

Managing Demand and Sales Dynamics in New Product Diffusion Under Supply Constraint

Teck-Hua Ho • Sergei Savin • Christian Terwiesch

The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104

Graduate School of Business, Columbia University, New York, New York, 10027

The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104

hoteck@wharton.upenn.edu • sv30@columbia.edu • terwiesch@wharton.upenn.edu

The Bass diffusion model is a well-known parametric approach to estimating new product demand trajectory over time. This paper generalizes the Bass model by allowing for a supply constraint. In the presence of a supply constraint, potential customers who are not able to obtain the new product join the waiting queue, generating backorders and potentially reversing their adoption decision, resulting in lost sales. Consequently, they do not generate the positive “word-of-mouth” that is typically assumed in the Bass model, leading to significant changes in the new product diffusion dynamics.

We study how a firm should manage its supply processes in a new product diffusion environment with backorders and lost sales. We consider a make-to-stock production environment and use optimal control theory to establish that it is never optimal to delay demand fulfillment. This result is interesting because immediate fulfillment may accelerate the diffusion process and thereby result in a greater loss of customers in the future. Using this result, we derive closed-form expressions for the resulting demand and sales dynamics over the product life cycle. We then use these expressions to investigate how the firm should determine the size of its capacity and the time to market its new product. We show that delaying a product launch to build up an initial inventory may be optimal and can be used as a substitute for capacity. Also, the optimal time to market and capacity increase with the coefficients of innovation and imitation in the adoption population. We compare our optimal capacity and time to market policies with those resulting from exogenous demand forecasts in order to quantify the value of endogenizing demand.

(Marketing-Operation Interface; Bass Diffusion Model; New Product Forecasting; Capacity Planning)

1. Introduction

When introducing a new product, a firm must trade off the cost of supply, including the cost of capacity and inventories, with the revenues from the product’s demand over its life cycle. An important operations decision when launching a new product is the sizing of capacity. Typically, capacity is determined by first specifying an exogenously defined demand trajectory for the new product over time. The question of how

this demand trajectory comes about is often left unanswered (e.g., Fine and Li 1988). Because the demand process is exogenous rather than endogenous to the model, the chosen level of capacity does not affect the demand dynamics.

In contrast to operations literature, marketing research has focused on developing accurate characterizations of the demand process. Specifically, it has long been argued that the demand and sales

of new products in the marketplace follow the patterns of social diffusion processes, similar to those in epidemiology and the natural sciences (see Mahajan et al. 1990 and 2000 for recent overviews). These models enable a firm to characterize the new product's demand process as a function of various internal and external factors (e.g., price, advertising, population characteristics, nature of innovation). They provide the empirical foundation for forecasting demand of a new product over its life cycle. These models, however, assume that the supply of new products is unlimited and never constrained.

Therefore, there is an apparent gap between the two streams of literature. On the one hand, the operations literature has taken the demand process as *given*, searching for the optimal amount of capacity to install. On the other hand, the marketing literature has looked at the demand process assuming that the diffusion process is *never capacity constrained*. This leaves an important question at the interface between the two unanswered: *How does a new product diffuse in the presence of a supply constraint?*

In the presence of a (binding) supply constraint, potential customers who are unable to obtain the new product immediately may either patiently wait for the product, a phenomenon referred to as backordering, or may impatiently abandon the adoption decision, leading to customer losses. In order to generalize the existing diffusion models to include these phenomena, we must distinguish between the demand process and the actual sales process, the latter being bounded by the minimum of the demand and the available supply.

A joint analysis of supply-related decisions and the corresponding demand dynamics also allows us to plan better operationally. For instance, current models of capacity sizing treat the life-cycle demand as given and independent of the actual sales. If, however, as postulated in the marketing literature, past sales do have an impact on future demand, the determination of the optimal capacity sizing requires an *endogenous* characterization of the demand process. Therefore, in addition to providing descriptive characterizations of the constrained demand and sales dynamics, we derive prescriptive results on how to

manage the new product's supply process. Specifically, we analyze how much the firm should invest in capacity and when it should launch the new product.

To determine how much capacity to install, the firm must trade off the cost of backordering and lost customers with the cost of overcapacity. In the presence of a short life cycle, the capacity decision is irreversible (the lead time for adding/reducing capacity is too long to allow for capacity adjustments to occur during the product life cycle.) The phenomenon of short life cycle with capacity shortages resulting from long capacity lead time prevails in high-tech industries, such as semiconductors, video game consoles, and pharmaceutical compounds. In these industries, supply shortages have been repeatedly reported and industry observers have speculated about the magnitude of their impact on life-cycle demand (e.g., Thomke 1999, Pisano 1997). In the absence of a joint analysis of supply-related decisions and demand dynamics, neither a quantification of sales losses nor an appropriate capacity recommendation is possible.

The firm does not have to launch the new product right after the plant is ready for production. In a make-to-stock (MTS) environment, it is possible to delay product launch in order to *preproduce* (to build inventory prior to starting the sales). Many high-tech companies preproduce to ensure a sufficient level of volume at launch. For example, Nintendo recently delayed the launch of GameCube to guarantee enough volume at launch (*Financial Times* April 19, 2001). Similarly, Microsoft postponed the launch of Xbox when they failed to meet the target of 700,000 boxes in initial inventory (*Financial Times* April 10, 2001).

As we will show, preproduction provides a substitute for installing capacity and thereby serves as a less costly mechanism for achieving the same life-cycle sales as with a higher capacity. However, preproduction delays revenue collection and leads to higher inventory costs. In this paper, we determine the optimal time to launch the new product and start the new product diffusion process in order to maximize the life-cycle profits.

Finally, one might argue that it may be optimal to *sell less than what is currently demanded*, even if there is

ample supply available. In the presence of a nonlinear diffusion dynamic with a positive feedback loop, such as the Bass diffusion model (Bass 1969), initially not selling a unit (even at the risk of losing this specific customer) has a desirable effect of slowing down the diffusion process. This leads to a reduced demand peak and thereby avoids a greater customer loss in the future. By characterizing the optimal sales plan we show that delayed demand fulfillment does not maximize life-cycle profits in a Bass-like diffusion environment. The operations decisions discussed above form a hierarchy as illustrated in Figure 1.

This paper makes three contributions to the operations and marketing literature. First, we derive closed-form expressions of demand and sales dynamics in a Bass-like diffusion environment with a supply constraint. To the authors' knowledge, this work is the first to do so. Second, we integrate capacity, time to market, and sales plan into a unified decision hierarchy. These interrelated decisions were treated separately in prior research. Third, we endogenize demand dynamics in determining the optimal capacity in a constrained diffusion environment. Prior research has treated demand exogenously to the capacity sizing decision.

