NORTH-HOLLAND

Timing, Diffusion, and Substitution of Successive Generations of Technological Innovations: The IBM Mainframe Case

VIJAY MAHAJAN and EITAN MULLER

ABSTRACT

Based on the behavioral assumptions of diffusion theory, this article proposes an extension of the Bass diffusion model that simultaneously captures the *substitution* pattern for each successive generation of a durable technological innovation, and the *diffusion* pattern of the base technology. Normative guidelines based on the model suggest that a firm should either introduce a new generation as soon as it is available or delay its introduction to a much later date at the maturity stage of the preceding generation. The decision depends on a number of factors including the relative size of the market potentials, gross profit margins, the diffusion and substitution parameters, and the discount factor of the firm. This "now or at maturity" rule is thus an extension and generalization of the "now or never" rule of Wilson and Norton [25]. Empirical and normative implications of the proposed model are explored for four successive generations on IBM mainframe computers: first generation (vacuum tubes); second generation (transistors); 360 family (integrated circuits); and 370 family (silicon chips). The model describes the growth of these generations well. The application of normative guidelines suggests that IBM introduced the two successive generations of 360 and 370 families too late, i.e., their time to market should have been shorter. Limitations and further extensions of the model and the application are discussed.

Introduction

The launch of a new product is a phase of new product development that commands a large commitment in time, money, and managerial resources. The costs of introduction include investments in production as well as in marketing. Whereas the magnitudes depend upon the product category, by far the largest investment in new product development is in full-scale launch [23]. Consequently, the new product launch requires careful planning to ensure the desired market success of a new product [12].

One of the key elements in the new product launch strategy is its introduction time. Given the possibility of cannibalizing the firm's existing product(s), and risks associated with premature or delayed product introduction, the timing questions are concerned with determining the optimal time for the new product introduction. Determination of optimal

VIJAY MAHAJAN is the John P. Harbin Centennial Chair in Business, Graduate School of Business, The University of Texas, Austin, Texas.

EITAN MULLER is Professor of Marketing, Leon Recanati Graduate School of Business Administration, Tel Aviv University, Tel Aviv, Israel.

Address correspondence to Eitan Muller, Leon Recanati Graduate School of Business Administration, Tel Aviv University, Tel Aviv 69978, Israel.

Technological Forecasting and Social Change 51, 109-132 (1996)

0040-1625/96/$15.00
655 Avenue of the Americas, New York, NY 10010
SSDI 0040-1625(95)00225-1

introduction time is especially critical for high-technology products, where the introduction of each successive generation of a product requires the firm to explicitly consider its impact on the demand for the preceding generations and vice versa. That is, although each successive generation of the product may create its new potential buyers, it may cause adopters or potential adopters of preceding generations to opt for it. Introduction time, therefore, may influence both the product's own diffusion and the diffusion of its preceding generations by the mechanism of substitution [13]. Hence, timing questions for successive generations of a product require simultaneous consideration both of substitution and of diffusion effects.

The objective of this study is to unify into one model the process of diffusion of the base technology and the process of substitution between the various generations of this technology in order both to predict sales and to determine optimal timing of new generations of the technology. An extension of the classical Bass diffusion model [1], our development is inspired by the pioneering works of Norton and Bass [18, 19] and Wilson and Norton [25]. Using a repeat-purchase framework for integrated circuit technologies, Norton and Bass [18] have proposed a model that relies on a substitution framework [e.g., 6] and diffusion model. The normative implications for the timing decisions of such a substitution model have been reported by Wilson and Norton [25]. Their analytical findings suggest that, for a monopolist, operating under a long planning horizon, a bang-bang or "now or never" is the optimal timing strategy. That is, a monopolist should either introduce a successive generation of a product as soon as it is available or else delay its introduction indefinitely.

The model proposed in this study is specifically developed for durable technological innovations. The normative implications for the timing decisions derived from this model suggest that, for a monopolist, a "now or at maturity" rule is the optimal timing strategy. That is, a firm should introduce a new generation as soon as it is available or else delay its introduction to a much later date at the maturity stage of the preceding generation. This date depends on a number of factors including the relative size of the market potentials, gross profit margins, the diffusion and substitution parameters and the discount factor of the firm. This "now or at maturity" rule is thus an extension and generalization of the "now or never" rule of Wilson and Norton [25]. The application of the model and the associated derived normative timing guidelines are illustrated for four generations of IBM mainframe computers: first generation (vacuum tubes), second generation (transistors), 360 family (integrated circuits), and 370 family (silicon chips).

The organization of this article is as follows. The next section proposes a model that incorporates the diffusion and substitution effects and depicts the overall growth of a durable technological innovation and the growth and the decline of its successive generations. Section 3 compares our approach with the two notable contributions of Norton and Bass [18] and Wilson and Norton [25]. We then derive the necessary analytics to recommend optimal introductory timing strategies for successive generations in section 4. Applications of the model and the introductory timing guidelines to the IBM data are reported in section 5. Finally, the article concludes with extensions of the model and the application, and the limitations of the model including the assumed monopoly market structure.

Substitution and Diffusion

MODEL DEVELOPMENT

To motivate the development of the model, consider the growth patterns for the four successive generations of IBM mainframe computers depicted in Figure 1 and Ta-

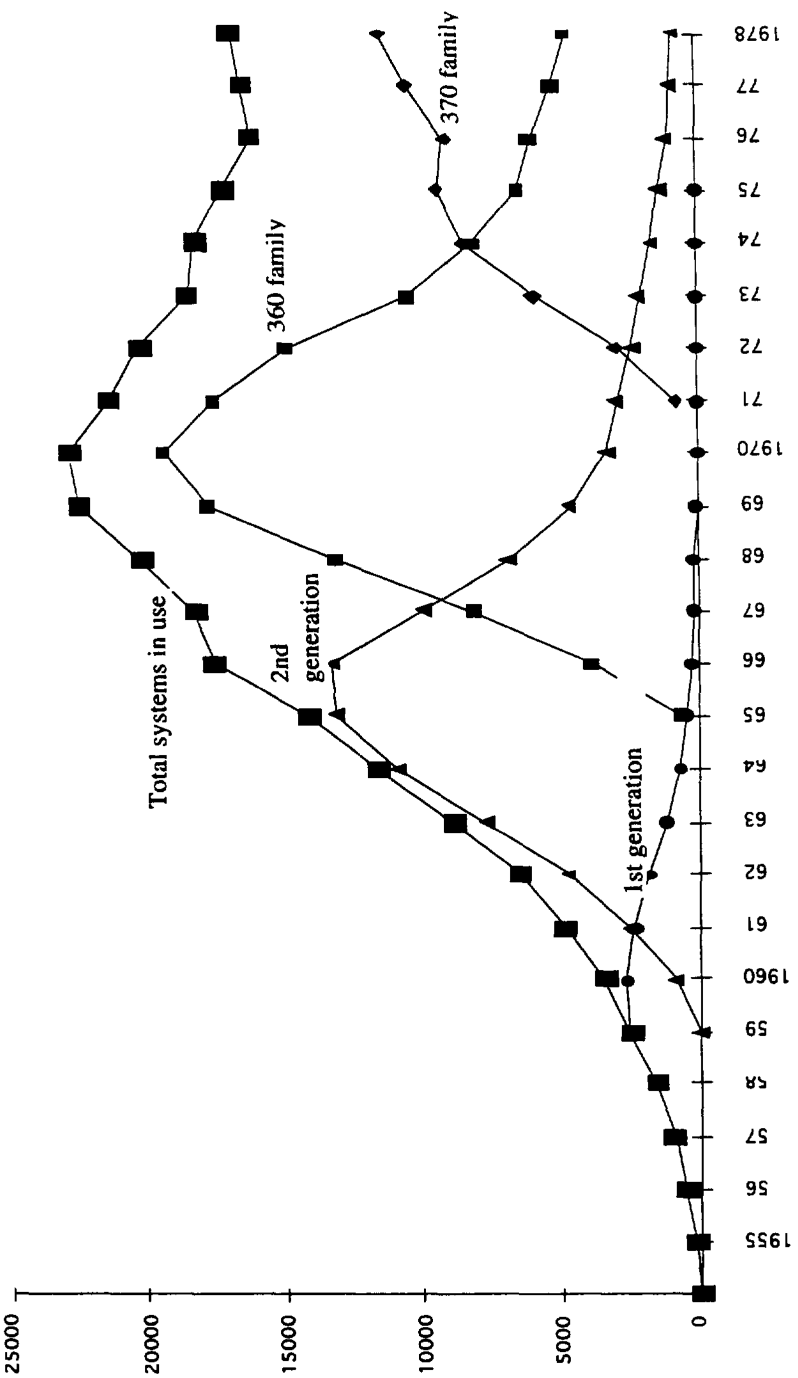

Fig. 1. Diffusion and substitution: IBM mainframe systems in use.

ble 1. These figures are based on data reported in Phister [20]. (For more information about IBM in these decades and later and an extensive history of the U.S. computer industry, see [3–5, 9, 17].) In addition to the total number, these patterns show the number of IBM mainframe *systems in use* per year, (and not current sales), for each of the four generations. These patterns capture the growth and decline of each generation and are amenable to substitution effects across generations and diffusion effects within each generation. Given these patterns, questions now relate to the development of a model that can describe and predict these patterns and the overall total growth of IBM mainframe systems, and assist in analyzing optimal introductory timing decisions.

