete sales model is described by the set of curves $\{g(s, z_m)\colon z_m \geq 0\}$. Letting Δ_x denote the first difference operator with respect to the variable x, the following assumption is made on the sales rate curves:

Assumption 1. *For all $z_m \geq 0$ and $s \in \mathcal{S}(z_m)$,*
(i) $\Delta_{z_m} g(s, z_m) \geq 0$,
(ii) $N(z_m) < \infty$ *and* $g(s, z_m) \leq N(z_m) - s$,
(iii) $(\partial/\partial s) g(s, z_m) \geq -1$.

Assumption 1(i) ensures that, all else equal, product sales rate is nondecreasing in product technology. Part (ii) accommodates realistic durable-good market scenarios in which the potential market size is finite and current period sales do not exceed total remaining market potential. Together with finite unit profit margins, this condition satisfies the technical requirement for bounded rewards in the analysis that follows. Condition (iii) of Assumption 1 limits the rate at which sales decrease and in a discrete-time framework guarantees that the sales rate from one period to the next does not decrease at a faster pace than sales accumulated within the period. It is intuitive that having sold one more unit in a previous period should not decrease the sales in the current period by more than one, or equivalently, selling one additional unit in an earlier period does not decrease the total sales in all the previous and current periods, combined. In the simple case where $g(s, z_m)$ is a linear decreasing function of cumulative sales s, a violation of (iii) implies that current period sales at every point s along the curve would exceed the remaining market potential, $N(z) - s$, and thus be infeasible. Clearly, (iii) is a reasonable assumption on the sales rate curves in our multiperiod sales framework. The three conditions of Assumption 1 accommodate a wide set of sales curves, including the innovation diffusion instance as demonstrated below.

The basis upon which to formulate the decision model as a dynamic program is now in place. At the state (s, z_m, z_r), the profit-to-go given the decision to introduce, $I(s, z_m, z_r)$, and the profit-to-go given the decision to wait, $W(s, z_m, z_r)$, are given by

$$I(s, z_m, z_r) = -K + \pi g(s, z_r) + \delta E V(s + g(s, z_r), z_r, z_r + \xi) \tag{1}$$

$$W(s, z_m, z_r) = \pi g(s, z_m) + \delta E V(s + g(s, z_m), z_m, z_r + \xi), \tag{2}$$

where the expectation is taken with respect to ξ. The optimum introduction policy is computed from the optimality equation:[1]

$$V(s, z_m, z_r) = \max\{I(s, z_m, z_r), W(s, z_m, z_r)\}. \tag{3}$$

[1] Unique existence of the optimal value function in the discounted Markov decision problem (1)–(3) is verified as follows (see e.g., Puterman 1994, Theorem 6.2.5). The state space is countable, the

In (3), the decision maker faces two choices at each state (s, z_m, z_r). She may either incur the fixed introduction cost and move to the higher sales curve by introducing a new generation that incorporates all available technology, or she can delay introduction and continue selling at the lower sales curve for another period. Delaying introduction allows the possibility of capturing more product technology at a later period for the same fixed cost. In addition, the introduction time for a new generation will determine the point along the new sales curve at which sales begin. Section 4 examines the optimal policy governing the decision maker's choices.

3.3. Relationship to Demand Diffusion

Let us now describe the relationship between the sales model $\{g(s, z_m): z_m \geq 0\}$ and standard demand diffusion models. Rather than restricting our analysis to a specific diffusion model, the sales model is designed for more generality within a discrete-time framework. It allows the current-period sales rate to depend on the cumulative sales level as well as the technology level of the current product generation. For the scenario considered in this paper, there is a natural link between this sales model and that of a typical (continuous-time) diffusion model. Consider the Bass diffusion model for a single innovative product:

$$\frac{dx}{dt} = \left(a + b\frac{x(t)}{N}\right)(N - x(t)). \tag{4}$$

Here, $x(t)$ is the cumulative product sales at time t, $N \geq 0$ is the total market potential for the product, and $a \geq 0$, $b \geq 0$ are the coefficients of innovation and imitation, respectively. Mahajan and Muller (1996) present an extension of the Bass model for the case of multiple product generations. Under the complete replacement scenario with a durable base product, their model can be simplified to express the relationship between two successive product generations as follows. Let $x_i(t)$ denote the cumulative sales of the ith generation by time t, $x(t)$ represent the total cumulative sales of all generations by time t, N_i represent the market potential through the ith generation, and t_i denote the introduction time of the ith generation. We then have

$$\frac{dx_i}{dt} = \begin{cases} \left(a + b\dfrac{x(t)}{N_i}\right)(N_i - x(t)) & \text{for } t_i \leq t < t_{i+1} \\ 0 & \text{for } t \geq t_{i+1}, \end{cases}$$

$$\frac{dx_{i+1}}{dt} = \begin{cases} 0 & \text{for } t < t_{i+1} \\ \left(a + b\dfrac{x(t)}{N_{i+1}}\right)(N_{i+1} - x(t)) & \text{for } t_{i+1} \leq t < t_{i+2}. \end{cases} \tag{5}$$

In (5), any two successive product generations separately follow a sales pattern of the form (4). Interaction between product generations is one-dimensional; the sales dynamics of successive generations i and $i+1$ are linked only through the cumulative sales $x(t_{i+1})$ at the time when generation $i+1$ is introduced.

For the case presented in this paper, a demand diffusion instance of the sales model $\{g(s, z_m): z_m \geq 0\}$ is specified by defining the curves $g(s, z_m)$ as follows:

$$g(s, z_m) = \left(a + b\frac{s}{N(z_m)}\right)(N(z_m) - s), \tag{6}$$

where a and b are coefficients of innovation and imitation, respectively. Because cumulative sales is tracked as a state variable, the decision model (1)–(3) clearly captures the interaction between product generations when sales curves are of the demand diffusion form (6). Moreover, an examination of (6) shows that the demand diffusion form satisfies Assumption 1 subject to a mild restriction on problem parameters; the sales curve $g(s, z_m)$ given by (6) satisfies Assumption 1(i) for all cases in which the market potential $N(z_m)$ is increasing in the product technology level z_m. Part (ii) of the assumption is satisfied by the definition of market potential in a demand diffusion setting. Because $g(s, z_m)$ is concave in s for (6), $(\partial/\partial s)g(s, z_m)$ is bounded below by its value at $s = N(z_m)$, and hence part (iii) of the assumption is satisfied for all cases with $a + b \leq 1$, a reasonable condition on the diffusion coefficients.

Thus, the decision model (1)–(3) provides a concise representation of the multiple product introduction problem faced by an innovative firm, and, in our setting, the model is consistent with and generalizes the models used in the demand diffusion literature. The

rewards are bounded, and $0 \leq \delta < 1$, where bounded rewards follows directly from Assumption 1(ii) and the condition $\pi < \infty$. Note that existence of the optimal value functions is implicit throughout the analysis of policy structure in §4.

next section examines the properties of an optimal introduction policy.