The rest of this paper is organized as follows. Section 2 reviews the relevant literature. Section 3

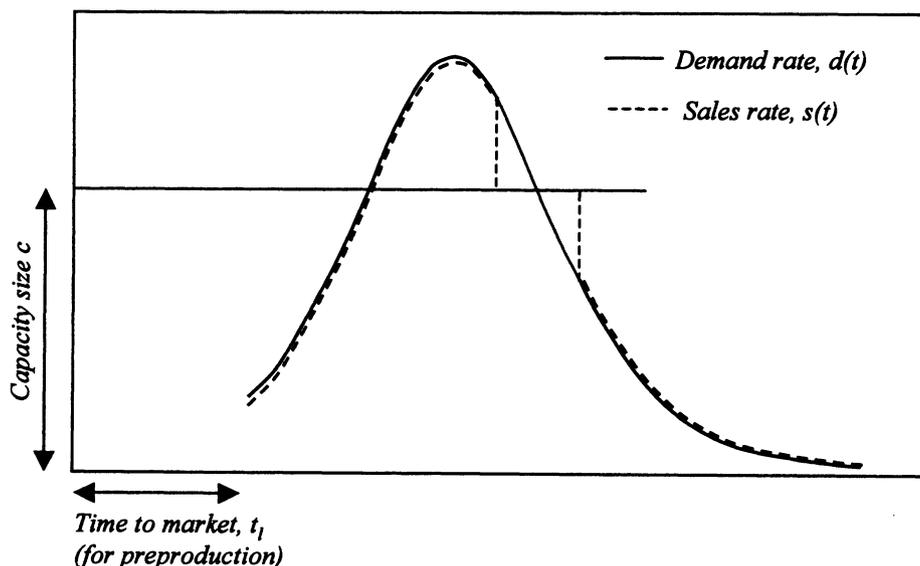
presents the model formulation. We determine the optimal sales plan in §4 and characterize the resulting demand and sales dynamics in §5. In §6 we determine the optimal time to market and capacity and quantify the effect of endogenizing demand on life-cycle profits. Section 7 concludes and suggests future research directions.

2. Related Literature

Our analysis builds on the traditional Bass model of new product diffusion (Bass 1969). The Bass model is recognized for its descriptive and predictive power, and indeed is used widely in marketing to forecast demand of new durable products. It predicts that new product demand is likely to follow specific patterns of social diffusion processes, similar to those in epidemiology and the natural sciences. The Bass model laid the foundation for many articles in marketing (see Mahajan et al. 1990, and 2000 for comprehensive overviews) and, more recently, in research that cuts across marketing and operations (see Fine and Li 1988, Kurawarwala and Matsuo 1996).

The Bass diffusion model posits that the population of potential adopters for a new product is subject to two means of communication: mass-media communication (external influence) and word-of-mouth

Figure 1 The Hierarchy of Decisions in a Constrained New Product Diffusion



communication (internal influence). The external influence affects potential adopters directly, while the internal influence relies on the interaction between customers who have already adopted the product and potential adopters. The Bass model is a mathematical model to capture these effects based on ideas from contagion models in epidemiology.

Although the congruency between the diffusion of a new product and the diffusion of an infectious disease is appealing, it is important to note one fundamental difference between the two. The reproductive capacity of a virus, defined as the number of offsprings that can be generated within one time period, grows proportionally with the diffusion of the disease. Obviously, this is not true for the availability of a new product in a supply chain. In a supply chain, there often exists a maximal production rate defined by the capacity of the plant.

This shortcoming of the Bass model was first addressed by Jain et al. (1991), who studied the diffusion of new telephones in Israel from 1949 to 1987. Waiting times for a new telephone were in excess of three years, and, in the absence of competition, customer losses did not occur. In the Jain et al. formulation, the level of capacity grows with the number of backorders, which may be suitable for a service environment where the lead time to expand capacity is short. Also, their supply constraint is always binding over the entire life cycle of the new product, and hence the sales trajectory is identical to the capacity level. These assumptions do not hold for most manufacturing environments, where customer losses are common, the lead time for changing capacity is long, and the supply constraint is not always binding. Our paper addresses these shortcomings by developing a general model of new product diffusion under supply constraint.

When making supply side decisions, operations managers often assume that the underlying life-cycle demand dynamics of a new product are independent of the product availability (see Luss 1982 for an overview of capacity sizing models). A classical approach to determining capacity under demand uncertainty is to use the newsvendor model, which converts a demand forecast into a supply plan by balancing the costs of excessive capacity with those

of capacity shortages. However, this approach often ignores the nonstationarity in demand inherent in new product diffusion. Addressing this problem, Kurawarwala and Matsuo (1998) present a model of procurement in which the demand process follows a Bass-type diffusion with known parameters of external and internal influence, but with unknown market size. Their model corresponds to an extension of a conventional newsvendor model and provides an example of how procurement policy can be influenced by new product diffusion dynamics. Finally, Fine and Li (1988) provide conditions under which a firm would switch from one supply process to another during the product life cycle. They assume demand dynamics with symmetrical growth and decline stages. The authors show that there are five possible process switching strategies, depending on the relative cost parameters of the processes. Their analysis relies on the assumption that process switching decisions will not influence the underlying demand dynamics; thus, they assume that demand is *exogenous* to the model.

We extend the existing literature by presenting a formal model of a new product's diffusion in the presence of a supply constraint. We thereby introduce important supply-chain phenomena, such as backordering and customer losses, into the field of new product diffusion. This represents the first *joint* analysis of supply and demand dynamics in a new product's supply chain. Building on this analysis, we address the managerial decisions of capacity sizing, time to market, and demand fulfillment policy.

3. Model Formulation

Consider a firm which plans an introduction of a new product. The firm faces a hierarchy of decisions. At the top of this hierarchy lies the capacity sizing decision, which is based on the trade-off between the cost of supply shortages and the cost of overcapacity. In the presence of short life cycles and long lead times for changing production capacity, the selected level of production capacity c remains the same throughout the life cycle of the product. Our analysis can be extended to include a general capacity "trajectory" $c(t)$; however, closed-form solutions would no longer be possible.

We assume the plant will be ready to start production at a known date, which we define as $t = 0$. Given a level of capacity, the firm must decide on the time to market $t_i \geq 0$. Delaying the product introduction may help the company to build inventory and thereby minimize the loss of sales due to insufficient product supply. At the same time, a delayed launch will move revenues further into the future as well as lead to an increase in inventory costs.

Finally, once the diffusion process has started, the firm can decide on how much to sell at each moment in time, which we denote as $s(t)$. In the presence of a nonlinear interaction between the potential adopters and those who already have bought the product, it is unclear whether selling as much as supply would permit is an optimal policy.