To capture the diffusion and substitution dynamics shown in Figure 1, we propose here a model that is an extension of the classical diffusion model suggested by Bass [1]. The Bass model captures the diffusion dynamics among potential adopters by considering adoptions due to the external sources of information and the word-of-mouth communication (for a review of diffusion models, see [15]).

From Figure 1 and Table 1 we abstract the following definitions:

1. Let $x_i(t)$ be the number of systems-in-use of the i-th generation. Note from Figure 1 that this number, unlike cumulative number of adopters, can and does decrease in time when a new generation $(i + 1)$ is present.
2. Let $x(t)$ be the total number of systems-in-use of all generations available at time t. Thus $x(t) = x_1(t) + x_2(t) + \ldots + x_i(t)$ where the i-th generation is the latest generation available at time t. Note from Figure 1 that unlike cumulative number of adopters, this number can decrease.
3. Let T_i be the introduction time of the i-th generation. In the IBM case, $T_2 = 1959$, $T_3 = 1965$ and $T_4 = 1971$.
4. Each generation of the product creates its own market potential N_i. Thus N_i is the eventual number of systems-in-use of the i-th generation if the $i + 1$ generation is not introduced. Formally,

$$N_i = \lim_{t \Rightarrow \infty} x_i(t) \qquad \text{given } T_{i+1} = \infty.$$

5. Let dx/dt denote the change over time in the total number of systems in use and likewise dx_i/dt is the change in the number of systems in use of the i-th generation (at time t).

Note that we assume a monopoly market structure and that each adopter adopts one unit of the product independently of its computing power. We elaborate on these assumptions later on. In addition, note that as long as there is only one generation available, $x_i(t)$ will also be the number of cumulative adopters, and dx_i/dt will denote unit sales. *This interpretation, however, does not hold in the multi-generation case.* In the latter case, the number of systems-in-use at time t is equal to the number of adopters that use this generation of the product at time t. The number of systems-in-use will usually be less than the cumulative number of adopters, as some of them may switch to the new generation when it is available for adoption. Similarly, dx_i/dt can be thought of as unit sales only if generation $i + 1$ has not yet appeared on the market. After its appearance, dx_i/dt is the change in the number of systems-in-use of the i-th generation. The definition of N_i, however, presumes that no further generation is introduced. Thus it is the total number of systems-in-use, and also the total cumulative number of adopters of the i-th generation over its life time. In practice, of course, we do not come to see the realization

TABLE 1
IBM Systems in Use, by Generation

T_i Time of introduction	Year	$x_1(t)$ 1st Generation	$x_2(t)$ 2nd Generation	$x_3(t)$ 360 Family	$x_4(t)$ 370 Family	$x(t)$ Total
	1955	190				190
	1956	560				560
	1957	1000				1000
	1958	1680				1680
T_2	1959	2542	3			2545
	1960	2640	880			3520
	1961	2350	2510			4860
	1962	1820	4725			6545
	1963	1170	7720			8890
	1964	750	10,940			11,690
T_3	1965	455	13,090	625		14,170
	1966	303	13,330	3881		17,514
	1967	203	9977	8125		18,305
	1968	170	6896	13,110		20,176
	1969	49	4646	17,687		22,382
	1970	29	3297	19,412		22,738
T_4	1971	14	2916	17,529	806	21,265
	1972	6	2384	14,909	2922	20,221
	1973	4	2079	10,475	5887	18,445
	1974	4	1676	8060	8440	18,180
	1975	3	1397	6450	9335	17,185
	1976		1107	5919	9046	16,072
	1977		894	5118	10,450	16,462
	1978		829	4641	11,348	16,818

The introduction times T_i are defined unambiguously, except T_2. Note that only 3 units were sold in 1959, and thus the introduction was probably in late 1959. In the true sense of the word "introduction," 1960 seems the better choice. Indeed the empirical estimation is slightly better with $T_2 = 1960$.

of N_i because the growth of x_i is reduced and later might altogether stop once the $i + 1$ generation is introduced.

Note that in general we do expect the market to grow from one generation to the next, i.e., that N_i will be monotonic in i. This, however, does not necessarily hold in all cases. There might be circumstances under which the market potential of the $i + 1$ generation will be smaller than the market potential of the i-th generation. The model will still be valid as long as the drop in the market potential is not too drastic as to force the term $N_i - x$ to be negative.

The model is based on the flow dynamics of Figure 2: in addition to mainframe system adoptions from potential adopters who have not yet adopted any of the product generations, adoptions for each generation i may also come from adopters of earlier generations who either switch to it (i.e., adopters of $(i - 1)$th generation) or leapfrog other intermediate generations to adopt it. Consider, for example, in Figure 2D for the fourth generation. Mainframe system adoptions for generation 4 may arise from: (1) adopters of generation 1 who leapfrog generations 2 and 3, (2) adopters of generation 2 who leap-frog generation 3, and (3) adopters of generation 3 who switch to it, and (4) first-time adopters who have not yet adopted any of the preceding generations.

With these definitions, the rate equations can be specified now to depict the growth and decline of each generation and the overall total growth of IBM mainframe systems.

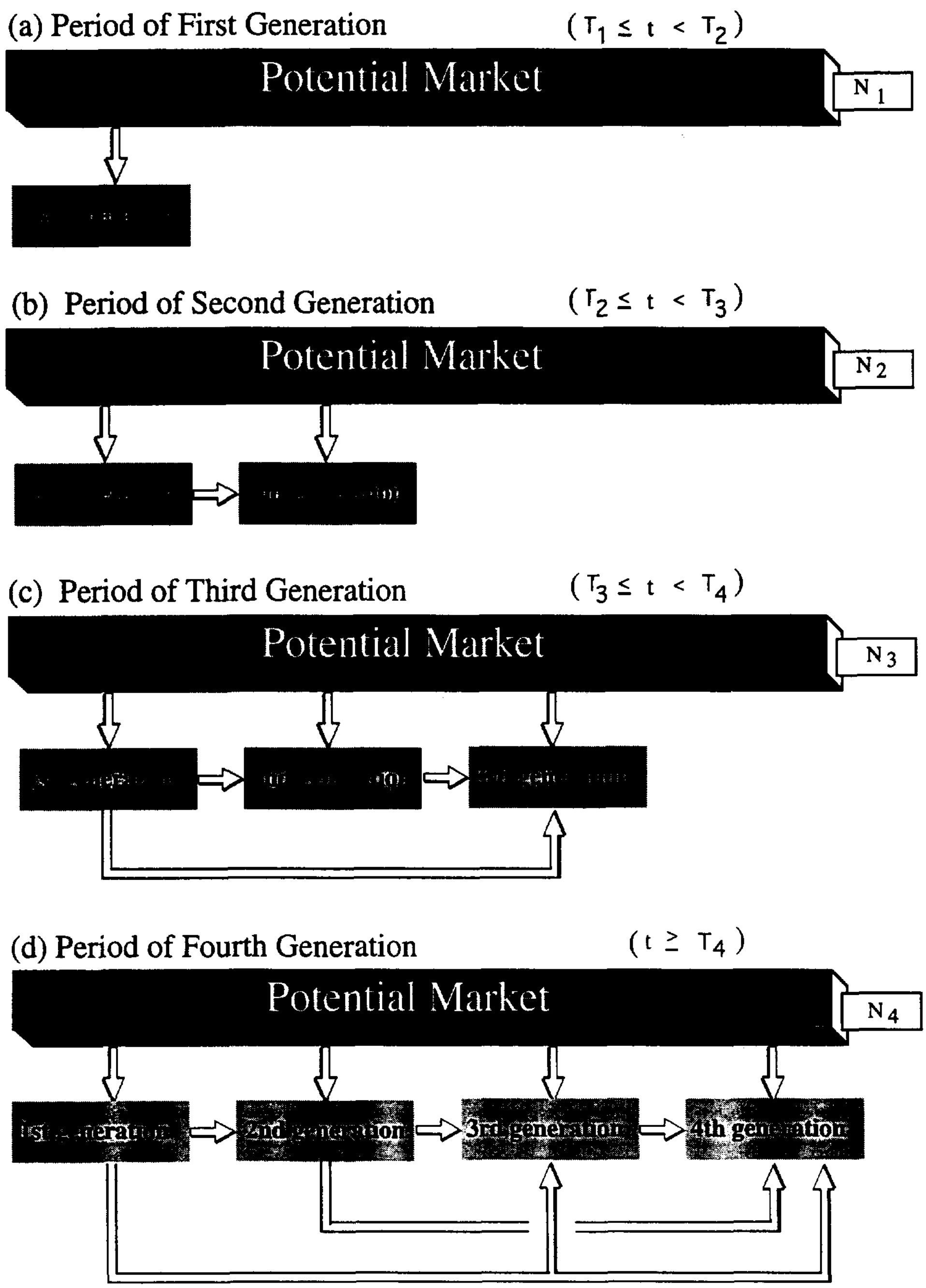

Fig. 2. Adopter flow dynamics across the various generations of a durable technological innovation.