4. Optimal Policy

We set out to characterize the structural properties of an optimal introduction policy for successive product generations. We begin by proving the existence of unique state-dependent thresholds $Z_m^*(s, z_r)$ such that it is optimal for the firm to introduce the next product generation whenever the technology level of the incumbent generation is below $Z_m^*(s, z_r)$. An alternative specification of the optimal policy considers thresholds $Z_r^*(s, z_m)$ in the available R&D technology. If the thresholds $Z_m^*(s, z_r)$ are monotonic in z_r, the two specifications have a one-to-one relationship and are equivalent.

Proof of the main theorem, which verifies the optimality of a threshold policy, is based on two results that describe monotonicity of the cumulative sales and their effect on profits. Define $\{s_i^j\}_{i=i_0}^{\infty}$ to be the sequence of cumulative sales starting from some period i_0 for a system j with incumbent technology fixed at level z_m and cumulative sales $s_{i_0}^j$ at period i_0, i.e., $s_{i+1}^j = s_i^j + g(s_i^j, z_m)$ for $i \geq i_0$. Consider two similar firms selling at the same fixed technology level z_m for all periods. The firms only differ in their cumulative sales level s at some period i_0. The first result states that as the two systems progress over time, the cumulative sales level of the firm with lower initial cumulative sales will never surpass the firm with higher initial cumulative sales.

Proposition 1. *Take fixed technology level $z_m \geq 0$ and cumulative sales levels $s_{i_0}^1, s_{i_0}^2 \in \mathcal{S}(z_m)$ with $s_{i_0}^2 \geq s_{i_0}^1$. Then $s_i^2 \geq s_i^1$ $\forall\, i \geq i_0$.*

Proof. Proceed by induction. We have that $s_{i_0}^2 \geq s_{i_0}^1$. Assume $s_i^2 \geq s_i^1$ holds for some period $i \geq i_0$. Then from Assumption 1(iii), $s_{i+1}^2 = s_i^2 + g(s_i^2, z_m) \geq s_i^1 + g(s_i^1, z_m) = s_{i+1}^1$. Thus, by induction the claim holds for all $i \geq i_0$. □

The next proposition relates the optimal profit-to-go functions for two firms with different cumulative sales levels. The result states that all else equal, the discounted optimal profit-to-go for a firm with lower cumulative sales will not exceed that of a firm with higher cumulative sales by more than the net value of their cumulative sales difference. That is, future benefits cannot make up for the current sales deficit.

Proposition 2. *Take $z_r \geq z_m \geq 0$, and $s_{i_0}^1, s_{i_0}^2 \in \mathcal{S}(z_m)$ with $s_{i_0}^2 \geq s_{i_0}^1$. Then, $V(s_{i_0}^1, z_m, z_r) - V(s_{i_0}^2, z_m, z_r) \leq \pi(s_{i_0}^2 - s_{i_0}^1)$.*

Proof. At some period i_0, consider System 1 in state $(s_{i_0}^1, z_m, z_r)$ and System 2 in state $(s_{i_0}^2, z_m, z_r)$ where $s_{i_0}^2 \geq s_{i_0}^1$. Let $\tilde{\pi}$ denote the optimal (introduction) policy for System 1, and let System 2 introduce new technology to replicate System 1's technology introductions under policy $\tilde{\pi}$. Let $\tilde{z}_i$ denote System 1's optimal incumbent technology level in period $i \geq i_0$ under policy $\tilde{\pi}$, and let s_i^1, s_i^2 represent the cumulative sales up to period i of Systems 1 and 2, respectively. Because both systems follow the same introduction policy, they both have the same introduction costs-to-go, $f(K)$. Let $E_{\tilde{\pi}}[\cdot]$ denote the expectation under policy $\tilde{\pi}$, and let $V_{\tilde{\pi}}(s_{i_0}^1, z_m, z_r)$ and $V_{\tilde{\pi}}(s_{i_0}^2, z_m, z_r)$ denote the period i_0 profit-to-go under policy $\tilde{\pi}$ for Systems 1 and 2, respectively. Then,

$$
\begin{aligned}
V_{\tilde{\pi}}(s_{i_0}^1, z_m, z_r) &= -f(K) + \pi E_{\tilde{\pi}} \sum_{i=i_0}^{\infty} \delta^{i-i_0} g(s_i^1, \tilde{z}_i) \\
&= -f(K) + \pi \lim_{n\to\infty} E_{\tilde{\pi}} \sum_{i=0}^{n-i_0} \delta^i (s_{i_0+i+1}^1 - s_{i_0+i}^1) \\
&= -f(K) + \pi \lim_{n\to\infty} E_{\tilde{\pi}} \Bigg[\delta^{n-i_0} s_{n+1}^1 - s_{i_0}^1 \\
&\qquad + \sum_{i=1}^{n-i_0} (\delta^{i-1} - \delta^i) s_{i_0+i}^1 \Bigg] \\
&\leq -f(K) + \pi \lim_{n\to\infty} E_{\tilde{\pi}} \Bigg[\delta^{n-i_0} s_{n+1}^2 - s_{i_0}^2 \\
&\qquad + \sum_{i=1}^{n-i_0} (\delta^{i-1} - \delta^i) s_{i_0+i}^2 \Bigg] + \pi(s_{i_0}^2 - s_{i_0}^1) \\
&= V_{\tilde{\pi}}(s_{i_0}^2, z_m, z_r) + \pi(s_{i_0}^2 - s_{i_0}^1),
\end{aligned}
$$

where the inequality follows from Proposition 1, since the two systems start at the same technology level and maintain equivalent technology introductions. Because $\tilde{\pi}$ was optimal for System 1 and possibly suboptimal for System 2, the inequality holds if $V_{\tilde{\pi}}(s_{i_0}^1, z_m, z_r)$ and $V_{\tilde{\pi}}(s_{i_0}^2, z_m, z_r)$ are replaced by $V(s_{i_0}^1, z_m, z_r)$ and $V(s_{i_0}^2, z_m, z_r)$, respectively. □

Next, we show optimality of a threshold policy governing introduction decisions. The threshold result is based on proof that $\Delta_{z_m}[W(s, z_m, z_r) - I(s, z_m, z_r)] \geq 0$ for all $(s, z_r) \geq (0, 0)$. Monotonicity of the difference $W(s, z_m, z_r) - I(s, z_m, z_r)$ in z_m ensures that if delaying introduction is optimal at some state $(\tilde{s}, \tilde{z}_m, \tilde{z}_r)$ (i.e, $W(\tilde{s}, \tilde{z}_m, \tilde{z}_r) - I(\tilde{s}, \tilde{z}_m, \tilde{z}_r) \geq 0$) then, all else equal, delaying introduction would be optimal $(W(\tilde{s}, z_m, \tilde{z}_r) - I(\tilde{s}, z_m, \tilde{z}_r) \geq 0)$ at any $z_m \geq \tilde{z}_m$. This implies that a threshold policy is optimal.