After defining the three decisions, namely, how much capacity to install, when to launch the product, and how much to sell at time t , we now describe the demand dynamics of our model. Let m denote the size of the target population of potential adopters.¹ In what follows, we use $D(t)$ and $S(t)$ to denote the cumulative demand and sales of the new product at time t , respectively. Table 1 summarizes our key notations.

At time t , a customer who was previously not ready to adopt may place an order. If the new product is available, the customer receives the product immediately. If not, she can either wait for the new product by joining the waiting list (backordering) or abandon the adoption decision by canceling the order. Consequently, the customer population can be divided into four groups. The first group consists of potential adopters who are not ready to adopt the product yet. The second group are adopters who have placed an order and already have received the new product. The third group are potential adopters on the waiting list and the fourth group are potential adopters who refuse to wait and hence cancel their orders (the so-called "lost" customers). We denote the size of the

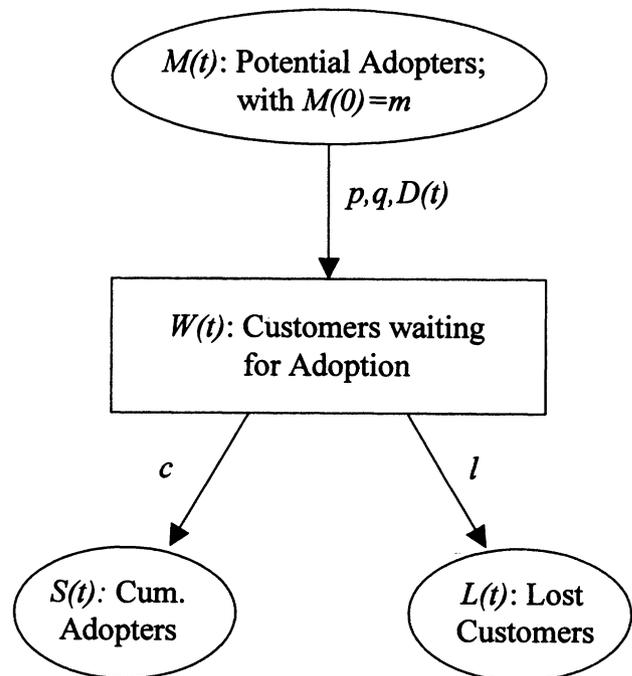
Table 1 Summary of Model Notation

c :	production capacity
t_i :	launch time
$s(t), S(t)$:	sales rate and cumulative sales at time t
m :	initial size of potential adopter population
$d(t), D(t)$:	demand rate and cumulative demand at time t
$W(t)$:	waiting customer population at time t
$L(t)$:	cumulative number of lost customers at time t
l :	rate of loss of waiting customers
p, q :	coefficients of innovation and imitation
$I(t)$:	inventory at time t
$r(t), R(t)$:	production rate and cumulative production at time t
$a(t)$:	profit margin at time t
H :	variable cost of acquiring and maintaining a unit of capacity
h :	unit inventory holding cost (per unit time)

third and fourth group at time t by $W(t)$ and $L(t)$, respectively. Figure 2 shows the interaction between the four customer groups.

At any moment in time, a consumer who is ready to adopt the new product can either join the adopters, the waiting list, or the group of lost customers. Thus,

Figure 2 The Four Customer Groups Under Constrained New Product Diffusion



¹ The variable m can be time dependent if the target population grows or declines over time. It can also vary with a firm's market mix variables such as price and level of advertising expenditure (e.g., Dodson and Muller 1978, Kalish 1985, Bass and Krishnan 1999). For simplicity, we assume a fixed target population.

we have:

$$D(t) = S(t) + W(t) + L(t). \quad (1)$$

If the product supply is unlimited (the firm is never capacity constrained), the waiting list will always be empty and there will be no lost customers. Thus, demand $D(t)$ and sales $S(t)$ are identical. In the presence of a supply constraint, potential adopters who are not able to obtain the product immediately join the waiting list $W(t)$. We assume waiting customers abandon their adoption decisions after, on average, $\frac{1}{l}$ units of time:

$$\frac{dL(t)}{dt} = lW(t). \quad (2)$$

This formulation allows us to capture the demand assumptions made by the existing operations models, namely backordering ($l = 0$) and customer loss ($l = +\infty$), as well as any intermediate case.

The demand process itself, which defines the arrival of customer orders, follows Bass-like dynamics. Thus, the consumer's adoption decision is influenced by two factors: the independent innovation dynamics and the interaction dynamics between adopters $S(t)$ and potential adopters who are still not ready to adopt the new product ($m - D(t)$). This interaction effect is also referred to as "internal influence" or "word of mouth":

$$\frac{dD(t)}{dt} = p[m - D(t)] + \frac{q}{m}S(t)[m - D(t)]. \quad (3)$$

Here, p and q are the coefficients of innovation and imitation, respectively.

By using the Bass model as the demand model, we assume a certain uniqueness of the product to be launched, either in the form of a new brand or a new product category (e.g., movies, video game console, Pentium III). In both cases, one can argue that customer loss can occur because consumers are impatient or engage in cross-brand or cross-category substitution. Note also that our model is sufficiently general to include the case of no customer loss by setting $l = 0$.²

² While Bass's original study estimated the model on data from new product categories (e.g., air conditioners, power lawn mowers), the model has been successfully applied at the level of brands within a category (e.g., Kurawarwala and Matsuo 1998, Sawhney and Eliashberg 1996, Parker and Gatignon 1994, Mahajan et al. 1993).

In order to connect (1), (2), and (3) to the supply process, we consider the cumulative production, $R(t)$, and the inventory of available products, $I(t)$. Note that because we allow the possibility for the firm to select the rate at which it sells, we cannot impose a standard restriction of $I(t)W(t) = 0$. The total production up to time t is either sold or put into inventory:

$$R(t) = I(t) + S(t). \quad (4)$$

The production rate can be expressed as:

$$r(t) = \frac{dR(t)}{dt} = \begin{cases} c, & t < t^*, \\ \frac{dD(t)}{dt}, & t \geq t^*. \end{cases} \quad (5)$$

The company produces at maximum capacity c until the time when demand drops below capacity ($t^* = \min(t \mid dD(t)/dt < c, d^2D(t)/dt^2 < 0)$). During the final phase of the diffusion ($t \geq t^*$) the firm produces according to the demand rate $dD(t)/dt$ in order to avoid unnecessary inventory. As the population of potential adopters, m , is finite, so is t^* for any positive production capacity c .³