Period of First Generation ($T_1 \leqslant t < T_2$)

Given that there is only one generation in the market place, the growth of generation 1 can be represented by the Bass model:

$$dx_1/dt = (a_1 + b_1x_1/N_1)(N_1 - x) \quad (1)$$

(recall that $x = \Sigma_{i=1}^{n}x_i$ where n is the latest available generation. In this case $x = x_1$).

Period of Second Generation ($T_2 \leqslant t < T_3$)

In this period, when there are two competing generations, non-adopters (i.e., $N_2 - x$) interact with previous adopters (i.e., x). As a result of these interactions a fraction α_2 of those who decided to adopt the base technology will purchase the new generation and a fraction $1 - \alpha_2$ will adopt the earlier generation. In addition there is a flow of consumers from the current adopters x_1 who upgrade (with presumably different parameters) by interacting with adopters of the new generation x_2. The two resulting equations are the following:

$$dx_2/dt = \alpha_2(a_2 + (b_1x_1 + b_2x_2)/N_2)(N_2 - x) + \alpha_2(a_2' + b_2'x_2/N_2)x_1 \quad (2)$$

$$dx_1/dt = (1 - \alpha_2)(a_2 + (b_1x_1 + b_2x_2)/N_2)(N_2 - x) - \alpha_2(a_2' + b_2'x_2/N_2)x_1 \quad (3)$$

Equations 2 and 3 are explained as follows: first, the market potential now becomes N_2 to take into account the fact that the introduction of a new generation has an effect on primary demand. However, whereas the introduction of the new generation increases the primary demand by $N_2 - N_1$, some of these new consumers may nonetheless buy the first generation. The first term in equations 2 and 3 is due to adoptions from the untapped market potential i.e., $N_2 - x$. Out of these adoptions, a fraction, α_2 adopt the second generation and the remainder $(1 - \alpha_2)$ still purchase the first generation. The second term in these two equations is due to adoptions from upgraders, who adopt with a different innovation and imitation parameters. For example, in our empirical analysis we estimate α_2 to be about 0.9, and thus $1 - \alpha_2$ is about 0.1. Thus there is a sharp drop in the sales of the first generation when the second generation is introduced. How sharp this drop is will be determined empirically by the parameter α. Note that in the transition from the first to the second generation, the coefficient of external influence (a) changes from a_1 to a_2 because this coefficient, as its name suggests, is directly related to the advertising efforts of the firm. Thus, if the firm introduces the new generation with an advertising campaign of the same intensity and efficiency as the previous generation, then this coefficient will remain unchanged. If however the new campaign is better or larger, then this coefficient will become larger. In addition, we assume that each generation of adopters has its own word-of-mouth coefficient, so that the introduction of the second generation of the product results in the new generation of adopters with its own word-of-mouth coefficient of b_2.

Period of Third Generation ($T_3 \leqslant t < T_4$)

In the three-generation world, a *leapfrogging* phenomenon could, in principle, exist. This incident occurs when adopters of generation 1 skip generation 2 altogether and adopt the latest technology instead. The additional leapfrogging parameter β should be given the following interpretation: The proportion of customers (firms or consumers) who own a first generation model of the product and who decide to upgrade is given by the parameter α_3. Given that they have made the decision to upgrade, then the proportion who upgrade to the *next* available generation (generation 2) is $1 - \beta_3$, and the proportion who leapfrogs to the *latest* available generation (generation 3) is β_3. The equations for the three-generation stage are therefore as follows:

$$dx_3/dx = \alpha_3(a_3 + \Sigma b_i x_i/N_3)(N_3 - x) + \alpha_3\beta_3(a_3' + b_3' x_3/N_3)x_1 + \alpha_3(a_3' + b_3' x_3/N_3)x_2 \quad (4)$$

$$dx_2/dt = (1 - \alpha_3)\beta_3(a_3 + \Sigma b_i x_i/N_3)(N_3 - x) - \alpha_3(a_3' + b_3' x_3/N_3)x_2 + \alpha_3(1 - \beta_3)(a_3' + b_3' x_3/N_3)x_1 \quad (5)$$

$$dx_1/dt = (1 - \alpha_3)(1 - \beta_3)(a_3 + \Sigma b_i x_i/N_3)(N_3 - x) - \alpha_3(a_3' + b_3' x_3/N_3)x_1 \quad (6)$$

In the same way, mutatis mutandis, one can construct the equations for the fourth generation ($t \geq T_4$). This is a straightforward extension of the three generation case, and therefore it is omitted from our discussion.

PROPERTIES OF THE MULTIPLE GENERATION SUBSTITUTION MODEL

We have based our model formulation on the well-known Bass diffusion model, taking into account both substitution and diffusion effects of multiple generations of the same base technology. Because equations 1 through 6 appear to differ markedly from the Bass equation, it is of interest to find out their true relationship to the latter.

We make the distinction between diffusion of the base technology and the substitution among the different generations (see Figure 1).

Whereas *substitution* is depicted in Figure 1 by system in use of each generation, the *diffusion* of the base technology is the total number of systems in use, regardless of the specific generation. In terms of our notation the diffusion of the base technology is given by $x = \Sigma_{i=1}^{4} x_i$. Summing equations 2 and 3 for the second generation we obtain:

$$dx/dt = (a_2 + (b_1 x_1 + b_2 x_2)/N_2)(N_2 - x) \quad (7)$$

In the same way, summing up the relevant equations for each generation we obtain:

$$dx/dt = (a_i + \Sigma b_j x_j/N_i)(N_i - x) \qquad T_i \leq t < T_{i+1} \quad (8)$$

If the diffusion parameters and the market potentials do not change across generations, equation 8 is precisely the Bass formulation. If some parameters do change, then equation 8 describes the *diffusion* of the base technology, whereas equations 1 through 6 describe the *substitution* between the different generations of this technology.

The general model presented earlier, however, contains more than 30 different parameters. This poses an empirical difficulty when estimating these parameters with the IBM data set that has about 60 data points. Thus we make some simplifying assumptions that reduce the number of parameters considerably. First, consistent with the previous discussion, we assume that the *diffusion* parameters basically represent the diffusion of the base technology and thus do not change from one generation to the next. This is also consistent with previous literature such as Norton and Bass (see the next section). Thus we assume that for all generations $a_i = a$, $b_i = b$, $a_i' = a'$, and $b_i' = b'$. Note that for this assumption to be valid, the base technology has to change *gradually* from one generation to the next as is the case, for example, in the diffusion of the different generations of PCs. In case of a discontinuous jump, this assumption will probably not carry through. In addition we found that when upgrading occurred firms usually upgraded to the latest available generation, and not to some intermediate generation, (i.e., we empirically verified that the parameters β are not significantly different from 1). Thus, without much loss in generality, we assume that for all generations $\beta_i = 1$.

Thus our model is logically consistent with previous models, and its main contributions lie in its ability to: (1) distinguish between the two processes of diffusion and substitution and (2) empirically and theoretically find optimal introduction times of the new generations of the base technology.

Relationship with Previous Research

As the development of our model was mainly inspired by the works of Norton and Bass [18] and Wilson and Norton [25], we discuss in this section the relationship between their works and the proposed model.

To appreciate the linkage between the model proposed in this study and the model presented by Norton and Bass [18], let N_i = market potential of generation i in terms of number of adopters; r_i = the average repeat purchase rate of an adopter; m_i = current potential for generation i ($m_i = r_i N_i$); T_2 = introduction time for generation 2; $F_i(t) = x_i(t)/N_i$ = the fraction of the total market potential that adopts generation i by time t; and $S_i(t)$ = shipments (or sales) of the i-th generation.

Given these definitions, consider the rate equations for two generations. According to Norton and Bass, the sales of the first generation are given by:

$$S_1(t) = F_1(t)m_1(1 - F_2(t - T_2)) \quad (9)$$

where $F_2(t - T_2) = 0$ for $t < T_2$.
Sales of generation 2 start from time T_2 and are given by:

$$S_2(t) = F_2(t - T_2)(m_2 + F_1(t)m_1) \quad (10)$$

Assuming the same model coefficients for both generations, $F_i(t)$ in equations 9 and 10 are assumed to be depicted by the Bass diffusion model. That is,

$$F_i(t) = (1 - e^{-(a+b)t})/(1 + (b/a)e^{-(a+b)t}) \quad (11)$$

Equations 9–11 describe the multiple generation diffusion and substitution model proposed by Norton and Bass [18]. Note that these equations describe sales of a frequently purchased (industrial) good. The sales in any period are the number of cumulative adopters multiplied by the quantity they purchase. Thus *sales, $S_i(t)$, equals cumulative adoption, multiplied by the repeat rate*. Whereas the model proposed by Norton and Bass is appropriate for repeat-purchase technological innovations, the model proposed in this article complements their work by suggesting an extension of the Bass model to study multiple generations of durable technological innovations that allows for partial leapfrogging, partial cannibalization, and clearly distinguishes between the two processes of diffusion and substitution.