Theorem 1. *For each* $(s, z_r) \geq (0, 0)$ *there exists a unique threshold* $Z_m^*(s, z_r) \leq z_r$ *such that the decision to introduce the next product generation is optimal if and only if incumbent technology level in the market* $z_m < Z_m^*(s, z_r)$.

Proof. It is sufficient to show that $\forall (s, z_r) \geq (0, 0)$, $\Delta_{z_m}[W(s, z_m, z_r) - I(s, z_m, z_r)] \geq 0$. Note that the value $I(s, z_m, z_r)$ does not depend on z_m. Hence, showing $\Delta_{z_m} W(s, z_m, z_r) \geq 0$ is sufficient. Consider the value iteration index k. Based on (1)–(3), we have

$$\begin{aligned} I^k(s, z_m, z_r) &= -K + \pi g(s, z_r) \\ &\qquad + \delta E V^{k-1}(s + g(s, z_r), z_r, z_r + \xi) \\ W^k(s, z_m, z_r) &= \pi g(s, z_m) \\ &\qquad + \delta E V^{k-1}(s + g(s, z_m), z_m, z_r + \xi) \\ V^k(s, z_m, z_r) &= \max\{I^k(s, z_m, z_r), W^k(s, z_m, z_r)\} \\ V^0(s, z_m, z_r) &= 0 \quad \forall (s, z_m, z_r) \geq (0, 0, z_m). \end{aligned}$$

$\Delta_{z_m} W^k(s, z_m, z_r) \geq 0$ holds trivially at $k = 0$ by the choice of $V^0(s, z_m, z_r)$. Assume that $\Delta_{z_m} W^k(s, z_m, z_r) \geq 0$ holds for some $k \geq 0$. Then,

$$\begin{aligned} &V^k(s + g(s, z_m), z_m + 1, z_r) - V^k(s + g(s, z_m), z_m, z_r) \\ &\quad = \max\{I^k(s + g(s, z_m), z_m + 1, z_r), \\ &\qquad\qquad W^k(s + g(s, z_m), z_m + 1, z_r)\} \\ &\qquad - \max\{I^k(s + g(s, z_m), z_m, z_r), \\ &\qquad\qquad W^k(s + g(s, z_m), z_m, z_r)\} \geq 0, \end{aligned}$$

where the inequality follows because

$$\begin{aligned} &W^k(s + g(s, z_m), z_m + 1, z_r) \\ &\qquad - W^k(s + g(s, z_m), z_m, z_r) \geq 0 \quad \text{and} \\ &I^k(s + g(s, z_m), z_m + 1, z_r) = I^k(s + g(s, z_m), z_m, z_r). \end{aligned}$$

Since $g(s, z_m + 1) - g(s, z_m) \geq 0$ by Assumption 1(i), the inequality above combined with Proposition 2 implies

$$\begin{aligned} &[V^k(s + g(s, z_m + 1), z_m + 1, z_r) - V^k(s + g(s, z_m), z_m, z_r)] \\ &\qquad \geq -\pi[g(s, z_m + 1) - g(s, z_m)]. \end{aligned} \tag{7}$$

Therefore,

$$\begin{aligned} &\Delta_{z_m} W^{k+1}(s, z_m, z_r) \\ &\quad = \pi[g(s, z_m + 1) - g(s, z_m)] \\ &\qquad + \delta E[V^k(s + g(s, z_m + 1), z_m + 1, z_r + \xi) \\ &\qquad\qquad - V^k(s + g(s, z_m), z_m, z_r + \xi)] \\ &\quad \geq \pi[g(s, z_m + 1) - g(s, z_m)] \\ &\qquad + \delta E[-\pi(g(s, z_m + 1) - g(s, z_m))] \\ &\quad = (1 - \delta) \cdot \pi[g(s, z_m + 1) - g(s, z_m)] \geq 0. \end{aligned}$$

This completes the induction step, so that for all k,

$$W^k(s + g(s, z_m), z_m + 1, z_r) - W^k(s + g(s, z_m), z_m, z_r) \geq 0.$$

Taking the limit in k, $\Delta_{z_m} W(s, z_m, z_r) \geq 0$. □

Theorem 1 establishes the optimality of a threshold policy. For a given technology z_r available in R&D, the plot of Z_m^* versus s defines a switching curve that divides the (s, z_m) state space into two regions. The space $\{(s, z_m): z_m < Z_m^*(s, z_r),\ s \geq 0\}$ is the region below the curve where the optimal decision is to introduce the next product generation. For the region on and above the curve, $\{(s, z_m): z_m \geq Z_m^*(s, z_r),\ s \geq 0\}$, the optimal decision is to delay introduction. Figure 1 describes a typical optimal switching curve for a firm with technology level z_r in R&D.

The complete optimal policy is described by a set of technology switching curves $\{Z_m^*(s, z_r),\ z_r \geq 1\}$, one for each possible state of R&D technology. Hence, for the multiple-introduction scenario, these curves span the (s, z_r) state space as shown in Figure 2.

Figure 1 Sample Z_m^* Switching Curve

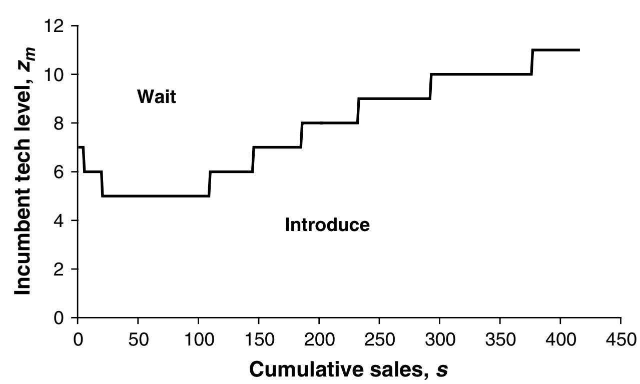
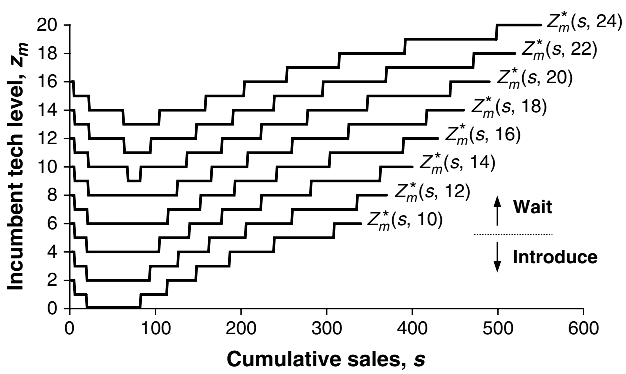


Figure 2 **Sample Optimal Policy: Z_m^* Switching Curves**



Reconsider the definition of the technology thresholds for the optimal policy. Comparing z_r and z_m, one would expect the switching curve Z_m^* (above which the firm does not introduce new technology) to be monotone nondecreasing in z_r. Indeed, Figure 2 depicts this. This monotonicity behavior is intuitive and is observed in all numerical scenarios examined by the authors.