For fixed values of production capacity c and launch time t_l , we choose sales rate $s(t)$ to maximize life-cycle discounted profits:

$$P(c, t_l) = \max_{s(t) \geq 0} \left(\int_{t_l}^{+\infty} (a(t)s(t) - hI(t))e^{-\theta t} dt \mid \{I(t_l) = ct_l\} \right), \quad (6)$$

where $a(t) > 0$ is the profit margin of the new product at time t and h is the inventory holding cost (per unit of inventory, per unit of time). The two terms in the objective function correspond to discounted life-cycle revenues and inventory costs, respectively. We observe that the expression for $P(c, t_l)$ can be simplified by shifting the time origin to t_l : $P(c, t_l) = e^{-\theta t_l} \bar{P}(c, t_l)$, where

$$\bar{P}(c, t_l) = \max_{s(t) \geq 0} \left(\int_0^{+\infty} (\bar{a}(t)\bar{s}(t) - h\bar{I}(t))e^{-\theta t} dt \mid \{\bar{I}(0) = ct_l\} \right), \quad (7)$$

³ In the next section we derive explicit expressions for t^* for any combination of p, q, m , and c .

and $\bar{a}(t) = a(t + t_l)$, $\bar{s}(t) = s(t + t_l)$, $\bar{I}(t) = I(t + t_l)$. In our analysis below we will drop the overbars from all these functions, thus, we will write $a(t)$ instead of $\bar{a}(t)$.

Once the optimal selling plan $s^*(t)$ is found, the company has to decide on the launch time $t_l \geq 0$. For a given launch time t_l , the discounted prelaunch inventory costs can be expressed as

$$h \int_0^{t_l} ct e^{-\theta t} dt = \frac{hc}{\theta} \left(\frac{1}{\theta} (1 - e^{-\theta t_l}) - t_l e^{-\theta t_l} \right). \quad (8)$$

Thus, the best launch time t_l , for given capacity c , can be found from

$$P^*(c) = \max_{t_l \geq 0} \left(P(c, t_l) - \frac{hc}{\theta} \left(\frac{1}{\theta} (1 - e^{-\theta t_l}) - t_l e^{-\theta t_l} \right) \right). \quad (9)$$

Finally, the overall production capacity c has to be selected:

$$\max_c = (P^*(c) - Hc), \quad (10)$$

where H denotes the variable cost of acquiring and maintaining a unit of production capacity. The sequence of expressions (6), (9), and (10) reflects the implied hierarchical structure of the company's decisions, reflected by Figure 1. We start by investigating the "tactical" problem (6).

4. Optimal Sales Plan

The tactical decision chooses the sales rate $s(t)$ to maximize profits for fixed values of capacity c and launch time t_l . This problem can be formulated within the optimal control framework as follows:

$$P(c, t_l) = \max_{s(t) \geq 0} \left(\int_0^{+\infty} (a(t)s(t) - hI(t))e^{-\theta t} dt \right)$$

s.t. $\frac{dD}{dt} = d(t), \quad (11)$

$$\frac{dS}{dt} = s(t), \quad (12)$$

$$\frac{d^2D}{dt^2} = \frac{q}{m}s(t)(m - D(t)) - d \left(p + \frac{q}{m}S(t) \right),$$

$$\frac{dL}{dt} = lW(t),$$

$$\frac{dW}{dt} = d(t) - s(t) - lW(t), \quad (13)$$

$$\frac{dI}{dt} = r(t) - s(t), \quad (14)$$

$$I(t), W(t) \geq 0, \quad (15)$$

$$D(0) = S(0) = L(0) = W(0) = 0,$$

$$I(0) = ct_l, d(0) = pm. \quad (16)$$

The first two equations are self-explanatory. The third one is the time derivative of (3), the fourth one is (2), and the last two are time derivatives of (1) and (4), respectively. We note that nonnegativity constraints on $I(t)$ and $W(t)$ imply that $r(t) \geq s(t)$ whenever $I(t) = 0$, and $d(t) \geq s(t)$ whenever $W(t) = 0$. The following result states the optimality of the maximum possible sales rate at any given t :

PROPOSITION 1. *For any profit margin $a(t) > 0$, holding cost $h > 0$, and launch time $t_l \geq 0$ in (16), the optimal sales rate is given by*

$$s^*(t) = \begin{cases} r(t), & W^*(t) > 0, \\ \min(r(t), d^*(t)), & I^*(t) = 0, W^*(t) = 0, \\ d^*(t), & I^*(t) > 0, \end{cases} \quad (17)$$

where $d^*(t)$, $I^*(t)$, and $W^*(t)$ are the optimal values of demand rate, inventory, and waiting pool size, respectively. Also, $I^*(t)W^*(t) = 0$ for all $t \geq 0$.

All proofs are presented in Ho et al. (2001). Proposition 1 suggests that when faced with the choice between selling an available unit immediately versus delaying the sale in order to reduce the degree of future shortages, the firm should always favor the immediate sale. This result is interesting because an immediate demand fulfillment policy will accelerate the new product diffusion process and lead to a higher demand peak, resulting in a greater loss of customers. Proposition 1 shows that this negative effect of customer loss due to demand acceleration is outweighed by the time benefit of immediate cash flow.

This result runs counter to a recent result by Kumar and Swaminathan (2000), who suggest delayed demand fulfillment may be optimal in constrained new product diffusion. They have independently proposed an extension to the Bass diffusion model to include a supply constraint. Their model minimizes lost sales and assumes that limited supply always

results in an immediate loss of unsatisfied demand. We introduce a more general model of new product diffusion, which, in addition to the lost sales, allows for backlogging of demand. In addition, the tactical sales planning in our modeling framework is driven by profit maximization, rather than minimization of lost sales. The use of life-cycle profits as the objective results in the optimality of an immediate demand fulfillment policy.