Wilson and Norton [25] proposed an extension of Kalish's [8] diffusion model to study the optimal timing decision for the second generation of a product (see also [7]). In order to demonstrate why our normative guidelines for timing decisions are different from theirs, let $x_1(t)$ and $x_2(t)$ be the cumulative (percentage) adoption of the two generations of the product up to time t. Let I(t) denote the fraction of the population that is aware of the first generation up $t = T_2$ (when the second generation is introduced) and of both first and second generations after time T_2. Following Kalish [8], Wilson and Norton [25] propose the following rate equations for the two generations (for convenience, we will drop the subscript t). Before the introduction of the second generation at time T_2, awareness and adoption are given for the first generation by:

$$dI/dt = (a_1 + b_1x_1)(1 - I) \quad (12)$$

$$dx_1/dt = m_1dI/dt, \qquad x(0) = 0 \quad (13)$$

After the introduction of the second generation, information is spread by adopters of both generations,

$$dI/dt = (a_2 + b_2(x_1 + x_2))(1 - I) \quad (14)$$

and adoption is proportional to the awareness level, i.e.,

$$dx_1/dt = m_2dI/dt, \qquad x_1(T_2) = m_1I(T_2) \quad (15)$$

$$dx_2/dt = m_3dI/dt, \qquad x_2(T_2) = 0 \quad (16)$$

where m_1 is the fraction of the population that would eventually have adopted the first generation had the second generation not been introduced. Similarly, m_2 and m_3 are the fractions of the population that would eventually have adopted the first and second generations, respectively, had they been introduced simultaneously.

One of the main findings of Wilson and Norton is that, if the planning horizon is sufficiently long, it is optimal for a monopolist to introduce the two generations simultaneously, or not to introduce the second generation at all. They aptly called this the "now or never" rule.

Although in our model development, we also assume a long planning horizon (it is in fact, infinite), our optimal rule derived in the next section is a modification of their rule. We propose a "now or at maturity" policy, suggesting that the firm should either introduce the second generation as soon as it is available or delay its introduction to the maturity stage of the preceding generation. This introduction date depends on factors such as the relative size of the market potentials of the various generations, gross profit margins, the diffusion and substitution parameters, and the discount factor of the firm.

The reasons for this difference in recommended policies are as follows: first, in deriving their timing results, Wilson and Norton did not consider discounting of profits. That is, the firm in their setting did not maximize net present value but rather profits at T_2, the introduction time of the second generation. The discount factor in our analysis, however, plays a major role. In our setting, the firm maximizes net present value.

Second, Wilson and Norton assume that the unit contribution margin of the second generation (g_2) is lower than that of the first generation (g_1) i.e., $g_2 < g_1$. If the reverse occurs, simultaneous introduction is found to be optimal [25, p. 7]. The case of nondiscounting and $g_2 > g_1$ is a degenerate case in their model development (i.e., all their analyses are done under the reverse assumption that $g_2 < g_1$). The data reported for IBM in column 5 of Table 2 (average gross profit margin), however, are as follows: $g_1 = 0.6$, $g_2 = 0.78$, $g_3 = 0.84$, and $g_4 = 0.85$. These numbers suggest that in the case of IBM, $g_i > g_{i-1}$ for all generations, i.e., the gross profit margin of IBM is increasing over time. Thus, the results developed by Wilson and Norton are not applicable for the IBM data and hence, there is a need for the model proposed in this article.

Finally, the variable I(t), the fraction of informed potential adopters, is central to the results of Wilson and Norton. Such data may be difficult to obtain. Hence, there is a need to develop a parsimonious model based on sales history, such as the one proposed in this article, which is amenable to empirical testing.

TABLE 2
Average Gross Profit Margins for IBM Systems

Year	Generation		Gross profit margins: Systems only	Average	Systems, rental, and services	Average
1956	1		0.56		0.47	
1957	1		0.55		0.46	
1958	1		0.62	$g_1 = 0.6$	0.46	$g_1 = 0.48$
1959	1		0.68		0.49	
1960	2		0.73		0.51	
1961	2		0.72		0.56	
1962	2		0.77	$g_2 = 0.78$	0.56	$g_2 = 0.59$
1963	2		0.83	$g_2/g_1 = 1.3$	0.57	$g_2/g_1 = 1.21$
1964	2		0.85		0.61	
1965	3	(360)	0.84		0.62	
1966	3		0.81		0.62	
1967	3		0.83		0.59	
1968	3		0.83	$g_3 = 0.84$	0.58	$g_3 = 0.61$
1969	3		0.84		0.61	
1970	3		0.86	$g_3/g_2 = 1.07$	0.63	$g_3/g_2 = 1.04$
1971	4	(370)	0.86		0.62	
1972	4		0.85		0.61	
1973	4		0.86	$g_4 = 0.85$	0.61	$g_4 = 0.62$
1974	4		0.86	$g_4/g_3 = 1.02$	0.62	$g_4/g_3 = 1.02$

Source: Phister [20, Table 2.1.311].

Optimal Introductory Timing Strategy

The problem of optimal introduction time for each generation is best described by a two-generation case. Consider a firm that has introduced a generation of its product at time T_1 and it is currently being adopted in the marketplace. In the meantime, it has developed a second generation of the product, and this generation is available at time T_a. The firm has to decide about the introduction time T_2 of the new generation. Thus:

$$T_1 < T_a \leqslant T_2 \tag{17}$$

The left-hand side of inequality (17), i.e., $T_1 < T_a$ holds by assumption. We assume that the second generation becomes available *after* the release time of the first generation. The right-hand side of inequality equation 17, i.e., $T_a \leqslant T_2$ holds because of obvious technological constraint—the product must be made available before consumers are able to adopt and use it. Note the possibility of equality between T_a and T_2. Indeed this is the issue addressed in this section. Time T_2 is a choice variable of the firm. Should the firm introduce the product as soon as it is available (set $T_a = T_2$) thereby foregoing sales and profits of the old model, or should it wait, delay receipt of cash flows from sales of the new generation, but prolong the flow of profits from the old generation?

Let u(t) be defined as follows:

$$u(t) = \begin{cases} 0 & \text{for} \quad 0 \leqslant t \leqslant T_2 \\ 1 & \text{for} \quad T_2 < t \end{cases} \tag{18}$$

Thus u(t) is a dummy variable that can be used to summarize equations 1–3 by the following two equations (with the simplifying assumptions made earlier):

$$dx_2/dt = u\alpha_2(a + bx/N_2)(N_2 - x) + u\alpha_2(a' + b'x_2/N_2)x_1 \tag{19}$$

$$dx_1/dt = (1 - u)(a + bx/N_1)(N_1 - x) + u(1 - \alpha_2)(a + bx/N_2)(N_2 - x) - u\alpha_2(a' + b'x_2/N_2)x_1 \tag{20}$$

Let g_i be the gross profit margin of generation i. The net present value of the venture, denoted by π is given by:

$$\pi = \int_0^\infty [g_1(1 - u)(a + bx/N_1)(N_1 - x) + u(g_2\alpha_2 + (1 - \alpha_2)g_1)(a + bx/N_2)(N_2 - x) + g_2u\alpha_2(a' + b'x_2/N_2)x_1]e^{-rt}dt \tag{21}$$

Let $\lambda(t)$ and $\mu(t)$ be the current multipliers of x_1 and x_2 respectively. The equations for the current Hamiltonian (H) and the current multipliers are as follows:

$$H = (1 - u)(g_1 + \lambda)(a + bx/N_1)(N_1 - x) + u(\alpha_2(g_2 + \mu) + (1 - \alpha_2)(g_1 + \lambda))(a + bx/N_2)(N_2 - x) + u\alpha_2(g_2 + \mu - \lambda)(a' + b'x_2/N_2)x_1 \tag{22}$$

$$\begin{aligned} d\lambda/dt - r\lambda &= -\partial H/\partial x_1 \\ &= -u\alpha_2(g_2 + \mu - \lambda)(a' + b'x_2/N_2) - u(\alpha_2(g_2 + \mu) \\ &\quad + (1 - \alpha_2)(g_1 + \lambda))(b - a - 2bx/N_2) \\ &\quad - (g_1 + \lambda)(1 - u)(b - a - 2bx/N_1) \end{aligned} \tag{23}$$

$$\begin{aligned} d\mu/dt - r\mu &= -\partial H/\partial x_2 \\ &= -b'u\alpha_2(g_2 + \mu - \lambda)x_1/N_2 - u(\alpha_2(g_2 + \mu) + (1 - \alpha_2)(g_1 + \lambda)) \\ &\quad \times (b - a - 2bx/N_2) \end{aligned} \tag{24}$$

Performing the differentiation of the Hamiltonian with respect to u, and evaluating the expression at time $t = 0$ we have the following proposition:

Proposition 1: If either the market potential of the second generation is large as compared to that of the first generation or the (current and future) profits gained from the second generation are larger than that of the first generation, the firm introduces the second generation as soon as possible ("now"). Otherwise the firm should shelve the new generation and introduce it at a later date.