Fixing s and examining Z_m^* versus z_r as in Figure 3, it is clear that monotonicity of Z_m^* in z_r implies existence of an equivalent set of thresholds Z_r^* as a function of z_m. In this interpretation, for given (s, z_m), $Z_r^*(s, z_m)$ designates the R&D technology level above which introduction of a new product generation is optimal. This observation is stated as the following proposition.

Proposition 3. *If $Z_m^*(s, z_r)$ is nondecreasing in z_r, then for each $(s, z_m) \geq (0, 0)$ there exists a threshold $Z_r^*(s, z_m) \geq z_m$ such that the decision to introduce the next product generation is optimal if and only if the technology in R&D $z_r > Z_r^*(s, z_m)$.*

Figure 3 **Switching Curve Monotonicity in z_m (s fixed)**

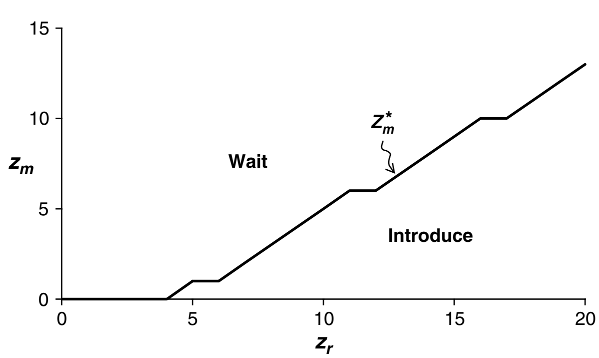


Figure 4 **Sample Z_r^* Switching Curves**

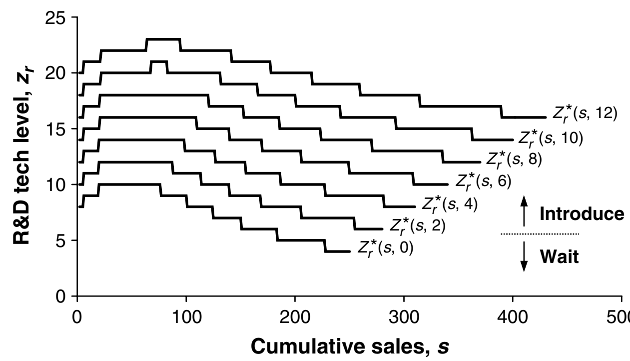


Figure 4 illustrates an optimal introduction policy represented in the form of Proposition 3. Here, the policy is specified by the set of switching curves $\{Z_r^*(s, z_m), z_m \geq 0\}$, one for each possible state of incumbent product technology z_m. Note that it is more natural to continue the present analysis in terms of the thresholds Z_r^*, and doing so does not affect structural insights because Z_r^* and Z_m^* are in one-to-one correspondence under the monotonicity assumption. Hence, we employ the Z_r^* thresholds of Proposition 3 for all further discussion of the optimal policy.

Further managerial insight on the optimal policy is gained by examining the effects of model parameter changes. We investigate such effects in the computational study below. We then revisit the conclusions of previous literature regarding the product-introduction time decisions for the IBM Mainframe Case.

5. Computational Study and Insights

In the previous section, we proved the optimality of a threshold policy for firms deciding when to introduce the next product generation. We now study the introduction time problem numerically to provide insight into the general pattern of product introductions and to understand how the optimal introduction thresholds and the related pace of product introduction are influenced by key model inputs. The numerical study focuses on the influences of a simple technology discovery rate, fixed product-introduction costs, and market parameters including the diffusion coefficients and a parameter describing the sensitivity of product market potential to changes in product technology.

The topic of uncertain demand is also reviewed with a presentation of model formulations that accommodate various forms of demand uncertainty.

Finally, we examine IBM's introductions of successive generations of their mainframe computer systems from 1954 to 1978. Insights are contrasted to those of Mahajan and Muller (1996) who previously studied the case using their "now or at maturity" product-introduction guidelines.

5.1. Influence of Key Model Parameters

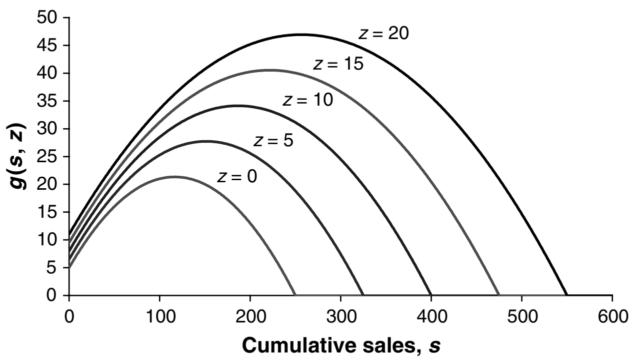
Although the analytical results of §4 apply to the more-general framework laid out in §3, for purposes of numerical investigation we begin with a simplified baseline scenario. Sales rate curves for the baseline scenario are generated within a discrete-time framework to approximate a demand diffusion process according to the form given in (6). Technology improvement is assumed to follow a simplified stochastic process in which available technology in R&D increases by one in each period with probability p. Product market potential, $N(z_m)$, is taken as a linear function of the incumbent product technology level. The market potential is obtained from the formula $N(z_m) = N(0) + mz_m$, where $N(0)$ is the market potential at the initial incumbent technology level and m represents the (linear) sensitivity of market potential to product technology. Parameter values for the baseline decision problem are given in Table 1. Figure 5 shows selected one-period sales curves $g(s, z)$ for the baseline scenario.

The baseline optimal policy is computed by solving the dynamic program (3). In solving (3) numerically, we use linear interpolation to handle cases in which the current period sales $g(s, z)$ is a noninteger multiple of the indexing unit used for cumulative sales. Linear interpolation allows discretization of cumulative sales s without altering the data for current sales $g(s, z)$. For the baseline scenario, numerical computation using this technique proceeds under the

Table 1 Baseline Data

Parameter	a	b	$N(0)$	δ	K	π	p	m
Value	0.02	0.30	250	0.9	20	0.75	0.2	15

Figure 5 Baseline Sales Curves



DP approximation:

$$\begin{aligned}
I(s, z_m, z_r) &= -K + \pi g(s, z_r) + \delta[g(s, z_r) - \underline{g}(s, z_r)] \\
&\quad \cdot [pV(s + \bar{g}(s, z_r), z_r, z_r + 1) \\
&\qquad + (1-p)V(s + \bar{g}(s, z_r), z_r, z_r)] \\
&\quad + \delta[\bar{g}(s, z_r) - g(s, z_r)] \\
&\quad \cdot [pV(s + \underline{g}(s, z_r), z_r, z_r + 1) \\
&\qquad + (1-p)V(s + \underline{g}(s, z_r), z_r, z_r)] \\
W(s, z_m, z_r) &= \pi g(s, z_m) + \delta[g(s, z_m) - \underline{g}(s, z_m)] \\
&\quad \cdot [pV(s + \bar{g}(s, z_m), z_m, z_r + 1) \\
&\qquad + (1-p)V(s + \bar{g}(s, z_m), z_m, z_r)] \\
&\quad + \delta[\bar{g}(s, z_m) - g(s, z_m)] \\
&\quad \cdot [pV(s + \underline{g}(s, z_m), z_m, z_r + 1) \\
&\qquad + (1-p)V(s + \underline{g}(s, z_m), z_m, z_r)] \\
V(s, z_m, z_r) &= \max\{I(s, z_m, z_r), W(s, z_m, z_r)\},
\end{aligned}$$