5. Supply-Constrained New Product Diffusion

In this section, we analyze the diffusion dynamics under the optimal sales plan established above. Our goal is twofold. First, we are interested in specifying the demand and sales dynamics $D(t)$ and $S(t)$ and comparing them to the unconstrained Bass demand dynamics. Second, we would like to obtain the expression for discounted profits (6), which we use in determining the optimal capacity and time to market. Below we provide separate analyses of the cases of patient ($l = 0$) and impatient ($l > 0$) customers. We therefore search for the solution to the system of differential equations (1), (2), (3), (4), (5), and (17) for particular values of production capacity c and launch time t_l subject to the following initial conditions:

$$W(0) = S(0) = L(0) = 0, I(0) = ct_l. \tag{18}$$

5.1. Patient Customers

In the case of patient customers, all unsatisfied orders are backlogged, $L(t) = 0$. The product diffusion is described by:

$$\begin{aligned} D(t) &= S(t) + W(t), \\ R(t) + ct_l &= S(t) + I(t), \\ \frac{dD(t)}{dt} &= p[m - D(t)] + \frac{q}{m}S(t)[m - D(t)], \\ \frac{dR(t)}{dt} &= \begin{cases} c, & t < t^*, \\ \frac{dD(t)}{dt}, & t \geq t^*. \end{cases} \end{aligned}$$

$$\frac{dS(t)}{dt} = \begin{cases} c, & W(t) > 0, \\ \min\left(c, \frac{dD(t)}{dt}\right), & I(t) = 0, \\ \frac{dD(t)}{dt}, & W(t) = 0, \\ & I(t) > 0. \end{cases} \tag{19}$$

with $t^* = \min(t \mid dD(t)/dt < c, d^2D(t)/dt^2 < 0)$. This set of equations is to be solved with the initial conditions $D(0) = S(0) = R(0) = 0$.

Below we analyze the new product diffusion process for any chosen capacity c and launch time t_l . In particular, we show that, depending on these two decisions, the diffusion can exhibit three different regimes. The first regime is observed when capacity and preproduction inventory are sufficiently high, and the presence of the limited production capacity is never felt by the diffusion process. This regime exhibits the classical Bass dynamics. The second regime is observed when the diffusion process begins with an unconstrained phase, then enters a constrained phase for a duration, and finishes with a second unconstrained phase. The third regime is observed when the product is launched immediately ($t_l = 0$) and the capacity c is lower than the initial demand rate. Consequently, the diffusion process starts with a constrained phase and switches to an unconstrained phase at a later point in time.

Regime 1: Unconstrained Diffusion (UD). In this regime, c and t_l are high enough to ensure that $W(t) = 0$ for every t . Our model then reduces to the classical Bass dynamics (3) with $D(t) = S(t)$ and $D(0) = S(0) = 0$. We note that even without preproduction ($t_l = 0$), the presence of limited supply will not constrain the diffusion process, provided that the production capacity is sufficiently high. The smallest capacity level ensuring that the Bass diffusion pattern is preserved is determined as follows. Let us define $\tau_+ = \max(\tau \mid c = d_{\text{Bass}}(\tau))$ as the last time when the Bass demand rate equals to c :

$$c = \frac{pm(q+p)^2 \exp((p+q)\tau_+)}{(q+p \exp((p+q)\tau_+))^2}, \tag{20}$$

so that

$$\tau_+ = \frac{1}{p+q} \ln\left(\frac{q}{p}\right) + \frac{1}{p+q} \ln\left(\frac{1 + \sqrt{1 - \frac{c}{c_0^*}}}{1 - \sqrt{1 - \frac{c}{c_0^*}}}\right), \tag{21}$$

where $c_o^* = m(p+q)^2/4q$ is the maximum demand rate under Bass diffusion. We note that

$$D_{\text{Bass}}(\tau_+) = \frac{m(q-p)}{2q} + \frac{m(p+q)}{2q} \sqrt{1 - \frac{c}{c_o^*}}$$

Then, Bass diffusion is preserved as long as $c\tau_+ \geq D_{\text{Bass}}(\tau_+)$, so that the combination of production and inventory is enough to satisfy the demand at all times. Thus, for $t_l = 0$, the smallest production rate necessary to sustain unconstrained Bass diffusion, $c_s^*(p, q, m)$, is determined as the capacity c satisfying the equation $c\tau_+ = D_{\text{Bass}}(\tau_+)$:

$$c \left(\frac{1}{p+q} \ln\left(\frac{q}{p}\right) + \frac{1}{p+q} \ln\left(\frac{1 + \sqrt{1 - \frac{c}{c_o^*}}}{1 - \sqrt{1 - \frac{c}{c_o^*}}}\right) \right) = \frac{m(q-p)}{2q} + \frac{m(p+q)}{2q} \sqrt{1 - \frac{c}{c_o^*}} \quad (22)$$

It follows that $c_s^*(p, q, m) < c_o^*(p, q, m)$: Because the inventory can be used to satisfy customer orders, the unconstrained diffusion can be preserved even if the production capacity c is smaller than the maximum demand rate in the Bass regime. This observation is illustrated by Figure 3.⁴

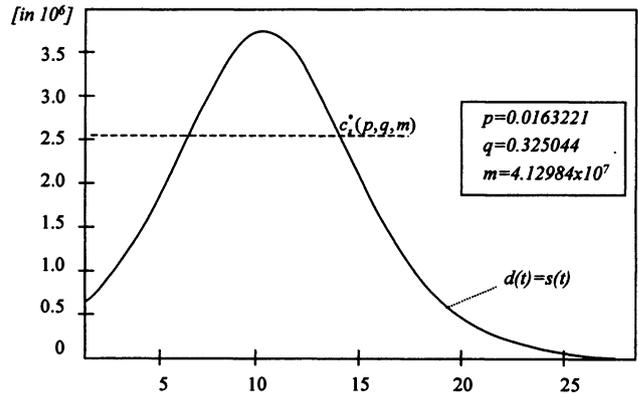
For $c < c_s^*(p, q, m)$, the Bass regime can be sustained only if $t_l > 0$, so that there is additional inventory present. The following statement defines the smallest value of t_l preserving the Bass diffusion regime for each c .

LEMMA 1. For a given level of production capacity c , the new product diffusion dynamics follow the Bass regime if and only if the launch time t_l exceeds the critical level $t_l^*(c)$ given by

$$t_l^*(c) = \begin{cases} 0, & c \geq c_s^*, \\ \frac{m(q-p)}{2qc} + \frac{m(p+q)}{2qc} \sqrt{1 - \frac{c}{c_o^*}} - \frac{1}{p+q} \ln\left(\frac{q}{p}\right), & c < c_s^*. \end{cases} \quad (23)$$

⁴To illustrate the shape of this curve, we use the average values of p, q , and m from Bass (1969).

Figure 3 Unconstrained Diffusion (UD) Under Make-to-Stock Production



The critical launch time is a nonincreasing function of c : $\partial t_l^*(c)/\partial c \leq 0$.

The relation $t_l^*(c)$ defines a critical curve in (c, t_l) space which separates the regions of constrained diffusion and Bass diffusion. Managerially, Lemma 1 provides the level of preproduction that avoids any supply shortages over the entire life cycle.

Regime 2: Initially Unconstrained Diffusion (IUD). According to Lemma 1, for any given level of production capacity c , if the launch delay is long enough the diffusion process will never sense the presence of limited supply of products. Below, we will look at the case when, for a given c , $0 < t_l < t_l^*(c)$. In this case, the prelaunch inventory is insufficient to support Bass diffusion regime over the entire life cycle of the product, and therefore a constrained diffusion will be observed.

