DIFFUSION OF INNOVATIONS UNDER SUPPLY CONSTRAINTS

SUNIL KUMAR

Graduate School of Business, Stanford University, Stanford, California 94305-5015, skumar@stanford.edu

JAYASHANKAR M. SWAMINATHAN

The Kenan-Flagler Business School, University of North Carolina, Chapel Hill, North Carolina 27599-3490, msj@unc.edu

In this paper we present a canonical setting that illustrates the need for explicitly modeling interactions between manufacturing and marketing/sales decisions in a firm. We consider a firm that sells an innovative product with a given market potential. The firm may not be able to meet demand due to capacity constraints. For such firms, we present a new model of demand, modified from the original model of Bass, to capture the effect of unmet past demand on future demand. We use this model to find production and sales plans that maximize profit during the lifetime of the product in a firm with a fixed production capacity. We conduct an extensive numerical study to establish that the trivial, myopic sales plan that sells the maximal amount possible at each time instant is not necessarily optimal. We show that a heuristic "build-up" policy, in which the firm does not sell at all for a period of time and builds up enough inventory to never lose sales once it begins selling, is a robust approximation to the optimal policy. Specializing to a lost-sales setting, we prove that the optimal sales plan is indeed of the "build-up" type. We explicitly characterize the optimal build-up period and analytically derive the optimal initial inventory and roll-out delay. Finally, we show that the insights obtained in the fixed capacity case continue to hold when the firm is able to dynamically change capacity.

(*Production/scheduling, planning: optimal sales plans. Marketing, new products: modifying Bass model for supply constraints. Dynamic programming/optimal control, application: application of maximum principle.*)

Received September 2000; revisions received September 2001, May 2002, December 2002; accepted December 2002.

Area of review: Manufacturing, Service, and Supply Chain Operations.

1. INTRODUCTION

We consider situations in which the rapid growth of demand for a product due to positive word of mouth spreading from past sales outstrips the firm's capacity to produce and supply the product. Examples include Apple's PowerMac G4 (*New York Post* 1999) where Motorola (the supplier of the G4 chips) was overwhelmed by the rapid growth of demand for the popular computer. In order to avoid such problems, companies may delay rolling out the product to build-up initial inventory. Examples include Sony delaying roll-out until it had one million units at launch of its new Playstation 2 (*Agence France Presse* 1999) and ID Four Ltd.'s increase of initial inventory of its software product in anticipation of rapid demand growth under supply constraints (*Business Wire* 1999). Such situations provide the motivation for modeling marketing-manufacturing interactions, in general, and the impact of supply constraints in the presence of word-of-mouth effects, in particular. In this paper, we study a single firm that sells an innovative product with a fixed market potential. The firm may be unable to supply the product at times due to capacity constraints. The objective of the firm is to maximize total profit during the lifetime of the product. We are interested in resolving when it is beneficial for the firm to delay roll-out and build-up inventory.

Among others, Bass (1969) proposed a diffusion model for adoption of new products by customers. In that model (and many subsequent models) the adoption of an innovative product by customers is driven by two sources: (a) direct communication from manufacturers or advertisers, and (b) word of mouth spreading from previous users. In a setting where the product is made by a monopolist, the demand as seen by the firm would be the same as the number of adoptions of the product by customers. In the Bass model, the instantaneous demand rate up to time t is given by $n(t) = [p + (q/m)N(t)](m - N(t))$, where $N(t) = \int_0^t n(s)\,ds$ is the cumulative demand up to time t, m is the market potential, and p, q are constants (positive and between 0 and 1) that represent the relative effects of mass media and word of mouth on the population.

A shortcoming of the Bass model is its inability to capture supply constraints. While supply constraints are less relevant when considering adoption of a product category by customers (the primary motivation for the Bass model), they are quite relevant in our setting, where the attempted adoption of the product immediately translates into demand on the sole producer. In this case, the assumption that everyone who attempted to purchase the product in the past would have been successful and, as a result, would spread positive word of mouth about the product, need not be valid. Under supply constraints, it is possible that when some customers attempt to purchase the product, the product may be unavailable. Thus, the cumulative sales up to time t, $S(t)$, may be quite different from the cumulative demand, $N(t)$. In this setting, one needs to take into account the fact that customers who have tried to buy but have been unsuccessful have no motivation to spread positive word of mouth about the product. We propose a modification of the Bass model that takes into account supply constraints and the resulting unmet demand which may be either backlogged or lost. In proposing this modification, we strive to retain one very desirable feature of the Bass model: a parsimonious mathematical description. In our model, the instantaneous demand at time t is given by $n(t) = [p + (q/m)S(t)](m - N(t))$, where the cumulative sales $S(t)$ satisfies the supply constraint. The key feature of our model is that future demand depends not only on past demand but also on past realized sales.

Using this model of demand, we begin by describing an optimal control problem that determines the production and sales plans that maximize the discounted profits over the product lifetime for a firm with fixed production capacity. The setup is general enough to allow for any degree of backlogging. That is, some fraction of customers who have asked for, but were not sold, a product join and continue to stay in a backlog queue. This fraction may vary from 0 (representing the case of lost sales) to 1 (representing the case of complete backlogging). Utilizing Pontryagin's maximum principle, we prove that the optimal sales plan either sells as much as possible or nothing at all at any given instant of time, in this very general model. Since explicitly characterizing the optimal policy is not analytically tractable, we numerically solve the control problem, by discretization, for a wide variety of parameter choices (resulting in nearly 450 distinct cases). This extensive numerical study establishes that the notion of a "sales plan" is not vacuous in this general setting. That is, one does not merely sell whatever is available. Rather, one sells according to a careful prescription. Our numerical study indicates that a heuristic policy which builds up inventory by not selling at all over an initial period and then switches to selling as much as possible (hereafter referred to as the *build-up* policy) is robust and very close to optimal (an average error of 0.3% over 432 experiments). Further, the *myopic policy* that always sells as much as possible can be far from optimal in some cases.

We then restrict attention to a stylized lost-sales setting. Using interchange arguments, we show the structure of the optimal sales plan is always of the build-up type in the lost-sales setting. We characterize precisely the duration of the build-up period and, consequently, provide a means to explicitly compute the optimal sales plan. We also consider the case in which the firm may choose not to advertise or even inform the market about the existence of the product during the inventory build-up period, and thus prevent generation of any demand. We compute the resulting optimal build-up period, which is better termed *roll-out delay* in this case. Taking the impact of supply constraints on demand into account affects not only operational decisions like sales plans but also strategic decisions like production capacity choice. With a numerical example, we show how using a sales plan derived from a traditional Bass model could lead to incorrect decisions about capacity sizing. The reader may be tempted to believe that the complexity of the sales plan is an artifact of fixing production capacity. To dispel this notion, we study the case when the firm can choose any production level, albeit at a convex increasing production cost, and show that the insights obtained in the fixed-capacity setting continue to be relevant in this more general setting.

The rest of the paper is organized as follows. In §2, we discuss relevant literature. In §3, we present the Bass model and introduce our modification of the Bass model when all demand cannot be met due to supply constraints. In §4, we present a general normative model for determining the optimal sales plan. We provide a theoretical result on the structure of the optimal sales plan using Pontryagin's maximum principle. In §5, we present the results of an extensive numerical study of the general model and compare the performance of alternative heuristic sales plans. In §6, we consider the lost-sales case and prove the optimality of the build-up policy. We also solve the related problem of calculating the optimal roll-out delay and discuss the choice of optimal fixed capacity. In §7, we discuss the case when production is unrestricted but the cost of production is convex increasing. We show that the insights obtained from the fixed-capacity case carry over. In particular, we show that the concept of a sales plan is not vacuous in this setting. We provide some concluding remarks in §8.

2. RELATED LITERATURE

Eliashberg and Steinberg (1993) provide an overview of the research that integrates marketing and manufacturing decisions. They indicate that researchers have in the past tried to address issues in two different ways—either by considering centralized control of both units or through mechanism design (incentive and penalties) for the two units. Our approach in this paper is to assume that a central unit has control over both the units, and we introduce and analyze models where the demand occurs as a diffusion process.

In the economics literature, Griliches (1957) and Mansfield (1961) have proposed diffusion models for the spread of technological innovation. These models have been adopted in the marketing literature by Bass (1969), among others, to model the adoption of innovative products. They have then been used to determine optimal timing for product introduction, pricing, and advertisement, among other issues (see Mahajan et al. 1990 for a survey of related papers). Several researchers in marketing have worked on modifying the Bass model to incorporate advertising, changes in market potential, multistage and flexible diffusion, among others. Table 3 of Mahajan et al. (1990) provides an extensive review of such modifications of the Bass model.

Bass (1969) and subsequent, related diffusion models were meant to model the adoption of a product category (such as washing machines) rather than a particular product (such as one model of a Kenmore washing machine). So, their inability to model supply constraints is not surprising. However, supply constraints have been indirectly acknowledged as far back as Griliches (1957).

Figure 1 of Griliches (1957) shows that adoption of hybrid corn in states is delayed because of the time it takes to grow a sufficient number of generations of hybrid corn. In other cases such as Mansfield (1961), the authors have been careful to assume that patents or other such limitations did not constrain the spread of the innovation. However, one could easily conceive of a situation where a particular innovation considered by Mansfield (1961), such as a high-speed bottle filler, is made by a single monopolist protected by patents and is capacity constrained. In the marketing literature, Simon and Sebastian (1987) note in their paper that supply constraints may have distorted the parameter estimates they obtained by using the Bass model. However, most of the research on the Bass model has avoided explicit modeling of supply-side constraints (such as capacity restrictions, inventory handling capabilities, etc.).