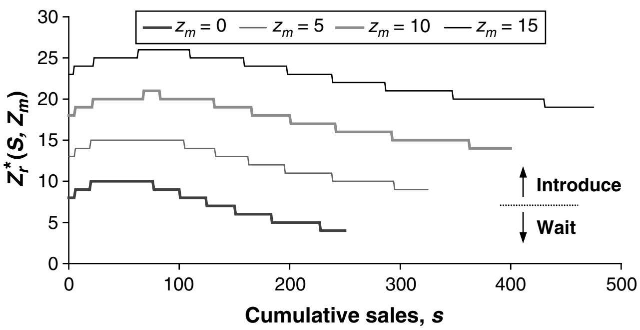
where, $\underline{g}(s, z)$ and $\bar{g}(s, z)$ represent the floor and ceiling of $g(s, z)$, respectively.

Numerical approximation generates the baseline set of technology switching curves illustrated in Figure 6. In reference to the baseline scenario, we first discuss the implications of the threshold policy for a firm deciding when to introduce a new product generation. We then analyze the effects of individual parameter changes on the optimal thresholds.

5.1.1. Pattern of Optimal Product Introductions.

The switching curves in Figure 6 suggest that optimal introduction of the next product generation may

Figure 6 Optimal Switching Curves for Baseline Scenario



be triggered in one of two ways: (1) through sufficient product sales at the current technology levels, or (2) through significant advances in available product technology. (1) implies that, even without further gains in R&D, a firm that continues to sell the current generation long enough may eventually find it optimal to introduce the technology on hand even though introducing at the same technology level was not profitable in the past. (2) implies that, regardless of the current generation's position along its sales curve, it may be optimal to introduce a new generation with large enough gains in R&D technology. Note that these observations generalize and extend the "now or at maturity" rule of Mahajan and Muller (1996) in that the optimal policy favors later introductions while still allowing introduction throughout the current product's horizon given sufficient technological advances.

The observed pattern of optimal product introductions supports the notion that, in the absence of credible competition, it may be profitable to postpone introduction of an available product generation due to strong sales of the current generation. Such introduction strategies have been observed in the semiconductor industry. Consider Intel Corporation's introduction of the Pentium microprocessor in 1993. The dominant producer of PC microprocessors, Intel continued to sell as many 486 processors as it could produce in late 1992. While originally slated for introduction in December 1992, Intel delayed the introduction of the Pentium processor until May 1993 even though no technical or production problems motivated the delay. As Brandt (1993) notes, analysts at the time estimated that Intel would add \$112 million to its 1993 profits by waiting for 486 revenues to begin to drop before releasing the Pentium chip. Intel's actions support our findings in highlighting the importance of considering sales of the current generation when timing a new introduction. The IBM example at the end of this section provides an example of the alternative situation where introduction timing coincides with the availability of sufficient product technology.

We next assess the effects of individual parameter changes on the optimal thresholds by systematically varying the input parameters for the baseline scenario.

5.1.2. Technology Discovery Probability. Consider the expected rate of technology discovery as measured (for the baseline scenario) by the technology discovery probability p. Though treated as an exogenous process in the model, in reality, a firm may have some control over the distribution of technological gains via its level of investment in R&D resources. Thus, understanding the policy implications of changing p adds insight to the problem of establishing a sound structure of R&D expenditure.

The policy effects of varying the discovery probability for the baseline scenario are shown in Figure 7. It is natural that under fixed introduction costs, a firm with a higher technology discovery rate will introduce new product generations when the gain in technology over the previous generation is larger. Hence, for a given technology lag between the product generation in the market and that available in R&D, an increase in the technology discovery probability should increase the attractiveness of the decision to wait versus introduce. This is precisely the relation reflected in the figure.

Figure 7 Influence of Technology Discovery Probability p

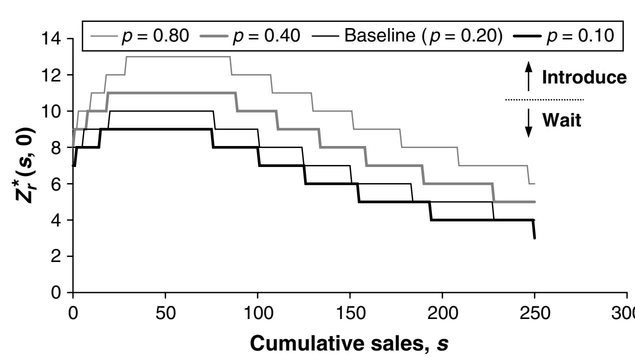


Table 2 Expected Introduction Point vs. p

p	0.1	0.2	0.4	0.8
$E[t^*]$	32.2	21.2	14.9	11.5
$s@E[t^*]$	250	247	218	166

Because varying p changes the expected pace of technology improvement, a stand-alone examination of the effects of p on introduction thresholds does not reveal its effects on the pace of product introduction. Further insight is gained by examining Table 2 below. For the different p values, Table 2 depicts the expected number of periods $E[t^*]$ and associated cumulative sales s at which the R&D technology will surpass the introduction threshold $Z_r^*(s,0)$. It is evident that although the technology introduction thresholds are increasing in the discovery probability (as shown in Figure 7), the expected rate, both in terms of time and sales, at which the firm will introduce new generations is also increasing; i.e., firms with a higher expected technology discovery rate are expected to introduce new product generations more frequently and with larger technology gains between generations.

5.1.3. Fixed Introduction Cost. Next, we examine the influence of the firm's cost structure on the optimal policy. As illustrated in Figure 8, a decrease in the fixed introduction cost K decreases the optimal introduction threshold at any given cumulative sales level. Hence, as intuition would suggest, a decrease in fixed introduction costs translates to more-frequent product introductions over time. Obviously, under the scenario with no introduction costs ($K=0$), introduction of the next product generation is always optimal as soon as new technology becomes available. This is reflected in the figure by $Z_r^*(s,0)$ constant at 1 across all s when $K=0$.

Let us turn now to the parameters that describe the product market. The modeled product sales dynamics will be affected by the diffusion coefficients a and b in (6) as well as the market potential parameter m, that determines the product market potential for a specific technology level. Uncertainties in the values of the diffusion coefficients a, b, and the parameter m are likely to exist because they would typically be estimated from the sales data of earlier generations. Thus, to establish an intelligent introduction policy, it is important that the decision maker understands the effect that varying these values has on the optimal policy.