Differentiating the Hamiltonian with respect to u and equating to zero, results in the following equation for the optimal introduction time (when it is not zero):

$$\begin{aligned} (g_1 + \lambda)(a + bx/N_1)(N_1 - x) = {} & (\alpha_2(g_2 + \mu) + (1 - \alpha_2)(g_1 + \lambda)) \\ & \times (a + bx/N_2)(N_2 - x) \\ & + \alpha_2(g_2 + \mu - \lambda)(a' + b'x_2/N_2)x_1 \end{aligned} \tag{25}$$

The interpretation of this optimality condition is as follows: the left-hand side of equation 25 represents the unit sales of generation 1 multiplied by the (current and future) profits per unit. These revenues have to be compared with the revenues from releasing the second generation (the right-hand side of equation 25) that consists of the unit sales from new adopters multiplied by the average margin per unit, as some of these adoptions are of the older generation, plus the sales that are due to upgraders. Because an upgrader is lost as a future customer of the previous generation, these sales are multiplied by the *net gain* in unit margin. As long as the profits from the first generation are greater, i.e., the left-hand side of equation 25 is larger than the right-hand side, the firm shelves the new generation. It releases the new generation at the time of equality in equation 25.

An easier interpretation of the condition, and a clue to when the second generation should be introduced (if not "now") can be found in the following numerical analysis.

We took a specific set of parameters to act as a "base case" for the analysis; then we systematically changed the value of the parameters, one at a time, and observed the resultant profits (net present value) as a function of the entry time of the second generation.

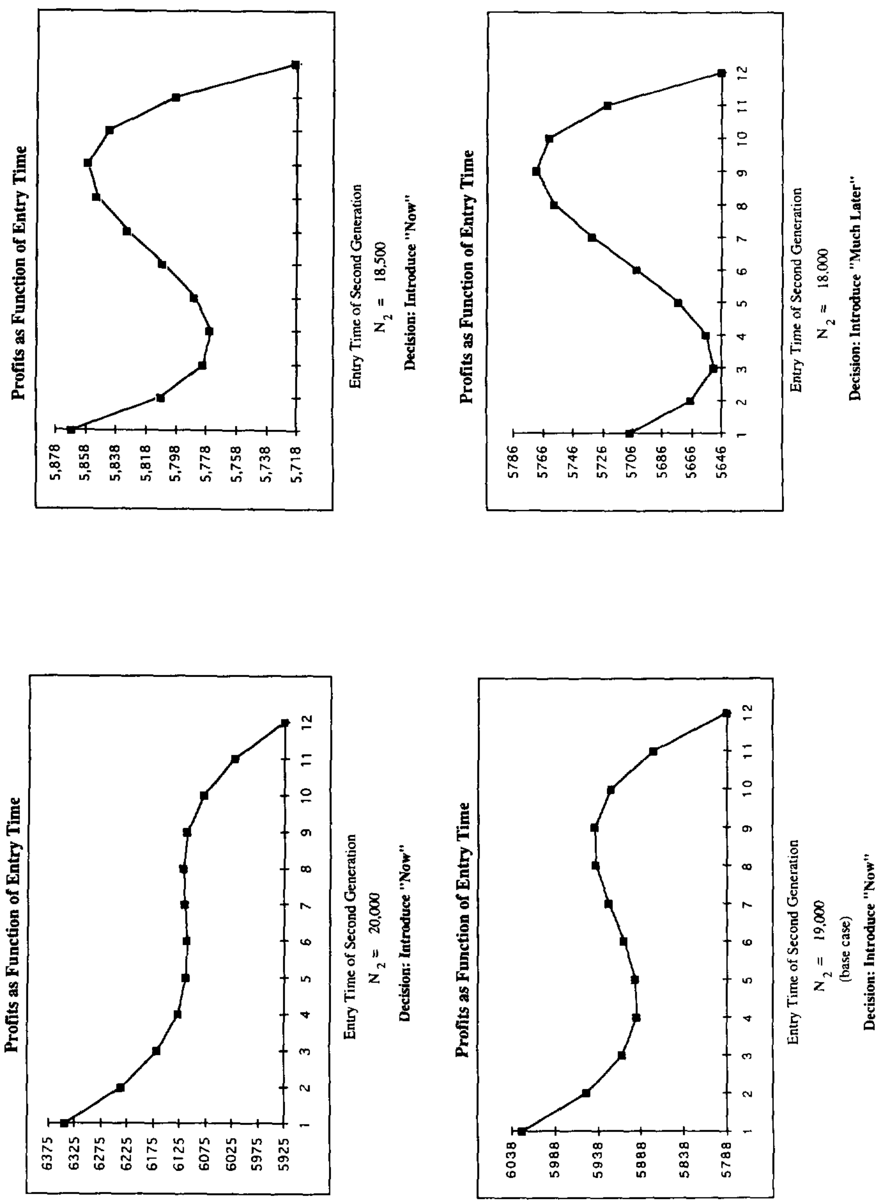

Fig. 3. Impact of entry time on profits using "now or much later" rule.

In addition, in order to minimize the possibility that the results depended on the specific base case, we have performed large number of runs in which the parameters were chosen with a random number generator (with some specific restrictions). The complete description is given in Appendix 1. The base case is as follows (recall that a and a′ are the coefficient of external influence of the nonadopters and adopters, b and b′ are the coefficients of internal influence of the nonadopters and adopters, respectively, g_1 and g_2 are the gross margins of the first and second generations, N_1 and N_2 are the market potential of the first and second generations, α_2 is the cannibalization parameter, and r is the cost of capital of the firm):

$$a = 0.02 \quad b = 0.30 \quad g_1 = 0.75 \quad N_1 = 15{,}000 \quad \alpha_2 = 0.80$$
$$a' = 0.04 \quad b' = 0.45 \quad g_2 = 0.80 \quad N_2 = 19{,}000 \quad r = 0.12$$

Figure 3 reveals the general pattern found. In Figure 3, the market potential of the second generation (N_2) is gradually decreased from 20,000 to 18,000. The profit function curve, including the maturity point by developing reasonable estimates of the diffusion parameters (see [16]).

IBM Mainframe Computers: Empirical Evidence

PARAMETER ESTIMATION

The process to be estimated is described by the equations in Appendix 2. These equations are a result of the restrictions on the parameters of our initial model. The process was estimated using the discrete regression analogy for these equations, and the nonlinear regression package of SPSS, PC version. This procedure yielded the following values (asymptotic standard errors are in parenthesis, * denotes significance at the 95% level, and ** denotes significance at the 90% level):

$a = -0.023$ (0.075)	$b = 0.600$ (0.196)*	$a' = 0.319$ (0.129)*	$b' = 0.425$ (0.187)*
$N_1 = 3150$ (2010)**	$N_2 = 17,641$ (4343)*	$N_3 = 21,419$ (956)*	$N_4 = 17,646$ (709)*
$\alpha_2 = 0.904$ (0.139)*	$\alpha_3 = 0.598$ (0.166)*	$\alpha_4 = 0.345$ (0.144)*	

and R^2 is 0.82.

Note that all the parameters are significant except the coefficient of external influence of the nonadopters (a). Also note that the market potentials are increasing, except for the fourth one, which declines by about 18% as compared to the market potential of the third generation. In addition, the cannibalization parameter (α) is declining with each generation. The parameter measures, in each generation, the percentage of new buyers who adopt the newest generation, as opposed to buying an older generation. Thus, as the years and generations progressed, more and more adopters were adopting an older generation *when a new one was already available*. Lastly note that $b > b'$ which implies that the internal influence is more effective in convincing new adopters to purchase the latest generation than it is in convincing adopters of previous generations to upgrade.

As mentioned in the introduction, there are two processes that work simultaneously: diffusion of the base technology and substitution between successive generations of this technology. The first process is captured by equation 8, whereas the latter by the equations in Appendix 2.

Diffusion

Once the parameters a, b, and N_i are estimated, equation 8 can be used to generate the estimate of the diffusion of the base technology.

The actual and fitted values are plotted in Figure 4. Note that the decline in the total number of IBM systems-in-use is well captured by the model. It should be pointed out that such a decline cannot be captured by an aggregate diffusion model such as the Bass model, because the Bass model always predicts an increase in the cumulative number of adopters. As the multiple generation model explicitly takes into consideration substitution across generations, it can capture such a phenomenon in the cumulative adoption pattern of the product. As is clear from Figure 4, the model describes the growth of IBM mainframe systems well.