Jain et al. (1991) consider supply constraints and present a modified Bass model where customers wait for future delivery in a queue if they do not get the product. This is often referred to as *backlogging* in the operations management literature, where customers wait if inventory is not available and are provided with the product when the inventory becomes available. Most of the modifications of the Bass model (including those above) are for forecasting or parameter estimation purposes and not for the design of optimal policies as studied in this paper. Horsky (1990), for one, does carry out optimal pricing policy design using his variant of the Bass model, and our paper is written in the same spirit as the implications section of Horsky (1990). Earlier papers like Dolan and Jeuland (1981) and Kalish (1985) also study optimal pricing in diffusions, but they too fail to consider supply constraints. Finally, we use optimal control framework which has been used by other researchers as well (see Feichtinger et al. 1994). However, the above papers do not study our particular setting and, further, the results proven in this paper (using interchange arguments) go beyond the standard optimal control methods.

There have been very few papers in operations management where the underlying demand is modeled using the Bass model. Kurawarwala and Matsuo (1996) consider a problem where the underlying demand is given by a Bass model but the parameters of the model are unknown. They study the problem of minimizing the total expected holding and stock-out costs during the horizon. El Ouardighi and Tapiero (1998) consider the impact of product quality on the diffusion and solve the problem of optimal quality. However, these papers do not take into account any supply constraints. Ho et al. (2002) study a Bass model with supply constraints, assert that a myopic sales plan is always optimal in their setting, and analyze performance under myopic sales plans. In this paper we consider a setting more general than that considered by Ho et al. (2002): Our model is capable of handling a variety of scenarios for backlogging unmet demand, ranging from complete backlogging to lost sales. We show that myopic sales plans need not be optimal in our setting, compare the performance of the myopic sales plans against other sales plans such as build-up plans, and prove that build-up plans are optimal in the lost-sales case.

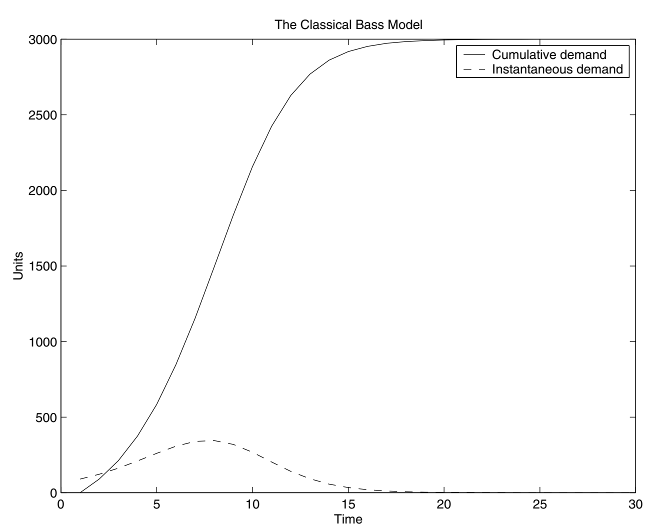
3. MODIFYING THE BASS MODEL FOR UNMET DEMAND

As mentioned in the introduction, the demand seen by a monopolist selling an innovative product can be considered to be given by the Bass model,

$$n(t) = p(m - N(t)) + \frac{q}{m} N(t)(m - N(t)). \tag{1}$$

The model is influenced by two parameters p and q, where p represents the impact of mass-media influences (also called the coefficient of innovation), and q represents the impact of the cumulative fraction who have already purchased the product (also called the coefficient of imitation). Thus, in a market with a fixed size m, the instantaneous demand for the product at time t, $n(t)$ expressed as a fraction of the remaining potential adopters $(m - N(t))$ has two components: one proportional to the constant contribution of mass media and the other proportional to the cumulative number of adopters up to time t, $N(t)$, representing the word-of-mouth effect. This model assumes that there are no repeat purchases by the adopters and that price does not affect adoption. For further details on the Bass model and its variants in marketing literature, see the survey paper by Mahajan et al. (1990). We will call (1) the "classical Bass model of demand" in the sequel. The classical Bass model gives rise to instantaneous and cumulative demand curves as shown in Figure 1. The parameter choices for the curves plotted in Figure 1 are $m = 3{,}000$, $p = 0.03$, $q = 0.4$. The instantaneous demand starts off due to the effect of innovators. Then it rapidly grows as the effect of imitators kicks in, reaches a maximum value, then decreases and finally tends to zero. The following lemma summarizes some useful properties related to the classical Bass model.

Figure 1. Cumulative and instantaneous demand in the classical Bass model.



Lemma 1. (i) *The cumulative demand* $N(t)$ *and instantaneous demand* $n(t)$ *are given by*

$$N(t) = m\left\{\frac{1-\exp(-(p+q)t)}{1+(q/p)\exp(-(p+q)t)}\right\}$$

and

$$n(t) = m\left\{\frac{p(p+q)^2\exp(-(p+q)t)}{(p+q\exp(-(p+q)t))^2}\right\}.$$

(ii) *The maximum value of demand is reached at time* t^* *where* $t^* = -\{1/(p+q)\}\ln p/q$.

(iii) *The maximum value of demand at* t^* *is given by* $n(t^*) = (m(p+q)^2)/4q$.

Proof. See Mahajan et al. (1990). □

An assumption of the classical Bass model is that there are no supply constraints and, hence, there are no limits on the firm's ability to meet demand. We propose a modification that attempts to relax this assumption. In light of the rapid growth of demand (as seen in Figure 1), it is possible that a large number of consumers may attempt to buy the product, but may be unsuccessful due to supply constraints. It is unreasonable to assume that these customers would continue to spread word about the product. It is far more reasonable to assume that any positive word-of-mouth effect would be due to only those customers who actually purchased a product. That is, the number of people who attempt to purchase a product by imitation at a given time t is influenced by the number of people who have successfully bought the product up to time t, and not necessarily by all the people who demanded the product up to time t. The word-of-mouth effect is better represented as being proportional to the cumulative sales $S(t)$. Therefore, we propose the following modification to the Bass model.

$$n(t) = p(m-N(t)) + \frac{q}{m}S(t)(m-N(t)). \tag{2}$$

Note that (2) is valid regardless of whether unmet demand $N(t)-S(t)$ is lost or is backlogged for eventual fulfilment. In deriving (2) we have attempted to retain one of the most useful features of the classical Bass model: a parsimonious analytical representation. Also note that in the modified Bass model, the firm's past sales, and consequently, its production and sales plans, directly impact the future demand for the product. This is a mixed blessing. It indeed connects production, sales, and inventory, that are usually operational issues in a firm, with future demand, usually a marketing issue. Thus, it provides a canonical model of marketing/manufacturing coordination. The flip side is that it makes the problem of finding the optimal production and sales plans much more challenging. We will attempt this in the sequel.

4. SALES AND PRODUCTION PLANS UNDER THE MODIFIED BASS MODEL

We begin by specifying a continuous-time model in which demand is given by the modified Bass model (2), there is a fixed production capacity c, and unmet demand is partially backlogged. Lost sales and complete backlogging are special cases of this model. The model assumes a fixed unit production cost $\alpha \geqslant 0$, a fixed unit selling price $\pi > 0$, a waiting cost $w \geqslant 0$ per unit backlogged per unit time, and a holding cost $h \geqslant 0$ per unit inventoried per unit time. Given a discount rate $\gamma \geqslant 0$, the objective is to maximize the discounted profit of the firm over the life cycle of the product. The resulting optimization problem that needs to be solved is

$$\max_{s(t),\, x(t);\, 0\leqslant t\leqslant T} J = \int_0^T e^{-\gamma t}\left[\pi s(t) - \alpha x(t) - wL(t) - hI(t)\right]dt \tag{3}$$

$$\text{s.t.}\quad \dot N(t) = n(t) \tag{4}$$

$$\dot S(t) = s(t) \tag{5}$$

$$\dot n(t) = -pn(t) + \frac{q}{m}\big[-n(t)S(t) + (m-N(t))s(t)\big] \tag{6}$$

$$\dot I(t) = x(t) - s(t) \tag{7}$$

$$\dot L(t) = \xi(n(t)-s(t)) - (1-\xi)L(t) \tag{8}$$

along with the inequality constraints

$$L(t) \geqslant 0 \quad \text{and} \quad L(0) = 0, \tag{9}$$

$$I(t) \geqslant 0 \quad \text{and} \quad I(0) = 0, \tag{10}$$

$$0 \leqslant s(t), \quad \text{and} \tag{11}$$

$$0 \leqslant x(t) \leqslant c \quad \text{for all} \quad t \geqslant 0. \tag{12}$$

As before, $S(t), N(t)$, and $n(t)$ are defined to be the cumulative sales, cumulative demand, and instantaneous demand. $s(t)$ denotes the instantaneous sales at time t and $x(t)$ denotes the instantaneous production at time t. $I(t)$ denotes the inventory of products that have been produced but not sold at time t and $L(t)$ denotes the number of backlogged orders at time t. The parameters c and T denote the production capacity and time horizon of the problem.[1]

The parameter $\xi \in [0,1]$ denotes the fraction of unmet demand that is backlogged. Equation (8) describes the dynamics associated with the backlogging process in the following manner. A fraction ξ of unmet demand is instantaneously backlogged and a fraction $1-\xi$ of the current backlogged customers instantaneously abandon the queue and never return. When $\xi = 1$ we have complete backlogging and when $\xi = 0$ we have complete lost sales. For all other choices of ξ we have partial backlogging. The constraints (4)–(8) define the dynamics of the system, with (6) being a convenient representation of (2). The inequality constraints (9)–(12) are natural. The following theorem characterizes all optimal policies. The use of an asterisk with a quantity denotes that the quantity is optimal. Proofs can be found in the appendix.