Because the finite amount of inventory is available at $t = 0$, it will be possible to sustain an unconstrained Bass diffusion for a finite duration. Consequently, the diffusion process goes through three distinct phases: (1) an initial unconstrained Bass diffusion (UP1), (2) a period of constrained diffusion (CP), and (3) a second unconstrained Bass diffusion (UP2). Below we provide a detailed analysis of each phase. Our main goal is to characterize the switching times between these diffusion phases and to derive demand and sales trajectories.

During UP1, the diffusion dynamics are described by

$$D(t) = S(t) = pm \left[\frac{\exp((p+q)t) - 1}{q + p \exp((p+q)t)} \right],$$

$$W(t) = 0. \tag{24}$$

Demand and sales rates are identical, and both are increasing with time: $s(t) = d(t)$, $ds(t)/dt > 0$. The UP1 lasts until the combination of production and inventory can no longer sustain an unconstrained Bass diffusion. Thus, the ending time of this phase, which we denote as τ_1 , is determined as

$$\tau_1 = \min \left(\tau \mid c(\tau + t_i) = m \left(1 - \frac{q+p}{q+p \exp((p+q)\tau)} \right) \right). \tag{25}$$

At $t = \tau_1$, the constrained phase (CP) begins. During the constrained phase, there are customers waiting ($W(t) > 0$) and the sales rate dS/dt is equal to capacity c . In this phase, the solution to (19) is subject to initial conditions

$$D(\tau_1) = D_1 = pm \left[\frac{\exp((p+q)\tau_1) - 1}{q + p \exp((p+q)\tau_1)} \right] = c\tau_1 + ct_i,$$

$$S(\tau_1) = D_1 \tag{26}$$

is given by

$$D(t) = m - (m - D_1) \times \exp \left[- \left(\left(p + q \frac{D_1}{m} \right) (t - \tau_1) + \frac{qc(t - \tau_1)^2}{2m} \right) \right],$$

$$S(t) = D_1 + c(t - \tau_1),$$

$$W(t) = -c(t - \tau_1) + (m - D_1) \times \left(1 - \exp \left(- \left(p + q \frac{D_1}{m} \right) (t - \tau_1) - \frac{qc(t - \tau_1)^2}{2m} \right) \right). \tag{27}$$

The constrained phase ends at time τ_2 when, for the first time after τ_1 , there are no customers waiting:

$$\tau_2 = \min(t \mid t > \tau_1, W(t) = 0). \tag{28}$$

From (27) we see that τ_2 is finite, since $\lim_{t \rightarrow \infty} W(t) < 0$. We observe that in the constrained phase, the sales rate $s(t) = c$ is constant and, in general, is not equal

to the demand rate $d(t)$. For $t > \tau_2$, the diffusion continues as the unconstrained Bass process (UP2):

$$D(t) = S(t) = m - \frac{(m - D_2)(p+q)}{q - \frac{q}{m}D_2 + (p + \frac{q}{m}D_2) \exp((p+q)(t - \tau_2))},$$

$$W(t) = 0, \tag{29}$$

where $D_2 = D(\tau_2)$. In UP2, demand and sales rates are equal again, and are decreasing functions of time: $s(t) = d(t)$, $ds(t)/dt < 0$. We observe that once Bass dynamics replace the “constrained” diffusion it never “switches” back. Thus, for all $t \geq \tau_2$, and $d(t)$ remains less than c .⁵

Denote by $\tau_B = 1/(p+q) \ln(\frac{q}{p})$ the time of maximum demand rate for Bass diffusion and by τ_1 the switching time between the unconstrained (UP1) and constrained (CP) phases in IUD regime, given by (25). Also, define

$$d_1 = \left(p + \frac{qc(\tau_1 + t_i)}{m} \right) (m - c(\tau_1 + t_i)),$$

$$v = \frac{qc}{m(p + qc(\tau_1 + t_i)/m)^2}.$$

Now we can use (25) and (27) to describe the demand and the sales processes in this regime:

LEMMA 2 (PEAK DEMAND AND SALES RATES). *The maximum demand rate in the IUD regime occurs at*

$$\tau_{\max}^D = \begin{cases} \frac{m}{qc} \left(\sqrt{\frac{qc}{m}} - p - \frac{qc}{m}(\tau_1 + t_i) \right), & \tau_1 < \frac{m}{qc} \left(\sqrt{\frac{qc}{m}} - p \right) - t_i, \\ \tau_1, & \frac{m}{qc} \left(\sqrt{\frac{qc}{m}} - p \right) - t_i \leq \tau_1 < \tau_B, \\ \tau_B, & \tau_1 \geq \tau_B, \end{cases} \tag{30}$$

⁵ Indeed, from the definition of τ_2 , for small ϵ , it follows that $d^2D/dt^2(t = \tau_2 - \epsilon) < 0$, and $d^2D/dt^2(\tau_2 + \epsilon) = d^2D/dt^2(\tau_2 - \epsilon) + q/m(m - D_2)(dD/dt(\tau_2 - \epsilon) - c) < d^2D/dt^2(\tau_2 - \epsilon) < 0$. However, the Bass curve for $d(t)$ has a unique maximum, and d^2D/dt^2 may switch sign only once. Then, from $d^2D/dt^2(\tau_2 + \epsilon) < 0$, it follows that $d^2D/dt^2 < 0$ for all $t \geq \tau_2$, and $d(t)$ remains less than c .

and is equal to

$$d(\tau_{\max}^D) = \begin{cases} \sqrt{v} \exp\left(-\frac{1}{2}\left(1 - \frac{1}{v}\right)\right) d_1, & \tau_1 < \frac{m}{qc} \left(\sqrt{\frac{qc}{m}} - p\right) - t_1, \\ d_1, & \frac{m}{qc} \left(\sqrt{\frac{qc}{m}} - p\right) - t_1 \leq \tau_1 < \tau_B, \\ c_o^*, & \tau_1 \geq \tau_B. \end{cases} \quad (31)$$

The maximum sales rate in IUD regime occurs at

$$\tau_{\max}^S = \begin{cases} \tau_1, & \tau_1 < \tau_B, \\ \tau_B, & \tau_1 \geq \tau_B, \end{cases} \quad (32)$$

and is equal to

$$s(\tau_{\max}^S) = \begin{cases} d_1, & \tau_1 < \tau_B, \\ c_o^*, & \tau_1 \geq \tau_B. \end{cases} \quad (33)$$

Several observations can be made with respect to results of Lemma 2. First of all, for all values of production capacity $\tau_{\max}^S \leq \tau_{\max}^D$, in particular, for $\tau_1 < m/(qc)(\sqrt{qc/m} - p) - t_1$, τ_{\max}^S is strictly less than τ_{\max}^D , while for $m/(qc)(\sqrt{qc/m} - p) - t_1 \leq \tau_1 < \tau_B$, demand and sales rates peak at the same time. More so, not only the peak times, but also the peak values for demand and sales rates coincide under these conditions. Finally, for $\tau_1 \geq \tau_B$, peak times and peak values for demand and sales rates coincide with those for unconstrained Bass diffusion. The properties of diffusion as described in the lemma above are illustrated for the case of $t_1 = 0$ in Figures 4a, 4b, and 4c.