5.1.4. Market Potential Parameter. As defined for the baseline scenario, an increase in m translates to larger gains in market potential per unit gain in technology. In turn, the sales rate at a given level of cumulative sales is more sensitive to increases in product technology when m is higher. Thus, an increase in m favors the decision to introduce over the decision to wait. Figure 9 confirms that all else equal the introduction threshold is decreasing in m. Hence, a firm is more likely to introduce new product generations with small gains in technology when product sales rates are highly technology dependent. Moreover, a firm that incorrectly estimates m too large would introduce new product generations more frequently than is optimal.

5.1.5. Demand Diffusion Coefficients. Given the product market potential, the diffusion coefficients a, b, characterize the product diffusion dynamics. Under

Figure 8 Influence of Fixed Cost K

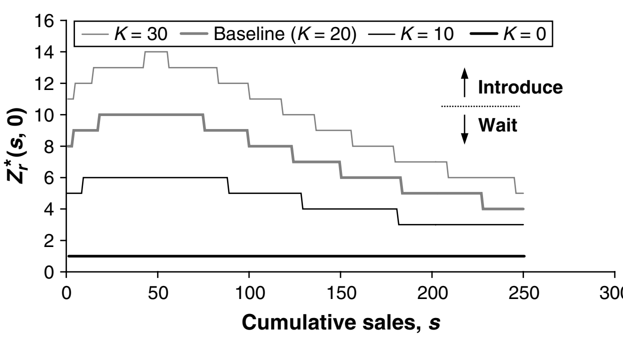


Figure 9 Influence of Market Potential Parameter m

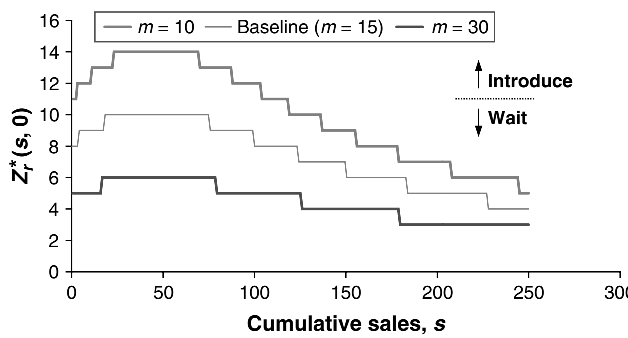


Figure 10 Sales Rate Curves for Different *a* Values ($b = 0.30$)

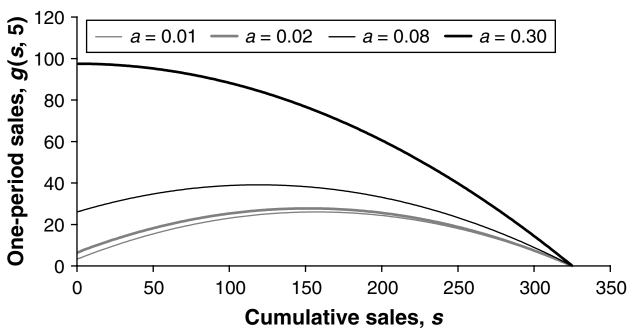
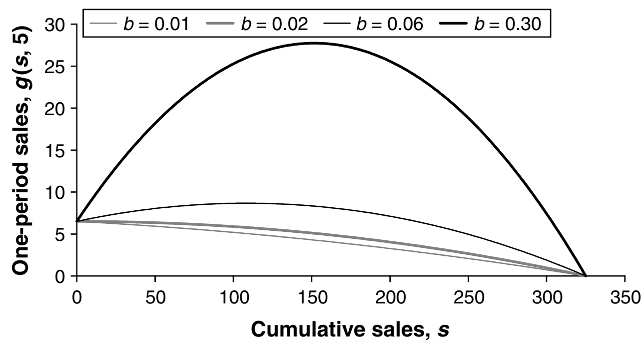


Figure 12 Sales Rate Curves for Different *b* Values ($a = 0.02$)



the sales model (6), a product with a higher coefficient of innovation a would exhibit a sales rate curve that starts with higher one-period sales, peaks earlier, and lies completely above that of a product with lower a. Figure 10 illustrates how the coefficient of innovation influences the product sales rate curves. The policy effect of these changes is summarized in Figure 11.

We find that a product with higher coefficient of innovation is associated with more-frequent product introductions. This is a logical result under discounting because the added sales revenues due to introducing the new product generation are realized earlier for a product with higher coefficient of innovation, hence increasing the attractiveness of introducing versus waiting. It follows that the policy obtained using a high estimate for a would direct the firm to introduce new product generations more frequently than is optimal.

Similar to the effect of a, a higher coefficient of imitation b translates to a sales rate curve that lies completely above that for lower b. Figures 12 and 13, respectively, show the influence of the coefficient of imitation on a product sales rate curve and the corresponding effects on the optimal switching curve. As with a, the introduction thresholds are decreasing in the product's coefficient of imitation b. Thus, a firm should introduce new product generations more frequently given a base technology that diffuses through its potential adopter population faster.

As shown in this numerical study, the optimal introduction policy exhibits a well-behaved structure under the specific case in which product sales follow the demand diffusion form (6), technology in R&D follows a simple jump process, and product market potential is linear in product technology. Table 3 summarizes the effects of changing model parameters. Moreover, the policy form resulting from our model is intuitive and helps to explain firm behavior, as we demonstrate further in the discussion of IBM's mainframe computer introductions (§5.3).

5.2. Uncertain Demand

Thus far, our attention has focused on deterministic models of demand to discern the influence of technology improvement and demand diffusion dynamics

Figure 11 Influence of Coefficient of Innovation *a*

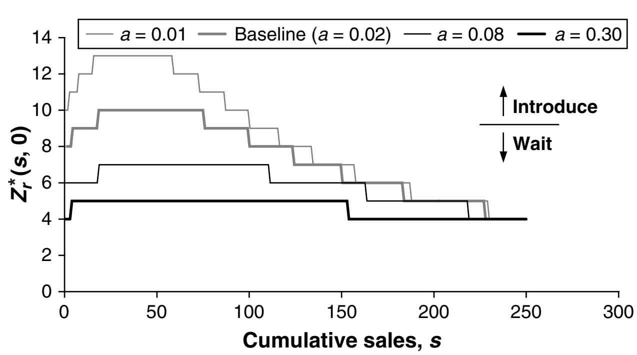


Figure 13 Influence of Coefficient of Imitation *b*

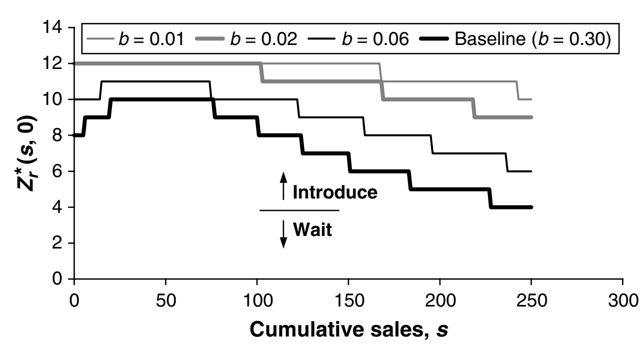


Table 3 Summary of Key Parameter Influences

Increase parameter	p	K	m	a	b
Effect on $Z_r^*(s, z_m)$	↑	↑	↓	↓	↓
E[introduction frequency]	↑	↓	↑	↑	↑

on a firm's pattern of successive introduction decisions. The demand diffusion framework constructed in §3 is, however, easily modified to accommodate demand uncertainty. Here, we describe two possible scenarios for uncertain demand along with the revised decision-model formulations. A brief numerical example is presented to demonstrate the persistence of the threshold policy structure established in §4.