Theorem 1. *At every instant of time t, an optimal sales plan $s^*(t)$ in (3)–(12) is of the following form. Either*

$$s^*(t) = 0 \tag{13}$$

or

$$s^*(t) = \begin{cases} x^*(t) & \text{if } I^*(t)=0 \quad \text{and} \quad L^*(t)>0, \\ n^*(t) & \text{if } L^*(t)=0 \quad \text{and} \quad I^*(t)>0, \quad \text{and} \\ \min(x^*(t), n^*(t)) & \text{if } I^*(t)+L^*(t)=0. \end{cases} \tag{14}$$

Furthermore, $I^*(t)L^*(t) > 0$ only if $S^*(t) = 0$. To put it another way, the optimal sales plan either sells nothing at all or as much as possible.

In order to obtain insights on the optimal production and sales policy, we numerically solve the finite dimension optimization problem obtained by considering the following discrete-time version.

$$\max_{s(t),\, x(t);\, t=0,1,\ldots,N} J = \sum_{t=0}^{N} \beta^t \left[\pi s(t) - \alpha x(t) - wL(t) - hI(t)\right] \tag{15}$$

$$\text{s.t.} \quad N(t+1) - N(t) = n(t) \tag{16}$$

$$S(t+1) - S(t) = s(t) \tag{17}$$

$$n(t) = p(m - N(t)) + \frac{q}{m} S(t)(m - N(t)) \tag{18}$$

$$I(t) = \sum_{s=0}^{t} x(s) - S(t) \tag{19}$$

$$L(t+1) = \xi(L(t) + n(t) - s(t)) \tag{20}$$

along with the inequality constraints

$$L(t) \geqslant 0 \quad \text{and} \quad L(0) = 0, \tag{21}$$

$$I(t) \geqslant 0 \quad \text{and} \quad I(0) = 0, \tag{22}$$

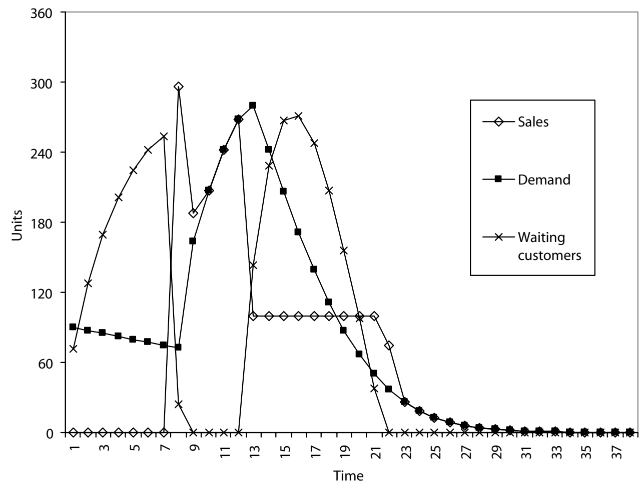
$$0 \leqslant s(t), \quad \text{and} \tag{23}$$

$$x(t) \leqslant c \quad \text{for all} \quad t \geqslant 0. \tag{24}$$

In the above formulation, we pick N to be large enough that the market potential m has been almost entirely exhausted, and we translate the discount rate γ to a suitable analog $\beta \in (0, 1)$ in the discrete case. For all the numerical investigations in this section, we assume that the parameters of the Bass model are given by $m = 3{,}000$, $p = 0.03$, $q = 0.4$, $c = 100$, and the per-unit cost $\alpha = 1$. We now illustrate the various forms that an optimal sales plan can take.

The optimal sales plan seen in Figure 2 can be explained as follows.[2] Initially, there are no sales, all production is used to build inventory, and all customers are backlogged. When there is a critical number of backlogged customers, the built-up inventory is used to "flush out" the backlog queue, instantaneously. From then on, the firm

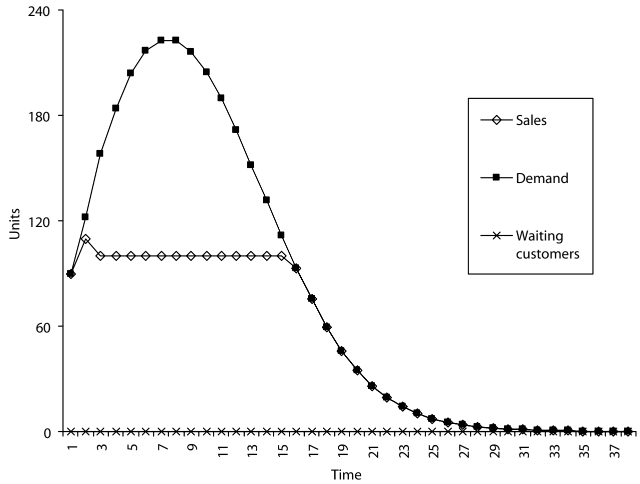
Figure 2. Example 1: $\beta = 0.995$, $h = 0.005$, $\pi = 1.3$, $w = 0.001$, and $\xi = 0.8$.



sells as much as possible: It meets all demand as long the inventory lasts, then it sells up to production capacity until demand falls below production capacity, after which it meets demand. In the presence of discounting and partial backlogging, it may appear counterintuitive that the firm chooses to deliberately avoid sales in the first few periods in order to build inventory. However, avoiding sales deliberately in the initial periods enables the firm to control the rate of demand growth. If the firm sold as much as possible initially, the consequent demand growth could cause excessive backlogging. This may lead to higher costs in the future, both due to waiting costs incurred by backlogging, as well as lost sales due to abandonment from the queue. The benefits of earlier sales may be outweighed by these costs. When the inventory on hand is sufficiently large and the number of backlogged customers is high, it is no longer beneficial for the firm to avoid sales. From that time onwards, the firm chooses to sell as much as possible. Note that when the firm starts selling as much as possible, the remaining market potential $m - N(t)$ is smaller than at the beginning, and hence the demand growth tends to be more manageable, resulting in smaller backlogs.

For the choice of parameters in Figure 2, the optimal policy is of the *build-up* variety. That is, the optimal policy does not sell at all for an initial period (building inventory) and then sells as much as possible. Clearly, this need not always be optimal. Figure 3 illustrates a case where the myopic sales plan which always sells as much as possible is optimal. To illustrate another possibility, we present Figure 4, where the optimal sales plan first sells as much as possible and then starts building inventory by not selling at all, and finally sells as much as possible once there is sufficient inventory. Figures 2–4 indicates the rich possibilities for optimal policies. Computing optimal plans explicitly is quite difficult in general. Build-up and myopic plans have simple, intuitive structures. Therefore, the firm could use build-up, myopic, or a combination merely as heuristic

Figure 3. $\beta = 0.99$, $h = 0.01$, $\pi = 1.3$, $w = 0$, and $\xi = 0$.

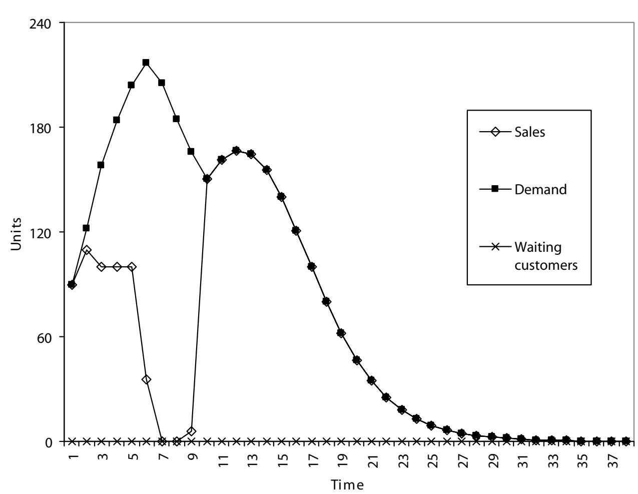


plans. It is necessary to quantify the deviation from optimality of these heuristics over a wide range of parameters in order to justify their use. In the next section, using a numerical study, we compare build-up, myopic, and optimal plans and provide insights.

5. NUMERICAL INVESTIGATION OF HEURISTICS

We initially investigate two heuristics that are consistent with Theorem 1. They are the *myopic* policy, which always sells as much as possible, and the *build-up* policy, which builds up inventory in the first few periods (hereafter referred to as the build-up period) and then starts selling as much as possible. In picking the build-up heuristic, we search among all integer build-up periods and choose the best alternative. Note that the myopic policy is a special case of the build-up heuristic where the build-up period is zero. However, whenever we use the term build-up we are restricting attention to the case where the build-up period is strictly positive.

Figure 4. $\beta = 0.99$, $h = 0.001$, $\pi = 1.3$, $w = 0.005$, and $\xi = 0$.



It is desirable to use appropriate parameter choices in our detailed numerical study of the discrete model (15)–(24). Empirical studies of the Bass model report a wide range for the parameters p and q. Based on meta-analysis of 213 data sets drawn from a wide variety of settings (from 15 articles), Sultan et al. (1990) report mean values of p and q to be 0.03 and 0.38, respectively. Therefore, we use $p = 0.03$ and $q = 0.4$ in our study to serve as representative values for potential applications. We set $m = 3{,}000$, a scale parameter, which results in the market being exhausted in approximately 30 periods (i.e., $T = 30$). Therefore, we choose production capacity to be the average demand per period, i.e., $c = 100$. If each period is two to four weeks, this corresponds to a product whose life cycle ranges from one to two years.