Regime 3: Initially Constrained Diffusion (ICD).

When $t_1 = 0$ and the production capacity c is smaller than the initial rate of the inflow of potential adopters pm , the diffusion initially proceeds in a constrained mode ($W(t) > 0$ for $0 < t < \tau_2$), later (at $t = \tau_2$) replaced by unconstrained Bass process ($W(t) = 0$ for $t \geq \tau_2$). These two phases are similar to the last two phases of the diffusion process for $pm < c < c_o^*(p, q, m)$. In particular, during the initial constrained period, the

Figure 4a Initially Unconstrained Diffusion (IUD), Regime 1

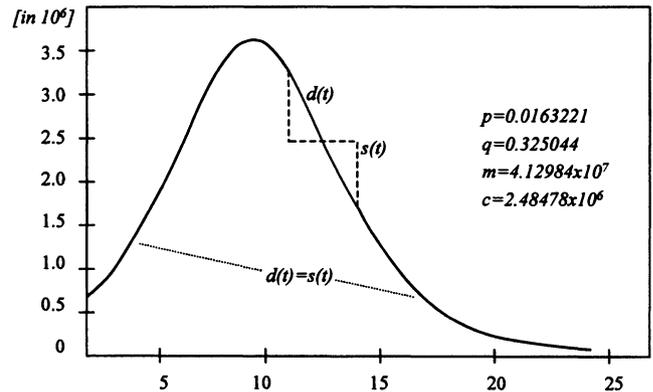


Figure 4b Initially Unconstrained Diffusion (IUD), Regime 2

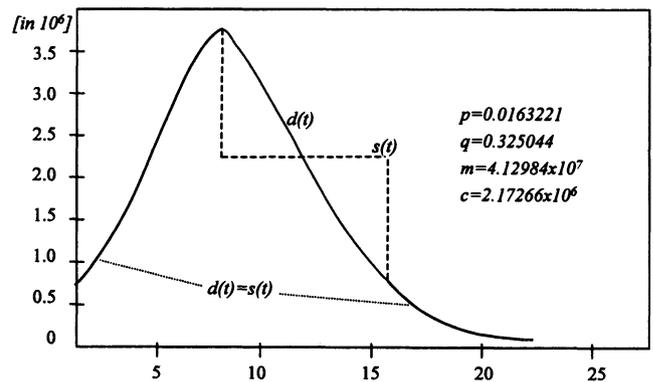
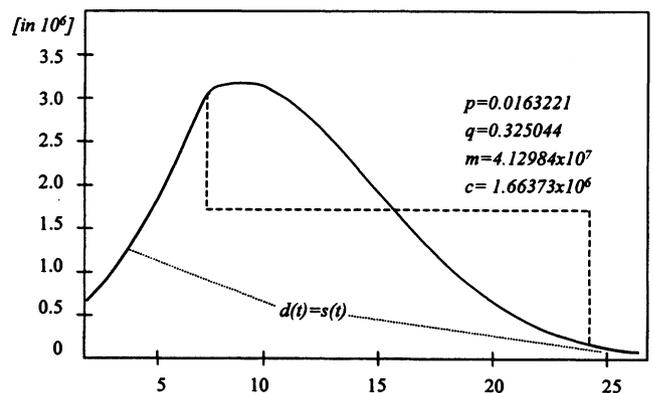


Figure 4c Initially Unconstrained Diffusion (IUD), Regime 3



diffusion dynamics are described by (27) with $\tau_1 = 0$, and is equal to $D_1 = 0$:

$$D(t) = m \left(1 - \exp \left[- \left(pt + \frac{qc}{m} t^2 \right) \right] \right),$$

$$S(t) = ct,$$

$$W(t) = -ct + m \left(1 - \exp \left(- \left(pt + \frac{qct^2}{2m} \right) \right) \right). \quad (34)$$

$$d(\tau_{\max}^D) = \begin{cases} pm, & 0 \leq c < c_0^S, \\ \sqrt{u} \exp \left(-\frac{1}{2} \left(1 - \frac{1}{u} \right) \right) pm, & c_0^S \leq c < pm. \end{cases} \quad (37)$$

The maximum sales rate is equal to c .

The “switching” time τ_2 is defined, as before, by

$$\tau_2 = \min(t \mid t > 0, W(t) = 0). \quad (35)$$

Note that, as in the constrained phase for the IUD regime, the rate of sales $s(t)$ is, in general, different from the demand rate $d(t)$. Similar to Lemma 1, the demand and the sales dynamics in the ICD regime can be described as follows:

LEMMA 3 (DEMAND AND SALES DYNAMICS IN ICD REGIME). Define $c_0^S = p^2m/q$, $u = c/c_0^S$. Then the maximum demand rate in the ICD regime occurs at

$$\tau_{\max}^D = \begin{cases} 0, & 0 \leq c < c_0^S, \\ \frac{1}{p} \frac{\sqrt{u}-1}{u}, & c_0^S \leq c < pm, \end{cases} \quad (36)$$

We note that, unlike the IUD regime, ICD demand and sales rates are very different from the Bass diffusion rates. This result is the reflection of the strongly constraining production capacity in this regime and is illustrated in Figures 3a, 3b, and 3c. In Figure 3a, the demand peak is identical to the Bass demand peak. In Figures 3b–3c, the demand peak is different from the Bass demand peak. While the demand and sales peaks coincide in Figure 3b, they do not in Figure 3c. Comparing these diffusion processes, we note that as the production capacity is decreased, so is the observed peak demand rate.