Consider the simple case where the overall pattern of demand follows a demand diffusion process, but the actual single-period sales rates may deviate from the value predicted. Here, the sales rate in each period i may be modeled as the demand diffusion forecast value $g(s, z_m)$ modified by a simple disturbance factor to reflect uncertainty in the actual sales for the period. In this case, the uncertain one-period sales rate is given by $g(s, z_m)\gamma_i$, where the γ_i are i.i.d. with distribution H on $\Re^+$ and $E\gamma_i < \infty$. The decision model (1)–(3) is reformulated as follows to accommodate this single-period demand uncertainty:

$$V(s, z_m, z_r) = \max\{I(s, z_m, z_r), W(s, z_m, z_r)\} \quad (8)$$

$$I(s, z_m, z_r) = -K + E_\gamma[\pi g(s, z_r)\gamma + \delta E_\xi V(s + g(s, z_r)\gamma, z_r, z_r + \xi)] \quad (9)$$

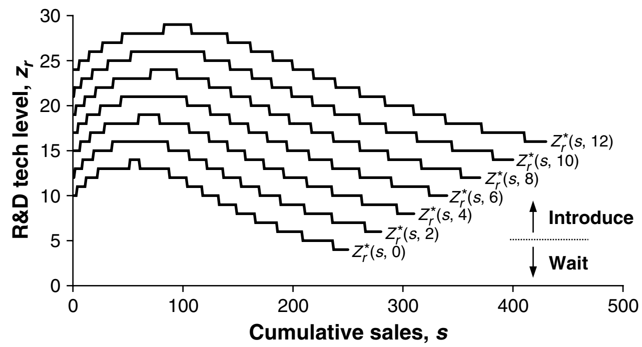
$$W(s, z_m, z_r) = E_\gamma[\pi g(s, z_m)\gamma + \delta E_\xi V(s + g(s, z_m)\gamma, z_m, z_r + \xi)]. \quad (10)$$

Building on the baseline scenario specified in Table 1, we define the following discrete distribution for the disturbance factors γ_i.

$$\gamma_i = \begin{cases} 1.2 & \text{w/prob } 0.2 \\ 1 & \text{w/prob } 0.5 \\ 0.8 & \text{w/prob } 0.3. \end{cases}$$

The decision problem (8)–(10) is solved numerically for the baseline data set with demand uncertainty specified as above. The resulting optimal policy has the familiar form summarized by the set of technology switching curves shown in Figure 14.

Figure 14 Uncertain Demand: Z_r^* Switching Curves



A firm may also wish to capture the uncertainty in overall market acceptance of a new product generation while taking into account potential correlation between the market success of two sequential generations. For the demand diffusion case, the estimation of a new generation's market potential $N(z)$ is a key source of such uncertainty. The present model framework can capture the market success uncertainty for a new generation using a Markov modulated demand formulation. A possible implementation is presented here for illustration.

Let $d \in D$ represent the state of overall market success for the current product generation (e.g., $D = \{\text{low, neutral, high}\}$). Define a transition probability matrix $P = [(p_{ij})]$, $i, j \in D$, where p_{ij} is the probability the next product generation has success level j given the current generation is experiencing success level i. Finally, let the market potential $N(z_m)$ for an incumbent product be scaled by a function $\phi(d)$, $0 < \phi(d) < \infty$, such that a generation's effective market potential is given by $\phi(d)N(z_m)$. Under this Markov modulated framework, the demand diffusion instance of the sales rate curves (6) is redefined to reflect the relationship between effective market potential and current state of demand:

$$g(d, s, z_m) = \left(a + b\frac{s}{\phi(d)N(z_m)}\right)(\phi(d)N(z_m) - s)^+. \quad (11)$$

The decision model (12)–(13) implements this Markov modulated demand framework for accommodating uncertainty in product success.

$$V(d,s,z_m,z_r) = \max\{I(d,s,z_m,z_r), W(d,s,z_m,z_r)\} \quad (12)$$

$$I(d,s,z_m,z_r) = -K + \sum_{d'\in D} p_{dd'}[\pi g(d',s,z_r) + \delta EV(d',s+g(d',s,z_r),z_r,z_r+\xi)] \quad (13)$$

$$W(d,s,z_m,z_r) = \pi g(d,s,z_m) + \delta EV(d,s+g(d,s,z_m),z_m,z_r+\xi). \quad (14)$$

Equations (8)–(10) and (12)–(13) represent two possible implementations of demand uncertainty within the present demand diffusion framework. The numerical example for (8)–(10) demonstrates that the optimal policy form established for the deterministic demand scenario also holds for simple uncertain demand scenarios. Persistence of this threshold policy form has also been verified for simple numerical cases of (12)–(13). Hence, adding demand uncertainty is likely to influence the actual decisions, but not the qualitative results of the model.

5.3. The Role of Technology Improvement: IBM Mainframe Introductions

Consider IBM's mainframe computer introduction policy from 1954 to 1978 as examined in Mahajan and Muller (1996). During this period, IBM introduced four successive generations of mainframe computer systems, each incorporating significant technological advance over the previous generation. Product technology progressed from vacuum-tubes in the first generation, to transistors in the second, integrated circuits in the third (360 family), and finally silicon chips in the fourth generation (370 family). Sales for each product generation followed a pattern of diffusion and substitution, which is described in detail by Mahajan and Muller.

Mahajan and Muller (1996) examine IBM's introduction timing decisions using their "now or at maturity" rule. They conclude that IBM introduced the 360 and 370 product families too late (at "maturity" rather than "now"). A shortcoming of their analysis lies in the implicit assumption that the appropriate technology was available and suitable for introduction at an earlier date. To the contrary, significant technological discovery occurred for the 360 family between Mahajan and Muller's suggested introduction date of 1963 (or 1964, depending on assumed cost of capital) and the actual introduction date of 1965. Brock (1975, pp. 16) notes: "As with the transistor, the integrated circuit was expensive at first and only competitive in applications where size was more important than cost.... A late 1965 report dated the competitiveness of integrated circuits at May 1964, when Fairchild Semiconductor made substantial price cuts in an attempt to stimulate commercial usage of the circuits." Furthermore, IBM matched the timing of this technology "discovery" with an April 1964 announcement of its new 360 product generation, which was based on integrated circuit technology. The timing of IBM's announcement supports the primary result of this paper, that a new generation of an innovative product must surpass a minimum technology threshold before introduction is optimal.