We normalize per-unit cost at $\alpha = 1$. We consider three options for the selling price $\pi = 1.1$, 1.2, and 1.3. We consider high, medium, and low values for holding costs $h =$ 0.01, 0.005, and 0.001, respectively, representing values ranging from 1% to 0.1% of unit production cost every two weeks. Similarly, we choose three values for the waiting cost, $w = 0.01$, 0.005, and 0.001. We study full backlogging ($\xi = 1$), lost sales ($\xi = 0$), and two partial-backlogging scenarios ($\xi = 0.5$, 0.8). Finally, all the above settings are studied for four different discount rates $\beta = 0.99$, 0.995, 0.997, and 1.0, capturing a wide range from 1% every two weeks to no discounting at all. This results in 432 cases.

Table 1 lists the maximum and average deviation of myopic and build-up heuristics from the optimal policy. The maximum deviation from optimality of build-up is less than 3.5% and the average deviation (across all 432 instances) is 0.30%. Although the average deviation of the myopic heuristic is 3.5%, the maximum deviation could be as high as 25%. Table 2 shows the improvement in terms of the percentage of cases where build-up outperforms the myopic heuristic (out of 108 cases for each choice of β), the maximum length of the build-up period, the maximum and average percentage improvement in profit using the build-up rather than myopic heuristic. In a total of 225 out of 432 cases, i.e., 52% of the cases, the build-up heuristic performs better than the myopic heuristic. Note that the maximum length of the build-up period can be as high

Table 1. Percentage deviation from optimality of build-up and myopic heuristics.

	$\beta = 0.99$	$\beta = 0.995$	$\beta = 0.997$	$\beta = 1.0$
Average deviation of build-up	0.04	0.14	0.28	0.73
Average deviation of myopic	0.08	1.69	3.45	7.77
Max deviation of build-up	0.84	1.59	2.32	3.32
Max deviation of myopic	2.09	9.78	14.86	25.62

Table 2. Comparison of myopic and build-up heuristics.

β	% Build-Up > Myopic	Max Build-Up Period	Max % (Average %) Improvement Over Myopic
0.99	12	5	1.26 (0.04)
0.995	49	10	9.63 (1.57)
0.997	64	10	14.86 (3.20)
1	83	12	25.62 (7.14)

as 12 (which is nearly 40% of the horizon), illustrating a significant prescriptive difference from the myopic policy. Furthermore, when the performance of the myopic policy is inferior to that of the build-up policy, it can be significantly worse.

The following excerpts from our numerical study illustrate the relative performance of the two heuristics as parameters of the model change. Tables 3–8 show the dominant heuristic for a representative set of parameters. Of course, neither heuristic need be optimal. The build-up policy trades off stimulating the market and having customers wait in the backlog queue (or losing customers altogether due to abandonment) against building inventory and discounting due to delayed sales. Therefore, one expects improvement in the performance of the build-up policy relative to the myopic policy when the holding cost h decreases, when the discount factor β increases, or the waiting cost w increases. The build-up policy also avoids disappointing customers by not overstimulating the market. Thus, when the cost of disappointing the customer (i.e., the profit lost due to abandonment from the queue or the cost of waiting) is high, the build-up heuristic is preferable. These insights are reflected in Tables 3–8. For example, when $\beta = 0.99$, there is no benefit to using a build-up policy, except when the margins (π) are high and holding costs are low. However, when $\beta = 1$ the build-up policy outperforms the myopic policy, except when margins are low or the customers are infinitely patient. For intermediate values of β the relative performance of the policies is determined by the holding cost, with high holding costs favoring the myopic policies, and vice versa. Finally, one would expect that the relative performance of the build-up policy increases as the capacity of the firm decreases. We will have more to say on this issue in §7.

As is common in optimal control, Theorem 1 prescribes extreme solutions such as the myopic and build-up heuristics analyzed above. It is conceivable that the firm would moderate such a prescription. The firm may choose to build up some inventory for the future while continuing to serve part of the market. The firm would also be loath to let the backlog become too large. We study the performance of such a plan, where the firm serves a fraction $0 < \delta < 1$ of the current demand if possible (i.e., $s(t) \leqslant \delta n(t)$) until the backlog reaches $\overline{L}$, (the maximum backlog that the manager is willing to suffer as a consequence of deliberately denying sales), after which the policy switches to the myopic rule. Table 9 shows the performance of this heuristic for three values of δ and a fixed choice of $\overline{L} = 50$ across a representative subset of 48 instances. The performance of this heuristic is inferior to both myopic and build-up heuristics when $\beta = 0.99$, but improves with respect to the myopic policy as β increases. Such behavior is plausible given that the moderate heuristic behaves as a combination of myopic and build-up heuristics.

In a recent paper, Ho et al. (2002) analyze a model similar to ours and assert that a myopic sales plan is always optimal for their model. Theorem 1 does allow for the myopic sales plan to be optimal, but the numerical results described in this section indicate that the myopic plan may be far from optimal. Although the models considered in Ho et al. (2002) and (3)–(12) are different, it is unlikely that the differences in the models alone account for the contrasting results. Resolving this issue is the subject of future work.

Given that the performance of the build-up heuristic improves when the holding cost is small and the cost of disappointing customers is high, we focus our attention on the special case of lost sales and prove that the build-up policy is indeed optimal for this special case.

6. THE SPECIAL CASE OF LOST SALES

In this section, we restrict attention to the case where $\beta = 1$, $h = 0$, $w = 0$, $\xi = 0$, and $\alpha = 0$. In this stylized setting, we can analytically obtain the optimal policy. Under these assumptions, with a fixed market potential, the problem of maximizing profits over the product life cycle is equivalent

Table 3. Dominant heuristic when $\xi = 0.8$, $\pi = 1.2$, and $h = 0.01$.

β	$w = 0.001$	$w = 0.005$	$w = 0.01$
0.99	Myopic	Myopic	Myopic
0.995	Myopic	Myopic	Myopic
0.997	Myopic	Myopic	Build-up
1	Build-up	Build-up	Build-up

Table 4. Dominant heuristic when $\xi = 0.8$, $\pi = 1.2$, and $h = 0.001$.

β	$w = 0.001$	$w = 0.005$	$w = 0.01$
0.99	Myopic	Myopic	Myopic
0.995	Build-up	Build-up	Build-up
0.997	Build-up	Build-up	Build-up
1	Build-up	Build-up	Build-up

Table 5. Dominant heuristic when $\xi = 0.8$, $w = 0.005$, and $h = 0.01$.

β	$\pi = 1.1$	$\pi = 1.2$	$\pi = 1.3$
0.99	Myopic	Myopic	Myopic
0.995	Myopic	Myopic	Build-up
0.997	Myopic	Myopic	Build-up
1	Myopic	Build-up	Build-up

Table 6. Dominant heuristic when $\xi = 0.8$, $w = 0.005$, and $h = 0.001$.

β	$\pi = 1.1$	$\pi = 1.2$	$\pi = 1.3$
0.99	Myopic	Myopic	Build-up
0.995	Build-up	Build-up	Build-up
0.997	Build-up	Build-up	Build-up
1	Build-up	Build-up	Build-up

to minimizing lost sales over the life cycle of the product. Since there are no inventory holding costs, and no costs to wasting inventory, *one* optimal production plan is to produce the entire market potential as soon as possible, that is, in m/c time units, by producing at capacity c until the m units are produced. The optimal sales plan, however, is not trivial. The problem of determining the optimal sales plan now becomes

$$\min_{s(t),\, 0\leqslant t\leqslant\infty} J = N(\infty) - S(\infty) \tag{25}$$

$$\text{s.t.} \quad \dot{N}(t) = n(t) \tag{26}$$

$$\dot{S}(t) = s(t) \tag{27}$$

$$\dot{n}(t) = -pn(t) + \frac{q}{m}\big[-n(t)S(t) + (m - N(t))s(t)\big] \tag{28}$$

along with the inequality constraints

$$s(t) \leqslant n(t), \tag{29}$$

$$s(t) \geqslant 0, \tag{30}$$

$$S(t) \leqslant ct \quad \text{for all} \quad t \geqslant 0. \tag{31}$$

6.1. Optimality of the Build-Up Policy

In this subsection, we will prove the optimality of the build-up policy when there is no initial inventory in the system. As before, use of the asterisk $(*)$ with a quantity indicates that the quantity is optimal.

THEOREM 2. *When demand is given by the modified Bass model* (2), *and when there is no initial inventory in the system, the optimal selling plan* (*specified by the optimal cumulative sales process* $S^*(\cdot)$) *is a build-up plan. That is, there exists a time* t_0, *which can be computed from the values of the market potential m, the coefficient of innovation p, the coefficient of imitation q, and the production capacity c, such that the optimal cumulative sales* $S^*(t_0) = 0$. *Furthermore,* $S^*(t) = N^*(t) - N^*(t_0)$ *for all* $t > t_0$. *That is, all lost sales are incurred* **only** *in* $[0, t_0]$.

Table 7. Dominant heuristic when $\pi = 1.2$, $w = 0.005$, and $h = 0.01$.

β	$\xi = 0$	$\xi = 0.5$	$\xi = 0.8$	$\xi = 1$
0.99	Myopic	Myopic	Myopic	Myopic
0.995	Myopic	Myopic	Myopic	Myopic
0.997	Myopic	Myopic	Myopic	Myopic
1	Build-up	Build-up	Build-up	Myopic

Table 8. Dominant heuristic when $\pi = 1.2$, $w = 0.005$, and $h = 0.001$.