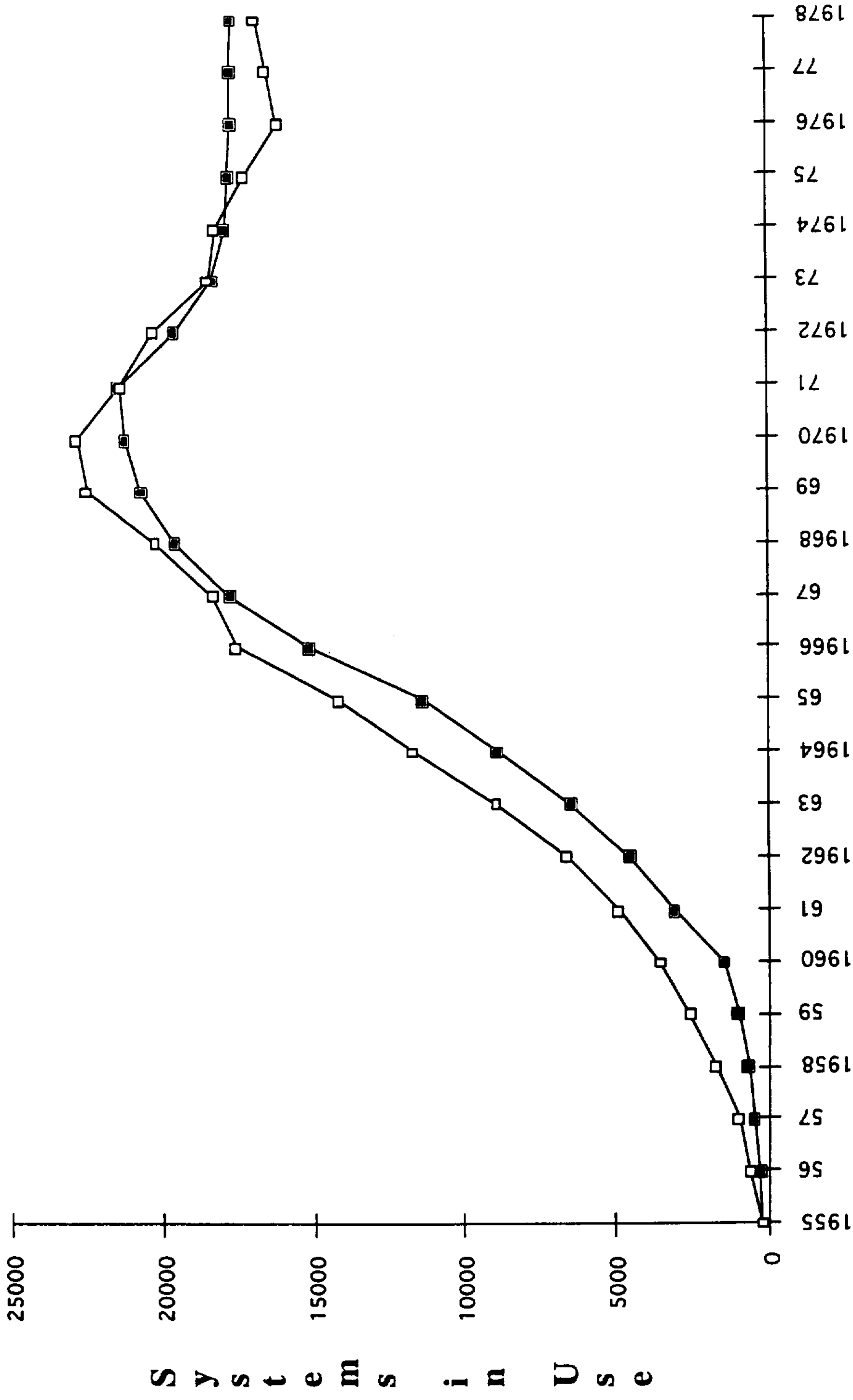

Fig. 4. IBM mainframes: total systems—observed vs. expected. pred x (■); obs x (□).

Substitution

Because the parameters a′, b′, and α_i are given, the equations in Appendix 2 are used to estimate the substitution between the different generations of IBM mainframes. The actual and fitted values for the four generations are plotted in Figure 5. These charts indicate that the model equations describe the growth and decline across generations well.

PREDICTIONS

One of the criticisms raised in using a Bass-type diffusion model is that it requires several data points to estimate the model parameters. In many cases, including ours, the product in question is a new version or generation of the same base technology. Data for previous generations of this technology, however, exist and can compensate for the lack of data points. In effect we use previous generation's data in order to impute the parameters of future generations' diffusion. This can be thought of as a special case of Lawrence and Lawton [20], who suggest using analogy as a way of determining diffusion parameters.

To demonstrate this point we estimated the Bass model (equation 1) using 6 years of data of the 370 family (1971 to 1976). The results are:

$a_4 = 0.17 \qquad b_4 = 0.97 \qquad N_4 = 9161 \qquad R^2 = 0.99$

(all parameters are significant at the 95% level).

We then used the parameters to predict sales of the 370 family for 1977 and 1978. Note that since the 370 family at that time was the most advanced generation, sales and change in systems-in-use coincide. These figures are reported in Table 3, second column. Next we performed our previous analyses twice: (1) we used data points for the previous three generations and only 4 years of the 370 family (1971–1973), (2) we used data points for the previous 3 generations and 6 years of the 370 family (1971–1976). These results are reported in Table 3, columns 3 and 4, respectively. All parameters were significant and closely approximated the parameter values obtained from the full data set reported earlier (of the differences in the parameter values between the full data set and the current ones, only α_3 and α_4 were significant at the 95% level). As compared to the actual data, our model's predictions were better when based on 6 years of data, as expected. Furthermore, it also did better than the benchmark Bass model when it was based on 4 years of the 370 family. Both sales and market potential are closer to the actual data. We therefore conclude that our approach can somewhat alleviate the problem of lack of the sufficient data points.

ANALYSIS OF TIMING DECISIONS

In this section we use the parameter estimates reported earlier for the IBM mainframes to assess the optimality of the new generation timing decisions for IBM.

From the earlier section, it is clear that in addition to the parameters reported, two more sets of parameters have to be estimated: gross profit margin and the cost of capital of IBM.

Let g_i denote the gross profit margin of the i-th generation. Two proxies for this parameter are shown in Table 2. The estimates based on sales (without rental and services) yield the following values: $g_1 = 0.6$, $g_2 = 0.78$, $g_3 = 0.84$, and $g_4 = 0.85$. The fact that the gross profit margin is increasing over time is very noticeable, and is probably due to increasing returns to scale, economies of scope, and experience effects (see [3, chapter 3; 4, chapters 3, 4]). Although both use the same source (Phister [20, Table 2.1.311]), the first profit margin is calculated for systems only, whereas the second combines systems sales, systems rental, and services. Although the gross profit margins are evidently not

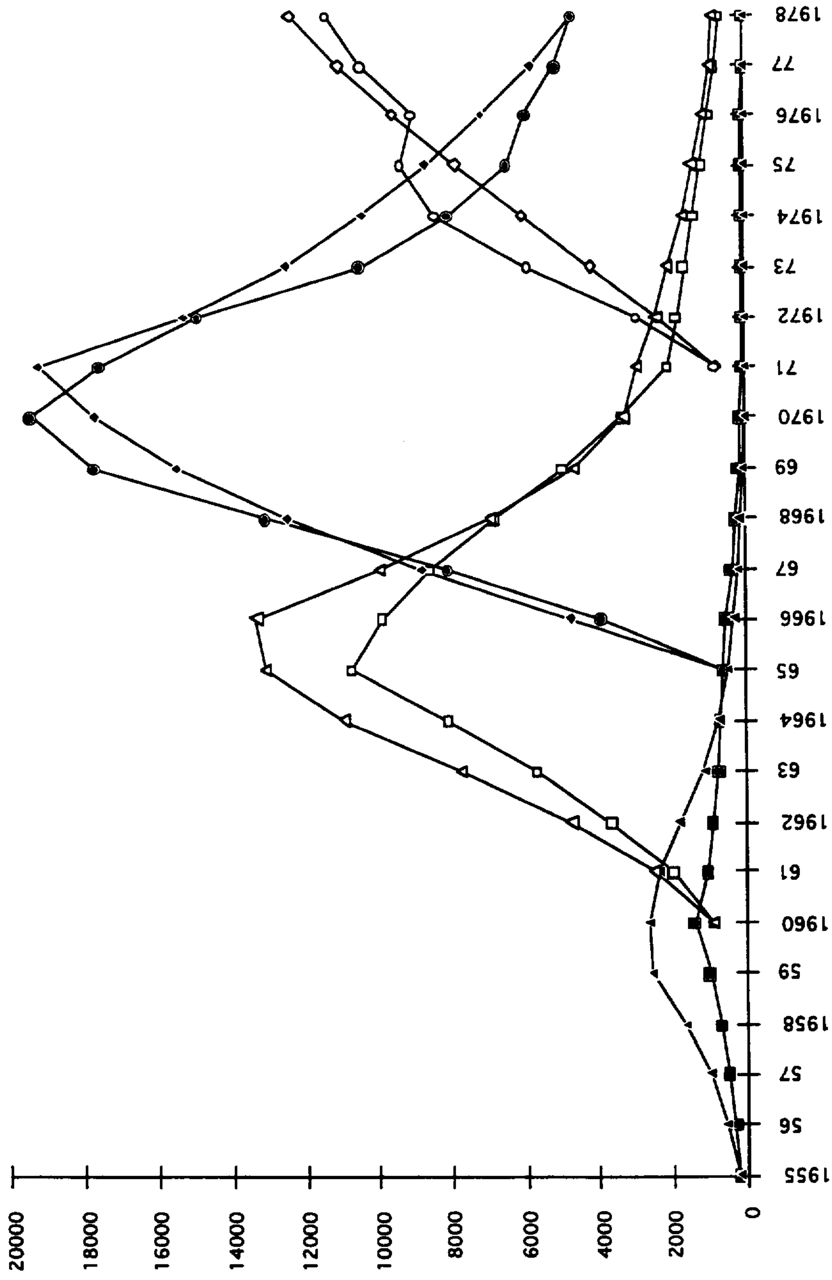

Fig. 5. IBM mainframes: observed vs. expected. pred x1 (■); pred x2 (□); pred x3 (◆); pred x4 (◇); obs x1 (▲); obs x2 (△); obs x3 (●); obs x4 (○).