By allowing for technology improvements, we have expanded on the analysis of Mahajan and Muller to more accurately model the dynamics that drive successive product introductions. As evident in the case of IBM, a firm must consider the effects of improving technology when deciding whether to introduce a new generation of its product. Under continuing technological progress, a strictly "now or never" or "now or at maturity" policy is not necessarily applicable to governing a firm's introduction decisions. Under such a scenario, the threshold policy of Theorem 1 should be incorporated into the firm's product-introduction strategy.

6. Conclusion

In this paper, we constructed a decision model to examine a firm's introduction timing problem for successive generations of an innovative good. The model accommodates a demand diffusion specification of sales while allowing for uncertain technology improvement over time. Using this model, we reframed the analysis of Mahajan and Muller (1996) to consider the effects of continuing R&D progress on a firm's product-introduction decisions.

We proved the optimality of a state-dependent threshold policy governing the firm's product-introduction decisions. It was shown that the firm must compare the incumbent technology level to that available in R&D to determine, for a given level of cumulative sales-to-date, whether introduction of a new product generation is optimal. Under a specific demand diffusion form of sales, the threshold policy expands the "now or at maturity" rule of Mahajan and Muller (1996) in that the policy favors later introductions, while still permitting introduction at intermediate times given sufficient advances in available technology. This policy more closely explains the actions observed in practice, as demonstrated through the examination of IBM's mainframe computer introduction policy. Numerical analysis further characterized the optimal policy and its dependence on model inputs under the demand diffusion scenario.

Although generating insight on the dynamics guiding firms' new product-introduction behavior was the primary objective of this work, one could imagine application of the decision model in an actual setting. This would require estimation of information similar to the data summarized in Table 1. Norton and Bass (1987), and Mahajan and Muller (1996) discuss methods of estimating the diffusion parameters a, b, and N by fitting sales data from previous product generations. The dependence of N on product technology can be estimated using properly designed choice models (e.g., logit, probit, and their refinements). Finally, parameters π, K, and δ can be obtained via information on the firm's internal cost structure.

A limitation of our analysis is that the model does not consider the effects of introduction timing on consumers' purchase strategies and resulting demand patterns. Significant increase in the firm's pace of product introductions may cause consumers to postpone purchase decisions in anticipation of forthcoming product improvements (e.g., Dhebar 1994, Kornish 2001). Endogenizing consumer behaviors to account for this phenomenon represents a possible extension to the current work.

Our model considers the trade-off between waiting for further technology advances or realizing increased market potential sooner. The model does not however capture the time-to-market concerns that may arise in a competitive market setting. As described in Hendricks and Singhal (1997) the cost of delay (and hence value of earlier introduction) in such environments can be significant. One possible direction for future research is to consider a game theoretic model with multiple firms. An alternative approach may be to follow Cohen et al. (1996) and impose a fixed window of time for new product introduction. Such a constraint implicitly captures time-to-market concerns. Further research is needed in this area to examine the effects of market competition more explicitly.

One can conceive of other extensions to the current analysis. Obvious directions are to generalize the framework to incorporate upgrade purchases or to permit simultaneous sale of multiple product generations with substitution effects. Allowing product price to fluctuate over time and differ between product generations would provide additional insight on firm policies. Our computational results suggest that the expected pace of product improvement can significantly affect the firm's optimal introduction policy. Extending the model to incorporate a decision variable for the expected pace of product innovation (e.g., as measured by R&D investment) would enable a more-complete analysis of firm policies.

References

Balcer, Y., S. A. Lippman. 1984. Technological expectations and adoption of improved technology. *J. Econom. Theory* **34** 292–318.

Bass, F. M. 1969. A new product growth model for consumer durables. *Management Sci.* **15** 215–227.

Brandt, R. 1993. Intel: What a tease–and what a strategy. *Bus. Week* (Feb. 22) 40.

Brock, G. W. 1975. *The U.S. Computer Industry: A Study of Market Power*. Ballinger, Cambridge, MA.

Chatterjee, R., J. Eliashberg. 1990. The innovation diffusion process in a heterogeneous population: A micromodeling approach. *Management Sci.* **36**(9) 1057–1079.

Cohen, M. A., J. Eliashberg, T. Ho. 1996. New product development: The performance and time-to-market tradeoff. *Management Sci.* **42**(2) 173–186.

Dhebar, A. 1994. Durable-goods monopolists, rational consumers, and improving products. *Marketing Sci.* **13**(1) 100–120.

Farzin, Y. H., K. J. M. Huisman, P. M. Kort. 1998. Optimal timing of technology adoption. *J. Econom. Dynam. Control* **22** 779–799.

Gjerde, K. A. P., S. A. Slotnick, M. J. Sobel. 2002. New product innovation with multiple features and technology constraints. *Management Sci.* **48**(10) 1268–1284.

Hendricks, K. B., V. R. Singhal. 1997. Delays in new product introductions and the market value of the firm: The consequences of being late to the market. *Management Sci.* **43**(4) 422–436.

Jensen, R. 1982. Adoption and diffusion of an innovation of uncertain profitability. *J. Econom. Theory* **27** 182–193.

Jun, D. B., Y. S. Park. 1999. A choice-based diffusion model for multiple generations of products. *Tech. Forecasting Soc. Change* **61** 45–58.

Kalish, S. 1985. A new product adoption model with price, advertising, and uncertainty. *Management Sci.* **31**(12) 1569–1585.

Kornish L. 2001. Pricing for a durable-goods monopolist under rapid sequential innovation. *Management Sci.* **47**(11) 1552–1561.

Krishnan, T. V., F. M. Bass, D. C. Jain. 1999. Optimal pricing strategy for new products. *Management Sci.* **45**(12) 1650–1663.

Kumar, S., J. M. Swaminathan. 2003. Diffusion of innovations under supply constraints. *Oper. Res.* **51**(6) 866–879.

Lilien, G. L., E. Yoon. 1990. The timing of competitive market entry: An exploratory study of new industrial products. *Management Sci.* **36**(5) 568–585.

Mahajan, V., E. Muller. 1996. Timing, diffusion, and substitution of successive generations of technological innovations: The IBM mainframe case. *Tech. Forecasting Soc. Change* **51** 109–132.

Mahajan, V., E. Muller, F. M. Bass. 1990. New product diffusion models in marketing: A review and directions for research. *J. Marketing* **54**(1) 1–26.

Norton, J. A., F. M. Bass. 1987. A diffusion theory model for adoption and substitution for successive generations of high-technology products. *Management Sci.* **33**(9) 1069–1086.

Puterman, M. L. 1994. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley and Sons, New York, 151.

Robinson, B., C. Lakhani. 1975. Dynamic price models for new-product planning. *Management Sci.* **21**(10) 1113–1122.

Wilson, L. O., J. A. Norton. 1989. Optimal entry timing for a product line extension. *Marketing Sci.* **8**(1) 1–17.