β	$\xi = 0$	$\xi = 0.5$	$\xi = 0.8$	$\xi = 1$
0.99	Myopic	Myopic	Myopic	Myopic
0.995	Build-up	Build-up	Build-up	Myopic
0.997	Build-up	Build-up	Build-up	Build-up
1	Build-up	Build-up	Build-up	Build-up

We will use optimal control theory and novel interchange arguments to establish this result via a sequence of lemmas. As mentioned earlier, proofs of all the lemmas are postponed to the appendix. We will first establish in Lemma 2 that in an optimal selling plan, the only acceptable choices for the instantaneous sales $s(t)$ are 0, c, or the instantaneous demand $n(t)$, and that an optimal policy switches between these choices only finitely many times. We do this by setting up the Hamiltonian for the optimal control problem and by arguing that it is affine in the control $s(t)$, in Lemma 2. This result is similar to Theorem 1. Next, in Lemma 3 we show that after some time the optimal policy is guaranteed to meet all future demand. Using this, we argue in Lemma 4 that an optimal policy could have only been selling $s^*(t) = 0$ just before it begins meeting all future demand. Lastly, we argue that the beginning of this period over which the optimal policy was selling $s^*(t) = 0$ before meeting all future demand is in fact at $t = 0$ in Lemma 5. Thus, we establish that an optimal policy is of the build-up type. Finally, the switch over time (when the policy switches from selling nothing to meeting all future demand) is explicitly computed in Lemma 6.

LEMMA 2. *The only possible choices for the optimal instantaneous sales* $s^*(t)$ *are* 0, c, *or the instantaneous demand* $n^*(t)$. *Furthermore, there are only finitely many switches between these choices.*

Having computed the Hamiltonian, we could attempt to explicitly compute the co-state trajectories and hence resolve how the switches between these choices occur. Rather than carry out this nearly intractable computation, we resort to an indirect interchange argument that rules out all switches except one from $s^*(t) = 0$ to $s^*(t) = n^*(t)$. We do this as follows. First, we argue that under an optimal selling plan, after some finite time the instantaneous demand falls below the production capacity and stays below it forever in Lemma 3.

Table 9. Percentage deviation from optimality averaged over w and ξ with $\overline{L} = 50$, $\pi = 1.2$, and $h = 0.005$.

	$\beta = 0.99$	$\beta = 0.995$	$\beta = 0.997$	$\beta = 1.0$
$\delta = 0.4$	0.19	0.34	1.96	6.80
$\delta = 0.6$	0.18	0.34	1.96	6.81
$\delta = 0.8$	0.11	0.35	2.01	6.88
Build-Up	0.01	0.08	0.06	0.30
Myopic	0.01	0.36	2.06	6.99

Lemma 3. *Under any optimal selling plan, there exists a time t'_f such that for all $t \geqslant t'_f$, $n^*(t) < c$. Therefore, for all $t \geqslant t'_f$, $s^*(t) = n^*(t)$.*

Thus, any optimal plan eventually switches to $s^*(t) = n^*(t)$. In the sequel we will use the following definition of t_f under any given optimal policy.

$$t_f = \inf\{s\colon n^*(t) < c, \text{ for all } t \geqslant s\}. \tag{32}$$

Next, we argue that a switch from $s^*(t) = c$ to the eventual $s^*(t) = n^*(t)$ is not optimal in Lemma 4. The basic idea of the proof is that a small perturbation of a policy that makes this switch results in a policy whose performance is no worse than that of the original policy. However, since the perturbed policy sells at levels other than 0, c, or $n(t)$, it cannot be optimal by Lemma 2. Thus, the c to $n(t)$ switch is ruled out.

Lemma 4. *Consider an optimal policy, and the corresponding t_f as specified by* (32). *Let $t_0 = \inf\{t \leqslant t_f\colon s^*(t) = n^*(t)\}$. Then either $t_0 = 0$ or for every δ sufficiently small, $s^*(t_0 - \delta) = 0$.*

Next, in Lemma 5 we rule out switches of the form $s(t) = x$, where x can be either c or $n^*(t)$, followed by a switch to $s^*(t) = 0$ and then to $s^*(t) = n^*(t)$, as established above in Lemma 4.

Lemma 5. *Consider an optimal policy, and the corresponding t_0 as specified by Lemma* 4. *Then,*

$$t_1 := \inf\{t \leqslant t_0\colon s^*(t) = 0\} = 0.$$

Lemma 5 above establishes that any optimal policy is a "build-up" policy. That is, there are no sales up to a time t_0 after which all the demand is met. The last unresolved issue in specifying an optimal policy is to give a procedure for explicitly computing the t_0 as specified by Lemma 4. We could reduce this to a one-parameter search problem, but we have a simpler way of doing this, as described in the following lemma.

Lemma 6. *If the parameters p, q, m, and c satisfy*

$$ct > m\left[\frac{1 - e^{-(p+q)(t)}}{1 + (q/p)e^{-(p+q)(t)}}\right], \quad \textit{for all} \quad t > 0, \tag{33}$$

then $t_0 = 0$. Otherwise, the optimal time to build-up, t_0, and the time t_f defined in (32) *satisfy the simultaneous equations*

$$ct_f = \bar{m}\left[\frac{1 - e^{-(p+\bar{q})(t_f - t_0)}}{1 + (\bar{q}/p)e^{-(p+\bar{q})(t_f - t_0)}}\right], \tag{34}$$

$$c = \bar{m}\left[\frac{p(p+\bar{q})^2 e^{-(p+\bar{q})(t_f - t_0)}}{(p + \bar{q}e^{-(p+\bar{q})(t_f - t_0)})^2}\right], \tag{35}$$

where

$$\bar{m} = me^{-pt_0}, \tag{36}$$

and

$$\bar{q} = qe^{-pt_0}. \tag{37}$$

Thus, we have to solve a set of simultaneous nonlinear equations to obtain both t_0 and t_f. Alternately, one can consider a class of policies parameterized by t_0 and optimize the objective function $N(t_f) - S(t_f)$ over this one parameter.

6.2. Initial Inventory and Delayed Roll-Out

The optimal solution obtained for the special case can also be interpreted as a *delayed roll-out* strategy. Basically, under such a strategy a firm may wait to accumulate enough inventory before bringing the product to market, while possibly losing some of the total potential market m because of the delay. Note that in our optimal policy derived in the previous subsection, during the build-up period of length t_0, the market potential reduces exponentially from m to $\bar{m} = m(1 - e^{-pt_0})$. So, one can interpret the build-up time in the optimal policy, derived in the previous subsection, as the optimal roll-out delay in a setting where the market potential decreases in an exponential manner during the roll-out delay.

In our model, we assumed that the firm does not have any inventory on hand at time $t = 0$. An alternative problem could be posed as to how much inventory a firm should have at time $t = 0$ in order to avoid any lost sales during the product life. This inventory level also determines the optimal inventory build-up period during which the firm chooses not to advertise or even inform the market about the existence of the product. It is more descriptive to label this build-up period the optimal roll-out delay. We can characterize this explicitly based on the following lemma.

Lemma 7. *The minimum amount of initial inventory i_0 required in order to avoid any lost sales during the lifetime of a product whose demand follows the Bass model is given by the solution to the following equation*:

$$i_0 + ct_2 = m\left\{\frac{1 - \exp(-(p+q)t_2)}{1 + (q/p)\exp(-(p+q)t_2)}\right\}, \tag{38}$$

where t_2 is the larger root of

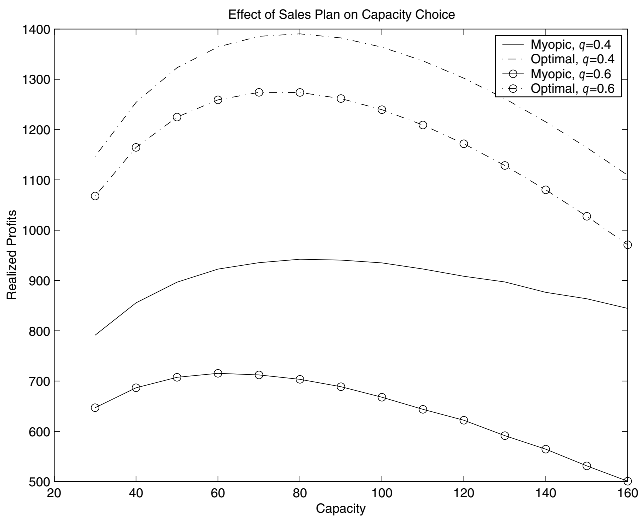
$$c = m\left\{\frac{p(p+q)^2 \exp(-(p+q)t_2)}{(p + q\exp(-(p+q)t_2))^2}\right\}. \tag{39}$$

Furthermore, the optimal roll-out delay is given by i_0/c.

6.3. Choosing the Optimal Capacity

In our previous analysis, we assumed that the capacity c was given. In general, one would expect that the firm would utilize some estimate of demand in order to determine what capacity level to install. In the modified Bass model, such capacity decisions will depend on the sales plan that is employed, since the sales plan determines demand. The capacity decision may be made assuming either that the sales plan that will be used is myopic, or that the optimal sales plan, which is a build-up sales plan for the lost-sales model by Theorem 2, is used. We conducted a limited computational study (using the parameters described in the

Figure 5. Effect of selling plan on capacity sizing decision when $p = 0.03$ and $m = 3{,}000$.



last subsection) to see how capacity decisions are different in these two settings. We assume that the cost of per-unit capacity is \$10 and the profit per sale is \$1 in Figure 5. The optimal capacity sizing decisions could be quite different depending on whether one used a myopic sales plan or the build-up plan which is optimal at each fixed capacity. For the case when $q = 0.6$, the optimal capacity choice under the myopic sales plan is smaller than the optimal capacity choice under the optimal build-up sales plan. The opposite is true when $q = 0.4$. The explanation for this effect is that the marginal benefit of a unit of additional capacity is different in the build-up plan than in the myopic plan. From a managerial standpoint, this indicates that if a firm utilizes a myopic sales plan, not only will it end up with lower sales for the product at a given capacity, but also that it could end up choosing an inappropriate level of production capacity, further exacerbating the negative effects of the choice of sales plans.