TABLE 3
Prediction of Market Potential and Sales of 370 Family

	Predictions based on the Bass model using 6 years of data of the 370 family	Predictions based on the proposed model using first 3 generations and 4 years of data of the 370 family	Predictions based on the proposed model using first 3 generations and 6 years of data of the 370 family	Actual data
1977 Sales	130	1873	1768	1404
1978 Sales	NA[a]	1153	1507	898
Market potential	9161	18,936	17,879	16,344
Mean absolute deviations (MAD)	3118	1105	836	–

[a] The Bass model yields a negative entry. The MAD computation for this model assumes a value of zero. The relatively poor performance of the Bass model is due to the fact that it yields an estimate for the market potential that is far too low.

the same in the two calculations, the ratios of g_i/g_{i-1} seem very similar. We report here the figures calculated on the basis of systems only. We also derived results based on the second method of calculating the gross profit margins (sales, rental, and services). The two methods yielded very similar results.

With respect to the estimation of the cost of capital of IBM at the introduction time, we use the standard tax-adjusted cost of capital that depends on the risk-free rate, the average market risk premium, the volatility coefficient, and the debt-to-equity ratio of IBM. Different estimates would be obtained, however, depending on the method used to estimate the previous parameters (e.g., the use of 3 months vs. 1 year T-bills to estimate the risk free rate, or the use of market value of IBM's debt to find the debt-to-equity ratio). Therefore, a rough estimate for the possible *range* of the cost of capital was obtained and was found to be 12% to 20%. A strong support for our calculation can be found in Levy and Sarnat [11, chapter 18] who calculated the cost of capital of IBM to be 15.4% in 1972 and 16.0% in 1973. We, therefore, computed the optimal time using different costs of capital covering this range.

We used the parameter estimates together with the estimates of the gross profit margins and the cost of capital to obtain introductory times T_2, T_3, and T_4 that yielded the greatest net present value, given these parameters.

Proposition 1 established the fact that there are only two categories of introduction times: either a "corner solution" where the new generation is introduced as soon as it is available or an "interior solution" where the new generation is introduced in a later stage of the preceding generation.

The figures in Table 1 are systems in use, and from these figures the sales patterns of the various generations can be obtained. From these patterns one can observe that the introduction time of the second generation T_2 occurred *at the peak* of sales of the first generation whereas T_3 and T_4 occurred *after the peak* of their respective generations. What we observe from our analysis, however, are the following three points:

1. T_2, the introduction time of the second generation falls into the category of "introduce now."
2. Introduction times T_3 and T_4 fall into the category of "introduce at maturity."
3. IBM introduced the 360 and 370 families too late.

TABLE 4
Actual and Optimal Introduction Dates

Cost of capital	Optimal time of introduction	
	T_3 (360 family)	T_4 (370 family)
$12 \leqslant r \leqslant 18$	1964	1970
$19 \leqslant r \leqslant 20$	1963	1969
Actual introduction date	1965	1971

Specifically from our empirical and numerical analysis, it follows that the optimal introductory times, as functions of the cost of capital are as follows (recall that the actual times of introduction are $T_3 = 1965$ and $T_4 = 1971$) (Table 4).

Thus for most values for cost of capital considered in the analysis, based on our estimates and empirical analysis, IBM was late in its introduction of the 360 and 370 families.

Time to market is defined as the time elapsed between making the decision to start product development and introduction of the product into the market place [2, 12, 21]. From a marketing point of view, we can thus conclude that the time to market of the 360 and 370 families of mainframe computers was too long, i.e., IBM should have shortened the product development time and introduced the 360 and 370 families sooner.

It is difficult to show direct empirical evidence that IBM was delaying product introduction deliberately. Nonetheless some indirect evidence exists. First, as we implied earlier, IBM should have shortened the product development time on its 360 and 370 families. In the case of the 360 family of mainframes, the following information is brought by Brock [3, pp. 192–193] in order to demonstrate the slow innovative pace of IBM: "The first version of the integrated circuit was developed without government funds by Texas Instruments and demonstrated to the Air Force in 1958. . . . Texas Instruments continued work on integrated circuits and built a small demonstration computer out of them in late 1961. . . . In April 1964, IBM announced the System 360 with hybrid circuits, an intermediate step between discrete components and integrated circuits. . . . IBM delivered the first hybrid System 360 in April 1965." Indeed, several more examples such as this are given in the same chapter, and Brock concludes that "IBM's contribution to technical progress has not been as large relative to industry advances as its market share would lead one to expect."

Furthermore, with respect to intentional delaying of new products, Brock gives the following examples [3, p. 103]: "In 1969, the management committee determined that a delay of announcement of NS (the System 370) from 3 to 6 months could yield as much as \$30 to \$50 million of additional Revenue. . . ." In discussion of IBM's "System Q" software plans to build a comprehensive new operating system for announcements in 1974 and delivery in 1975, it was determined that the system would be postponed if "the current generation systems are doing so well in the field that there is no compelling business reason for revealing our intentions."

Discussion and Conclusions

This article has proposed a model that simultaneously captures the adoption and substitution pattern for each successive generation of a durable technological innovation. Normative guidelines based on the model suggest that a firm should either introduce a new generation as soon as it is available or else delay its introduction until the maturity stage in the life-cycle of the current generation. The application of the model and of the associated normative results to the four generations of the IBM mainframe systems suggests that IBM introduced the two successive generations, 360 family and 370 family, too late, i.e., their time to market should have been shorter.

In applying the proposed model to the IBM data we assumed that (1) each potential adopter adopted exactly one unit of a given generation and (2) the various generations were introduced in a monopolistic market structure.

The first assumption (i.e., adoption of one unit per adopter) also implies that when an IBM system was replaced by a more advanced IBM generation, the adopter again bought one unit. However, we only partially considered the fact that generations were different in their computing power. It could be argued that because each successive generation of mainframe computers could solve more problems at a greater speed, and could store and retrieve information faster and more efficiently, the adopters of two or more units could replace them with a single unit of the new generation. That is, in order to examine diffusion over time we need to develop "equivalent" units of IBM systems adjusted for their computing power and for other characteristics.

This certainly is a valid criticism. Our contention, however, is that advanced and improved characteristics of a new generation affect the timing decisions of potential adopters. That is, adopters' expectations about the performance of a new generation affect their decisions of when and if to adopt it. For example, in buying a new model of a CD player with advanced features, one does not replace the old CD with one-half of the new one. That is, consistent with the diffusion research, the replacement of the old model by a single unit of the new model should be counted as one unit of adoption irrespective of the differences in features of the two generations. The availability of advanced features certainly may affect a potential adopter's decision of when and whether to adopt the new model. But when adopted, it counted as a single adoption regardless of its advanced features.

With respect to the second assumption underlying our application to the IBM data, we have assumed that the various generations were introduced in a monopolistic market structure. Whereas in many markets optimal timing may well be a function of competition and order of entry effects, these aspects were not considered in this study. IBM's share of the mainframe market during the years of analysis ranged from about 60% to 80%. Although there was a decline in IBM's market share in the 1950's, IBM maintained a consistent and high market share during the 1960s and 1970s. In the last 10 years of our analysis, IBM's share of the number of mainframe systems-in-use rose from 61% in 1968 to more than 66% in 1978. IBM dominated the market during the years included in our analysis. We, however, have neither formulated nor tested a model under an oligopolistic market structure. Although certainly a very fruitful opportunity for further research, we believe that this limitation does not undermine the major findings of our application. Our analyses reported indicate that IBM was late in introducing the 360 and 370 families. Because pressure from competitors in an oligopolistic market clearly results in new product introductions earlier than in a monopolistic market, our timing results are only strengthened by the fact that we considered IBM as operating in a monopolistic market. If the firm is deemed late under this scenario, it will certainly be deemed late as an oligopolist. In fact, marketing, production, and managerial problems associated with the introduction of the 360 family are highlighted in a recent book by Thomas J. Watson Jr. [24]. He called overcoming these introductory problems and introducing the 360 family "the greatest triumph of my business career."

The authors thank Nirit Avnimelech, Anne Coughlan, Jehoshua Eliashberg, Moshe Givon, Morton Kamien, and an anonymous reviewer for their helpful comments and suggestions. The authors also appreciated feedback from Rosanne Cole, W. L. Nichols, and P. D. Van Waes, all of IBM, and Douglas Ziemer of Texas Instruments on earlier versions of this article.