7. UNRESTRICTED PRODUCTION WITH CONVEX PRODUCTION COST

In §4, we assumed that the production capacity c is fixed throughout the life cycle of the product. An alternative setting is one in which a firm might be able to increase production using overtime or outsourcing, albeit at a higher production cost. In this section, we model such a setting. To be specific, we assume that if the firm chooses to produce at level $x(t)$ at time t, it incurs a *production cost* $C(x(t))$ where $C(\cdot)$ is a nonnegative, convex, increasing function. Such a model captures both endogenous capacity changes as well as the ability to outsource production and the associated costs. Our goal is to explore whether the insights obtained in the fixed capacity setting are valid in this more general setting. To be precise, we consider the following variant of the optimal control problem (3)–(12). The key changes from (3)–(12) are the inclusion of the production cost function in the objective function and the relaxation of the capacity constraint (12)–(49).

$$\max_{s(t),x(t);0\leqslant t\leqslant T} J = \int_0^T e^{-\gamma t}\big[\pi s(t) - C(x(t)) - wL(t) - hI(t)\big]\,dt \tag{40}$$

$$\text{s.t.}\quad \dot N(t) = n(t) \tag{41}$$

$$\dot S(t) = s(t) \tag{42}$$

$$\dot n(t) = -pn(t) + \frac{q}{m}\big[-n(t)S(t) + (m - N(t))s(t)\big] \tag{43}$$

$$\dot I(t) = x(t) - s(t) \tag{44}$$

$$\dot L(t) = \xi(n(t) - s(t)) - (1-\xi)L(t) \tag{45}$$

along with the inequality constraints

$$L(t) \geqslant 0 \quad \text{and} \quad L(0) = 0, \tag{46}$$

$$I(t) \geqslant 0 \quad \text{and} \quad I(0) = 0, \tag{47}$$

$$0 \leqslant s(t), \quad \text{and} \tag{48}$$

$$0 \leqslant x(t) \quad \text{for all} \quad t \geqslant 0. \tag{49}$$

In order to numerically compute optimal strategies using a nonlinear programming routine, we discretize the problem in the usual way, as in (15)–(24). The details are omitted for brevity. As before, for all the numerical investigations in this section, we assume that the parameters of the Bass model are given by $m = 3{,}000$, $p = 0.03$, and $q = 0.4$. We set the unit selling price at $\pi = 1.2$. We assume that the production cost function $C(x)$ is given by $C(x) = ax^b$ where $a > 0$ and $b > 1$ are constants. This function is convex increasing for all $a, x > 0$ and $b > 1$, and by changing a, b we can control the shape of the function.

Figures 6–8 illustrate the three types of policies observed by us in our numerical investigations. In each of these figures, we set the waiting cost $w = 0.005$, the holding cost $h = 0.005$, the selling price $\pi = 1.2$, the degree of backlogging $\xi = 0.8$, and the discount rate $\beta = 0.995$. We vary the shape of the production cost function $C(x) = ax^b$ by successively choosing $b = 1.05$, 1.1, and 1.15, while keeping a fixed at 0.6.[3]

When $b = 1.05$ (Figure 6), the marginal cost of production does not increase too much with the production level x. In this case, it is optimal to sell as much as the demand (i.e., $s^*(t) = n^*(t)$) at all times. There is never any backlogging and, consequently, there are no lost sales. When b is increased to 1.1, the marginal cost of production increases with x more rapidly, and hence, the optimal sales plan avoids increasing the production level x excessively. In this case, it is no longer optimal to set $s^*(t) = n^*(t)$ and, consequently, there is backlogging. The optimal sales plan turns out to be a build-up plan, as illustrated in Figure 7. In Figure 7, until Period 5, there are no sales and inventory is built up. Beyond the point the firm sells as much as

Figure 6. $C(x) = 0.6x^{1.05}$: Myopic sales plan without backlogging.

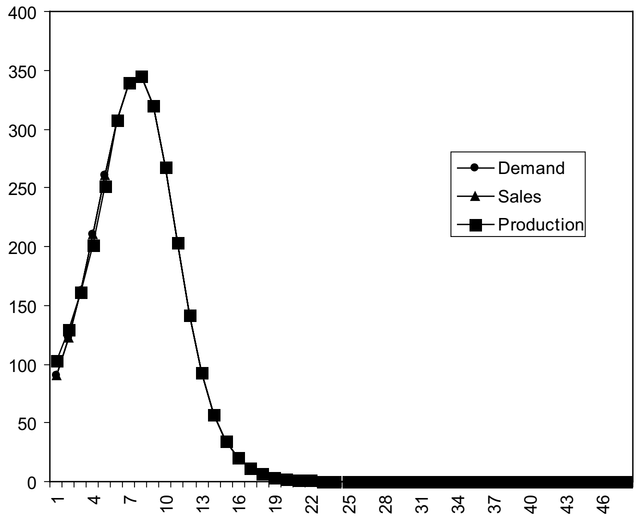
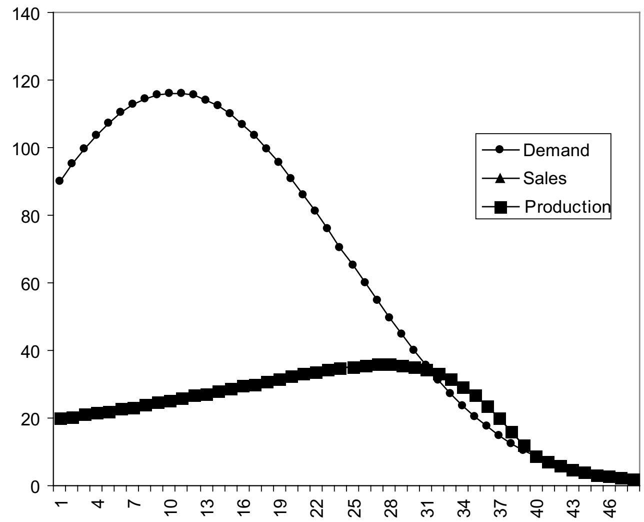


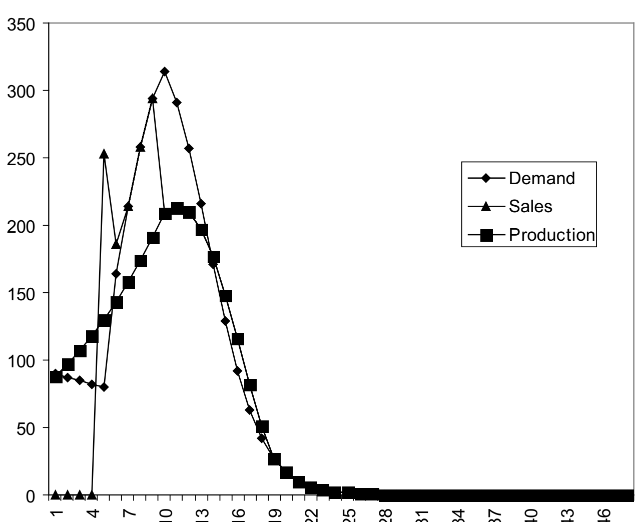
Figure 8. $C(x) = 0.6x^{1.15}$: Myopic sales plan with backlogging.



possible—initially meeting demand until inventory runs out and then selling as much as produced. When b is increased even further to 1.15, the marginal cost of production is so large that the production levels chosen are very small. Consequently, the ability to build-up inventory is limited. Hence, the optimal sales plan is myopic (i.e., it sells as much as is produced, with $s^*(t) = x^*(t)$ for all $t \geqslant 0$), and it incurs a large number of lost sales.

Table 10 illustrates the general pattern of the observed optimal sales plans for 20 cases with different values of a and b when $w = h = 0.005$, $\beta = 0.995$, $\pi = 1.2$, and $\xi = 0.8$. A myopic sales plan with no resulting backlog (denoted MWB) is optimal when the marginal cost of production is low. When the marginal production cost is increased via either a or b, the optimal sales plan is a build-up plan (denoted BU). When the marginal production cost becomes very high, the optimal sales plan is dictated by the production policy and is a myopic plan that builds up no inventory and incurs backlogging and lost sales (denoted MB). The above experiments demonstrate that the optimal sales plan need not be myopic and that the build-up plan may be optimal, albeit over a limited range of cost parameters even when there are no capacity constraints and the firm incurs convex production costs. However, when production capacity tends to be relatively cheap (as when $b = 1.05$) and the firm is less capacity constrained, the myopic sales plan is optimal.

Figure 7. $C(x) = 0.6x^{1.1}$: Build-up sales plan.



8. CONCLUSIONS

We have presented a canonical model of marketing-manufacturing interaction that captures the effect of supply constraints on the demand for innovative products. In extensive numerical studies, we have used this model to compute the optimal sales plans for monopolistic firms. We have established that build-up sales plans, which have a simple intuitive structure, are robust heuristics over a wide range of parameter settings. We also have rigorously proved that build-up sales plans are optimal when the firm is interested in minimizing lost sales. We have provided a brief illustration of the effect of relaxing capacity constraints and have shown that the shape of the production cost function can influence the choice of sales plans in a nontrivial way. We believe that the model and analysis in this paper bring out

Table 10. Form of the optimal sales plan when production cost $C(x) = ax^b$ and $w = h = 0.005$, $\beta = 0.995$, $\pi = 1.2$, and $\xi = 0.8$.

b	$a = 0.4$	$a = 0.5$	$a = 0.6$	$a = 0.7$	$a = 0.8$
1.05	MWB	MWB	MWB	MWB	MWB
1.1	MWB	MWB	BU	BU	BU
1.15	BU	BU	MB	MB	MB
1.2	BU	MB	MB	MB	MB

some of the key characteristics of marketing-manufacturing interplay within a firm at the tactical as well as the strategic level.

We have assumed that the parameters that specify the effect of mass media as well as word of mouth, and the market potential, are known and fixed. This is a restrictive assumption, especially in light of the fact that our proposed policy requires the knowledge of these parameters. We have also assumed that there is no randomness in the system. In the case when the parameters that determine demand generation are unknown, especially in the setting where the realized demand observed by the firm is corrupted by some stochastic noise, we will need to build adaptive versions of our policy that are able to learn the parameters while attempting to optimally control the system. On a more minor note, we have assumed that the single product that the firm produces and sells is not discrete in nature, but is infinitely divisible, and perform our theoretical analysis in continuous time. We admit that analytical tractability is a consideration in the choice of assumptions. The analysis is considerably challenging even in this simple setting, and since the purpose of the analysis is to serve as a stepping stone towards the development of a more complete theory of marketing-manufacturing interaction, we feel that these assumptions are justified. Relaxing several of the assumptions in this paper will be the subject of future research.

APPENDIX

Proof of Theorem 1. We will apply the maximum principle to prove this result. The reader is advised to refer to Sethi and Thompson (2000), §4.1 and §3.3, for the results we use. Denote the state vector $(N(t), S(t), n(t), I(t), L(t))'$ by $\mathbf{x}(t)$, the control vector $(s(t), x(t))'$ by $\mathbf{u}(t)$, and the running cost $\pi s(t) - \alpha x(t) - wL(t) - hI(t)$ by $\phi(\mathbf{x}, \mathbf{u})$. We can then write (3)–(12) as $\max \int_0^T e^{-\gamma t}\phi(\mathbf{x}, \mathbf{u})\, dt$ subject to the constraints $\dot{\mathbf{x}}(t) = f(\mathbf{x}, \mathbf{u})$ (corresponding to (4)–(8)) with the inequality constraints $g(\mathbf{u}) \geqslant 0$ (corresponding to (11)–(12) and $\mathbf{x}_4, \mathbf{x}_5 \geqslant 0$ (corresponding to (9)–(10)). Define the Hamiltonian of the system as $H(\mathbf{x}, \mathbf{u}, \lambda) = \phi(\mathbf{x}, \mathbf{u}) + \lambda' f(\mathbf{x}, \mathbf{u})$. From §4.1 and §3.3 of Sethi and Thompson (2000), we have

$$H(\mathbf{x}^*(t), \mathbf{u}^*(t), \lambda^*(t)) \geqslant H(\mathbf{x}^*(t), \mathbf{u}(t), \lambda^*(t)), \quad \text{for all} \quad t \in [0, T]$$

for the optimal state $\mathbf{x}^*(t)$ and optimal Lagrange multiplier $\lambda^*(t)$. The control $s(t)$ enters linearly into the Hamiltonian for fixed $\mathbf{x}^*(t)$ and $\lambda^*(t)$. Hence, an extremal choice is optimal for $s^*(t)$. When the coefficient multiplying $s^*(t)$ is negative, it is optimal to choose $s^*(t) = 0$, arrving at (13). When the coefficient is positive, the choice of $s^*(t)$ is not bounded above by an explicit constraint. However, when $I^*(t) = 0$ we must have $\dot{I}^*(t) \geqslant 0$ for the control to be admissible. Hence, we have the implicit constraint $s^*(t) \leqslant x^*(t)$. Similarly, when $L^*(t) = 0$ we have the implicit constraint $s^*(t) \leqslant n^*(t)$. Since a maximal choice is optimal for $s^*(t)$ when the coefficient multiplying it in the Hamiltonian is positive, we arrive at the choices listed in (14). Finally, consider the case when $I^*(t)L^*(t) > 0$ and $s^*(t) > 0$. The implicit constraints listed above do not apply to $s^*(t)$. Hence $s^*(t)$ is unbounded. If this condition persists for any interval of time $[t, t+\delta)$, with $\delta > 0$, it will lead to an infinite profit. Since profit is bounded above by πm, the only sustainable condition is that $s^*(t) = 0$ when $I^*(t)L^*(t) > 0$. □

Proof of Lemma 2. This follows immediately from Theorem 1 because $L^*(t) \equiv 0$ for all $t \geqslant 0$. The continuity of the Hamiltonian yields the result that there are only finitely many switches. □

Proof of Lemma 3. Note that under any policy, since $0 \leqslant S(t) \leqslant m$ for all $t \geqslant 0$, we must have

$$p(m - N(t)) \leqslant \dot{N}(t) = n(t) \leqslant (p+q)(m - N(t)), \quad \text{for all} \quad t \geqslant 0. \tag{50}$$

Thus, we have two exponential bounds on the cumulative demand under any policy,

$$m(1 - e^{-pt}) \leqslant N(t) \leqslant m(1 - e^{-(p+q)t}). \tag{51}$$

From the left-hand inequality of (51), we know that under any policy there exists a time t_f' such that

$$N(t) \geqslant m - \frac{c}{p+q}, \quad \text{for all} \quad t > t_f'.$$

Then, from the right-hand inequality of (50), we obtain the result. □

Proof of Lemma 4. Consider an optimal policy that sells at the capacity c from t_1 to t_0. We will construct another policy for which $s(t) = 0$ for $t_1 \leqslant t \leqslant t_2$. Further, $s(t) = c$ for $t_2 \leqslant t \leqslant t_3$ and $s(t) = n(t)$ for $t \geqslant t_3$, where t_3 is the smallest time beyond which the modified policy can sell all the demand. We show that such a sales plan is better than the supposedly optimal sales plan $s^*(t)$ for which $s^*(t) = c$ for $t_1 \leqslant t \leqslant t_0$. Thus, a policy that makes the c to n transition in its instantaneous sales cannot be optimal.

Consider t_2 sufficiently small such that a feasible sales plan $s(t)$ can be developed so that $s(t) = 0$ for $t_1 \leqslant t \leqslant t_2$ and $N^*(t_2) > N(t_2)$. Now we prove the existence of such a t_2. From the definition of t_1 and the constraints on the control, we know that $S(t_1) = S^*(t_1) = ct_1$. Since the new policy sells zero amount from t_1 to t_2 the feasibility of the policy directly follows. Consider, $\overline{N}(t) = N^*(t) - N(t)$ in the interval $[t_1, t_2]$.

$$\begin{aligned}\frac{\partial \overline{N}}{\partial t} &= \left(p + \frac{q}{m}S^*\right)(m - N^*) - \left(p + \frac{q}{m}S\right)(m - N)\\ &= -p\overline{N} + q(S^* - S) - \frac{q}{m}(S^*N^* - SN)\\ &= -p\overline{N} + qc(t - t_1) - \frac{q}{m}c(tN^* - t_1N)\end{aligned}$$

and

$$\frac{\partial^2 \overline{N}(t)}{\partial t^2} = -p(n^* - n) + qc - \frac{q}{m} c(N^* + tn^* - t_1 n).$$

At $t = t_1$, we have $(\partial \overline{N}/\partial t) = 0$ and $(\partial^2 \overline{N}(t))/\partial t^2 > 0$ since $N^* \leqslant m$. Therefore, there exists a $t_2 = t_1 + \delta$ such that $(\partial \overline{N}/\partial t) > 0$ in the interval $t_1 \leqslant t \leqslant t_2$. Thus, $\overline{N}(t_2) > 0$.

Consider t_3 such that $s(t) = n(t)$ for $t \geqslant t_3$. Now, consider $t_5 \geqslant t_3$ such that $S(t_5) = ct_5$. Such a t_5 must exist because, if not, it would be possible to meet all future demand at a point earlier than t_3. By definition, $S(t_5) \geqslant S^*(t_5)$ because $S^*(t_5) \leqslant ct_5$. Note that $t_3 \leqslant t_0$ because there is more inventory on hand in the modified policy (after time t_1), the time from which all demand is sold must be earlier. As a result, $S(t_3) < S^*(t_3)$. Therefore, there must exist a point t_6, $t_3 \leqslant t_6 \leqslant t_5$ such that $S(t_6) = S^*(t_6)$. If $N^*(t_6) \geqslant N(t_6)$, then we are done because then $N^*(t_6) - S^*(t_6) \geqslant N(t_6) - S(t_6)$, and since no lost sales occur in the modified policy beyond t_3, this implies that S^* cannot be optimal.