References

1. Bass, Frank M., A New Product Growth Model for Consumer Durables, *Management Science* 15 (January), 215-227 (1969).
2. Bayus, Barry L., High Definition Television: Assessing Demand Forecasts for Next Generation Consumer Durable, *Management Science* 39 (November), 1319-1333 (1993).
3. Brock, Gerald W., *The U.S. Computer Industry: A Study of Market Power*, Ballinger, Cambridge, MA, 1975.
4. Dorfman, Nancy S., *Innovation and Market Structure: Lessons from the Computer and Semiconductor Industries*, Ballinger, Cambridge, MA, 1987.
5. Ferguson, Charles H., and Morris, Charles R., *Computer Wars*, Times Books, New York, 1993.
6. Fisher, J. C., and Pry, R. H., A Simple Substitution Model of Technological Change, *Technological Forecasting and Social Change* 3, 75-88 (1971).
7. Jain, Dipak, and Rao, Ram. Effects of Price on the Demand for Durables: Modeling, Estimation, and Findings, *Journal of Business and Economic Statistics* 2 (April), 163-170 (1990).
8. Kalish, Shlomo, A New Product Adoption Model with Price, Advertising, and Uncertainty, *Management Science* 31 (December, 1569-1585 (1985).
9. Katz, Barbara, and Phillips, Almarin, The Computer Industry, in *Government and Technical Process*. Richard Nelson, ed., Pergamon, New York, 1982.
10. Lawrence, K. D., and Lawton, W. H. Applications of Diffusion Models: Some Empirical Results, in *New Product Forecasting*. Y. Wind, V. Mahajan, and R. C. Cardozo, eds., MA: Lexington Books, Lexington, MA, 1981.
11. Levy, Haim, and Sarnat, Marshall, *Capital Investment and Financial Decisions*, Prentice Hall, Englewood Cliffs, NJ, 1982.
12. Lilien, Gary L., and Yoon, Eunsang, The Timing of Competitive Market Entry: An Exploratory Study of New Industrial Products, *Management Science* 36 (May), 568-585 (1990).
13. Linstone, Harold A., and Sahal, Devendra, *Technological Substitution*, Elsevier, New York, 1976.
14. Mahajan, Vijay, Mason, Charlotte H., and Srinivasan, V., An Evaluation of Estimation Procedures for New Product Diffusion Models, in *Innovation Diffusion Models of New Product Introduction*. Vijay Mahajan and Yoram Wind, eds., Ballinger, Cambridge, MA, 1986.
15. Mahajan, Vijay, Muller, Eitan, and Bass, Frank M. New Product Diffusion Models in Marketing: A Review and Directions for Research, *Journal of Marketing* 54 (January), 1-26 (1990).
16. Mahajan, Vijay, and Sharma, Subhash, A Simple Algebraic Estimation Procedure for Innovation Diffusion Models of New Product Acceptance, *Technological Forecasting and Social Change* 30, 331-345 (1986).
17. Modis, Theodore, Technological Substitution in the Computer Industry, *Technological Forecasting and Social Change* 43, 157-167 (1993).
18. Norton, John A., and Bass, Frank M. A Diffusion Theory Model of Adoption and Substitution for Successive Generations of High Technology Products, *Management Science* 33 (September), 1069-1086 (1987).
19. Norton, John A., and Bass, Frank M., Evolution of Technological Generations: The Law of Capture, *Sloan Management Review* 33, 66-77 (1992).
20. Phister, Montgomery, *Data Processing, Technology, and Economics*, Santa Monica, Santa Monica, CA, 1976.
21. Putsis, Jr., William P., Why Put Off Until Tomorrow What You Can Do Today: Incentives and the Timing of New Product Introduction, *Journal of Product Innovation Management* 10, 195-203 (1993).
22. Sultan, Fareena, Farley, John U., and Lehman, Donald, A Meta Analysis of Diffusion Models, *Journal of Marketing Research*, 25, 70-77 (1990).
23. Urban, Glen L., and Hauser, John R., *Design and Marketing of New Products*, Prentice-Hall, Englewood Cliffs, NJ, 1993.
24. Watson, Jr., Thomas J., and Petre, Peter, *Father, Son & Co.: My Life at IBM and Beyond*, Bantam, New York, 1990.
25. Wilson, Lynn O., and Norton, John A., Optimal Entry Time for a Product Line Extension, *Marketing Science* 8 (Winter), 1-17 (1989).

Received 18 April 1995; accepted 4 August 1995

Appendix 1

In this appendix we describe the details of our numerical analysis. The base case we used was as follows:

$a = 0.02$	$b = 0.30$	$g_1 = 0.75$	$N_1 = 15{,}000$	$\alpha_2 = 0.80$
$a' = 0.04$	$b' = 0.45$	$g_2 = 0.80$	$N_2 = 19{,}000$	$tr = 0.12$

The analysis was performed in two stages: In both stages we set the following restrictions on the parameter values:

$0.01 \leqslant a \leqslant 0.4$	$0.1 \leqslant b \leqslant 0.6$	$0 \leqslant g_i \leqslant 1$
$0.01 \leqslant a' \leqslant 0.4$	$0.1 \leqslant b' \leqslant 0.6$	$0.1 \leqslant \alpha_2 \leqslant 1$
$10{,}000 \leqslant N_i \leqslant 30{,}000$		$5\% \leqslant r \leqslant 15\%$

In the first stage, we divided the previous restrictions for each parameter into 10 (or 11) equal intervals except for N_i (21 intervals) and r (26 intervals). We then systematically changed the value of the parameters of the base case, one parameter at a time. For every value of each parameter, we ran the numerical analysis to discover the optimal time and the shape of the NPV as a funcition of the time of entry. Altogether 140 runs were performed. In every run and for every parameter (except for the parameter "a") we found the phenomenon reported in the text, namely that as we changed one of the parameters continuously, a local maximum began to develop around the eighth, ninth, or tenth period. As long as the profits of entering in the first period exceeded this local "hump," the optimal decision was to introduce the second generation as soon as possible. As soon as this local maximum surpassed the profits of entering right away, i.e., it became a global maximum, the optimal entry time made a quantum leap to the eighth, ninth, or tenth period. The one exception was the parameter "a" that when increased continuously to large values (such as 0.3 or 0.4), it caused the optimal entry time to decrease *continuously* (to $T_2 = 3$). However, we know of no empirical study that has reported such large values of this parameter (see, for example, [14, 22]).

In the second stage of this empirical study, in order to minimize dependence on the values of the base case, we performed 100 runs in which the parameters were chosen at random, using the @RAND function of the Lotus 123 spreadsheet (on each of the parameters separately). In order to incorporate the fact that the new generation is perceived as an improvement by the consumers (or at least not seen as a regression), we added the restriction that the second market potential is at least as large as the first one. The results of the 100 runs lend support to our "now or at maturity" proposition:

Strategy	*Frequency*
Introduce now ($T_2 = 1$)	34
Somewhat delay ($2 \leqslant T_2 \leqslant 7$)	9
Introduce at maturity ($8 \leqslant T_2$)	57

Therefore, based on these results we conclude that the optimal policy of the firm is either to introduce the new generation as soon as possible or delay its introduction to a much later phase in the life-cycle of the first generation.

Appendix 2

In this appendix we specify the equations we used in the empirical analysis. All variables and parameters conform to the definitions given in the second section.

Period of First Generation ($T_1 \leqslant t < T_2$):

$$dx_1/dt = (a + bx/N_1)(N_1 - x)$$

Period of Second Generation ($T_2 \leqslant t < T_3$):

$$dx_2/dt = \alpha_2 (a + bx/N_2)(N_2 - x) + \alpha_2 (a' + b'x_2/N_2)x_1$$

$$dx_1/dt = (1 - \alpha_2)(a + bx/N_2)(N_2 - x) - \alpha_2 (a' + b'x_2/N_2)x_1$$

Period of Third Generation ($T_3 \leqslant t < T_4$):

$$dx_3/dt = \alpha_3 (a + bx/N_3)(N_3 - x) + \alpha_3 (a' + b'x_3/N_3)(x_1 + x_2)$$

$$dx_2/dt = (1 - \alpha_3)(a + bx/N_3)(N_3 - x) - \alpha_3 (a' + b'x_3/N_3)x_2$$

$$dx_1/dt = -\alpha_3 (a' + b'x_3/N_3)x_1$$

Period of Fourth Generation ($T_4 \leqslant t$):

$$dx_4/dt = \alpha_4 (a + bx/N_4)(N_4 - x) + \alpha_4 (a' + b'x_4/N_4)(x_1 + x_2 + x_3)$$

$$dx_3/dt = (1 - \alpha_4)(a + bx/N_4)(N_4 - x) - \alpha_4 (a' + b'x_4/N_4)x_3$$

$$dx_2/dt = -\alpha_4 (a' + b'x_4/N_4)x_2$$

$$dx_1/dt = -\alpha_4 (a' + b'x_4/N_4)x_1$$