Consider t_7, $t_2 \leqslant t_7 \leqslant t_6$ where $N(t_7) = N^*(t_7)$ and $N(t) < N^*(t)$ for all $t \in [t_2, t_7)$. Note if no such t_7 existed, then $N^*(t_6) > N(t_6)$, since $\overline{N}(t_2) > 0$. Now,

$$\frac{\partial \overline{N}}{\partial t} = \left(p + \frac{q}{m} S^*\right)(m - N^*) - \left(p + \frac{q}{m} S\right)(m - N).$$

At $t = t_7$

$$\frac{\partial \overline{N}}{\partial t} = q(S^* - S) - \frac{q}{m}(N)(S^* - S) = q(S^* - S)\left(1 - \frac{N}{m}\right) \geqslant 0.$$

However, $\overline{N}(t_2) > 0$. Therefore, no such t_7 exists and, hence, S^* is not optimal as argued above. □

Proof of Lemma 5. We prove the result by contradiction. Suppose that $t_1 > 0$ for an optimal policy. Now consider a modification of this supposed optimal policy as follows. Consider a $t_2 < t_1$ sufficiently close to t_1 (exactly how close will be made explicit in what follows), and a policy (labeled without the $*$ superscript) that has $s(t) = 0$ for all $t_2 \leqslant t \leqslant t_1$ and that sells $s(t) = n(t)$ *as soon as it possibly can*. Let $t_3 := \inf\{s\colon s(t) = n(t) \text{ for all } t \geqslant s\}$. Now, there must exist a $t_5 > t_3$ under the modified policy such that $S(t_5) = ct_5$. If no such time existed, then it would be possible to start selling $s(t) = n(t)$ before t_3 (since we did not have to build up as much inventory as we did), leading to a contradiction. Now, if $N(t_5) < N^*(t_5)$, then the supposed optimal policy cannot be optimal because $S^*(t_5) \leqslant ct_5 = S(t_5)$ and, hence,

$$\begin{aligned} &N(\infty) - S(\infty) \\ &\quad = N(t_5) - S(t_5) < N^*(t_5) - S^*(t_5) \leqslant N^*(t_0) - S^*(t_0) \\ &\quad = N^*(\infty) - S^*(\infty). \end{aligned}$$

In order to complete the proof, we establish the following claim that

$$N(t) < N^*(t) \quad \text{for all} \quad t \geqslant t_3. \tag{52}$$

First we claim that $N(t_3) < N^*(t_3)$. If not, there exists a $t \in [t_1, t_3]$ such that $\overline{N}(t) = 0$ where, as before, $\overline{N}(t) := N^*(t) - N(t)$. As before, we have

$$\frac{\partial \overline{N}}{\partial t} = -p\overline{N} + q(S^* - S) - \frac{q}{m}(S^* N^* - SN).$$

Now $\overline{N}(t_1) > 0$ because at $t = t_2^+$, we have $\overline{N}(t) = 0$, $\partial \overline{N}/\partial t = 0$ and

$$\frac{\partial^2 \overline{N}}{\partial t^2} > 0.$$

Hence, if we pick t_2 sufficiently close to t_1, we must have $\overline{N}(t_1) > 0$. So if we define $t_4 := \inf\{t \in [t_1, t_3]\colon \overline{N}(t) = 0\}$, we must have at $t = t_4$

$$\frac{\partial \overline{N}}{\partial t} = q\left(1 - \frac{N^*(t_4)}{m}\right)[S^*(t_4) - S(t_4)] < 0.$$

Note that $t_3 \leqslant t_0$ because there is more inventory on hand in the modified policy (after time t_2) and as a result, the time from which all demand is sold must be earlier. Therefore, if * is optimal, then $S^*(t_4) \geqslant S(t_4)$ which contradicts above inequality. So, we must have $N(t_3) < N^*(t_3)$. Now, as before consider $\overline{N}(t)$ for $t > t_3$. Suppose (52) does not hold. Then there exists a $t > t_3$ such that $\overline{N}(t) = 0$. Let $t_6 := \inf\{t > t_3\colon \overline{N}(t) = 0\}$. At $t = t_6$, since $\overline{N}(t_3) > 0$, we must have

$$\frac{\partial \overline{N}}{\partial t} = q\left(1 - \frac{N^*(t_6)}{m}\right)[S^*(t_6) - S(t_6)] < 0,$$

which implies that $S^*(t_6) < S(t_6)$, which contradicts the optimality of the supposed optimal policy because $N(t_6) = N^*(t_6)$ and, hence,

$$\begin{aligned} N(\infty) - S(\infty) &= N(t_6) - S(t_6) < N^*(t_6) - S^*(t_6) \\ &\leqslant N^*(t_0) - S^*(t_0) = N^*(\infty) - S^*(\infty). \end{aligned}$$

Thus (52) holds, and hence, as argued before, the supposed optimal policy cannot be optimal if $t_1 > 0$. This establishes the result. □

Proof of Lemma 6. If (33) is satisfied, then in the unmodified Bass model we must have $N(t) < ct$ for all $t > 0$. Hence, there is always sufficient capacity to meet the demand. Hence, there is no reason to avoid selling and $S^*(t) = N(t)$ for all $t > 0$.

Now consider systems that violate (33) at some $t > 0$. In this case, there will be lost sales, and the modified Bass model applies. We know from the modified Bass model Equation (2) that $N^*(t_0) = m(1 - e^{-pt_0})$. If we define $\overline{N}(t) := N^*(t + t_0) - N^*(t_0)$, the fact that for all $t \geqslant t_f$, $S^*(t) = N^*(t) - N^*(t_0)$ and simple algebra yield

$$\frac{d\overline{N}}{dt} = n^*(t + t_0) = \left(p + \frac{\bar{q}}{\bar{m}} \overline{N}(t)\right)(\bar{m} - \overline{N}(t)),$$

where $\bar{m}$ and $\bar{q}$ are given by (36) and (37), respectively. Now, the optimal choice of t_0 is one such that the inventory resulting from the build-up period is just depleted at the time when it is no longer needed. That time, of course, is t_f, the time after which demand never exceeds production capacity. That is, we require that

$$S^*(t_f) = N^*(t_f) - N^*(t_0) = ct_f.$$

Also, note that at t_f, $n^*(t_f) = c$. Recognizing that the dynamics of $\overline{N}$ are the same as the unmodified Bass model, and thus using Lemma 1, we obtain the result. □

Proof of Lemma 7. Since there are no lost sales, the demand process follows the classical Bass model. The minimum amount of initial inventory required is such that at the time at which the firm runs out of inventory, the instantaneous demand falls to c and remains below c thereafter. That is, the firms cease to need inventory just as it is depleted. From Lemma 1, the time t_2 at which the instantaneous demand falls to c and remains below c from thereon is given by the larger root of the equation

$$c = m\left\{\frac{p(p+q)^2\exp(-(p+q)t_2)}{(p+q\exp(-(p+q)t_2))^2}\right\},$$

since $n(t) = c$ at potentially two time instants in the classical Bass model (see Figure 1). Further, the optimal amount of initial inventory chosen should be such that total production plus this inventory up to time t_2, $i_0 + ct_2$, should equal the cumulative demand until that time, which is given by

$$N(t_2) = m\left\{\frac{1-\exp(-(p+q)t_2)}{1+(q/p)\exp(-(p+q)t_2)}\right\}$$

from Lemma 1. □

ENDNOTES

1. For any $\epsilon > 0$, we can choose T to be so large that $m(1-e^{-pT}) \leqslant \epsilon$ and, thus, ensure that the remaining market potential is smaller than ϵ at time T and, hence, can be ignored.
2. To make the figure easier to understand, inventory and production are not shown in Figure 2. However, they can be inferred from the sales, demand, and waiting customers.
3. As an aside, we note that the choices result in unit production costs being less than the unit selling price for production levels less than 104,740; 1,023.22; and 101.1, respectively.

ACKNOWLEDGMENTS

The authors thank the anonymous associate editor and two referees for their valuable comments on this paper, which have helped in improving the content and presentation of this paper. They thank the following for their comments on this work: Michael Harrison, Evan Porteus, Suresh Sethi, Seenu Srinivasan, Miguel Villas-Boas, Russ Winer, Richard Steinberg, and participants at the Haas Marketing Seminar, UC Berkeley; Management Science and Engineering Seminar at Stanford; and the Manufacturing and Service Operations Management Conference at Ann Arbor. Finally, they thank their research assistant Yuanjing Zhou for performing numerical computations related to this paper.

REFERENCES

Agence France Presse. 1999. Sony Delays PlayStation 2 launch until next March. (September 13).

Bass, F. M. 1969. A new product growth for model consumer durables. *Management Sci.* **15**(5) 215–227.

Business Wire. 1999. ID Four Ltd. increases initial inventory of discover Y2K software. (March 26).

Dolan, R. J., A. P. Jeuland. 1981. Experience curves and dynamic demand models: Implications for optimal pricing strategies. *J. Marketing* **45**(1) 52–63.

Eliashberg, J., R. Steinberg. 1993. Marketing-production joint decision-making. *Handbooks in OR & MS*, Vol. 5. Elsevier Science Publications, Amsterdam, The Netherlands, 827–880.

El Ouaridighi, F., C. S. Tapiero. 1998. Quality and the diffusion of innovations. *Eur. J. Oper. Res.* **106** 31–38.

Feichtinger, G., R. F. Hartl, S. P. Sethi. 1994. Dynamic optimal control models in advertising: Recent developments. *Management Sci.* **40**(2) 195–226.

Griliches, Z. 1957. Hybrid corn: An exploration in the economics of technological change. *Econometrica* **25**(4) 501–522.

Ho, T., S. Savin, C. Terwiesch. 2002. Managing demand and sale dynamics in new product diffusion under supply constraint. *Management Sci.* **48**(2) 187–206.

Horsky, D. 1990. A diffusion model incorporating product benefits, price, income and information. *Marketing Sci.* **9**(4) 342–364.

Jain, D., V. Mahajan, E. Muller. 1991. Innovation diffusion in the presence of supply restrictions. *Marketing Sci.* **10**(1) 83–90.

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

Kurawarwala, A. A., H. Matsuo. 1996. Forecasting and inventory management of short life-cycle products. *Oper. Res.* **44**(1) 131–150.

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.

Mansfield, E. 1961. Technical change and the rate of imitation. *Econometrica* **29**(4) 741–766.

New York Post. 1999. Shortage chips Apple net—Supplier delays will hold results down. (September 21).

Sethi, S. P., G. L. Thompson. 2000. *Optimal Control Theory*, 2nd ed. Kluwer Academic Publishers, Boston, MA.

Simon, H., K. Sebastian. 1987. Diffusion and advertising: The German telephone campaign. *Management Sci.* **33**(4) 451–466.

Sultan, F., J. U. Farley, D. R. Lehmann. 1990. A meta-analysis of diffusion models. *J. Marketing* **27**(1) 70–77.