The properties of the three regimes described above are summarized in Table 2. As these results indi-

Table 2 Demand and Sales Dynamics for the Three Diffusion Regimes

Regime	Initially Constrained Diffusion (ICD)		Initially Unconstrained Diffusion (IUD)		Unconstrained Diffusion (UD)	
Behavior	Capacity is constrained from the beginning; sales fall with no peak	Capacity is constrained right from the beginning; however, there is still a demand peak	Demand peak is different from the Bass demand peak. Demand and sales peak do not coincide	Demand peak is different from the Bass demand peak. Demand and sales peak coincide	Demand peak coincides with the Bass demand peak	Because of the inventory, the traditional Bass diffusion is possible, even if c is below the demand peak
Peak demand rate	pm	$\sqrt{cq}m \cdot e^{-\frac{1}{2} \left(1 - \frac{p^2m}{cq} \right)}$	$(m - c(\tau_1 + t_1)) \sqrt{\frac{qc}{m}} \cdot e^{-\frac{1}{2} \left(1 - \frac{m(p+qc(\tau_1+t_1)/m^2)}{qc} \right)}$	$\left(p + \frac{qc(\tau_1+t_1)}{m} \right) \times (m - c(\tau_1 + t_1))$	$\frac{m(p+q)^2}{4q}$	$\frac{m(p+q)^2}{4q}$
Time of demand peak	0	$\frac{1}{p} \frac{\sqrt{cq/p^2m} - 1}{cq/p^2m}$	$\frac{m}{qc} \left(\sqrt{\frac{qc}{m}} - p \right) - \tau_1 - t_1$	τ_1	$\frac{1}{p+q} \ln \left(\frac{q}{p} \right)$	$\frac{1}{p+q} \ln \left(\frac{q}{p} \right)$
Peak sales rate	c	c	$\left(p + \frac{qc(\tau_1+t_1)}{m} \right) \times (m - c(\tau_1 + t_1))$ Define τ_1 as given by (20)	$\left(p + \frac{qc(\tau_1+t_1)}{m} \right) \times (m - c(\tau_1 + t_1))$ Define τ_1 as given by (20)	$\frac{m(p+q)^2}{4q}$ Define c_s^* as in (17)	$\frac{m(p+q)^2}{4q}$ Define c_s^* as in (17)

cate, the presence of supply constraints in product diffusion may have a significant impact on the position and the heights of the observed peaks in sales and demand. This in turn has substantial implications for the estimation of the diffusion parameters from observed sales and demand.

5.2. Impatient Customers

In the general case, when waiting for the new product makes some customers revise their adoption decision ($l > 0$), sales revenue is not only delayed, but also partially lost. Below we derive sales and demand trajectories and compute the number of lost customers. The solution to the diffusion equations (1), (2), (3), (4), (5), and (17) subject to initial conditions (18) can be described as follows:

PROPOSITION 2. *New product diffusion dynamics subject to customer loss exhibit the same diffusion regimes outlined in Lemmas 1–3. Diffusion dynamics in the unconstrained phases remain unchanged, while constrained phases are now described by*

$$\begin{aligned}
 D(t, l) &= m - (m - D_1) \\
 &\quad \times \exp\left[-\left(\left(p + q\frac{D_1}{m}\right)(t - \tau_1) + \frac{qc}{m}\frac{(t - \tau_1)^2}{2}\right)\right], \\
 S(t, l) &= D_1 + c(t - \tau_1), \\
 W(t, l) &= -\frac{c}{l}(1 - \exp(-l(t - \tau_1))) \\
 &\quad + (m - D_1)\exp(-l(t - \tau_1)) \\
 &\quad \times \left(1 - \exp\left(-\left(\tilde{p}(t - \tau_1) + \frac{qc(t - \tau_1)^2}{2m}\right)\right)\right) \\
 &\quad + (m - D_1)\exp(-l(t - \tau_1)) \\
 &\quad \times \left(l\sqrt{\frac{2\pi m}{qc}}\exp\left(\frac{m\tilde{p}^2}{2qc}\right)\right. \\
 &\quad \times \left(\Phi\left(\sqrt{\frac{qc}{m}}(t - \tau_1) + \sqrt{\frac{m}{qc}}\tilde{p}\right)\right. \\
 &\quad \left.\left. - \Phi\left(\sqrt{\frac{m}{qc}}\tilde{p}\right)\right)\right), \\
 L(t, l) &= D(t, l) - S(t, l) - W(t, l),
 \end{aligned}
 \tag{38}$$

where $\Phi(x) = 1/\sqrt{2\pi} \int_{-\infty}^x \exp(-\frac{s^2}{2}) ds$ is the standard normal CDF, $\tilde{p} = p + q\frac{D_1}{m} - l$, and $D_1 = c(\tau_1 + t_1)$ with τ_1 from (25). The constrained phase ends at

$$\tau_2(l) = \min(t \mid t > \tau_1, W(t, l) = 0). \tag{39}$$

Proposition 2 provides a complete characterization of the diffusion dynamics under supply constraint. Figures 3 and 4a–c show that the demand and sales dynamics generally do not coincide. Note that sales dynamics follow demand dynamics in certain parts of the life cycle and mirror capacity in the remaining parts.

A nice by-product of the above characterizations is that they enable the firm to track the fraction of lost customers at any time. This metric can be used by the firm to quantify the lost market opportunities and to improve capacity planning for future product launches.

In the presence of customer loss, the duration of the constrained phase depends on the loss parameter l :

PROPOSITION 3. *The length of the constrained phase $T_c(l) = \tau_2(l) - \tau_1$ is the smallest positive solution to*

$$\begin{aligned}
 c\frac{\exp(lT_c) - 1}{l} &= (m - D_1)\left(1 - \exp\left(-\left(\tilde{p}T_c + \frac{qcT_c^2}{2m}\right)\right)\right) \\
 &\quad + l(m - D_1)\sqrt{\frac{2\pi m}{qc}}\exp\left(\frac{m\tilde{p}^2}{2qc}\right) \\
 &\quad \times \left(\Phi\left(\sqrt{\frac{qc}{m}}T_c + \sqrt{\frac{m}{qc}}\tilde{p}\right) - \Phi\left(\sqrt{\frac{m}{qc}}\tilde{p}\right)\right)
 \end{aligned}
 \tag{40}$$

with D_1 and \tilde{p} defined in Proposition 2. $T_c(l)$ is a decreasing function of l :

$$\frac{\partial T_c(l)}{\partial l} < 0. \tag{41}$$

Proposition 3 suggests that the duration of the constrained phase decreases as customer impatience increases. If customer impatience reflects the degree of competition in the industry, the length of the constrained phase decreases with the intensity of competition. For example, in Jain et al. (1991), customers wait for three years for the installation of their telephone supplied by a monopolist. Consequently, the

length of the constrained phase is almost the length of the product life cycle. In general, this result indicates that a higher level of capacity and preproduction may be necessary in more competitive industries.

In the case of infinitely impatient customers, any unsatisfied demand is lost, and τ_2 is the earliest time after τ_1 when the demand rate $d(t)$ becomes equal to sales rate $s(t) = c$:

$$T_c(\infty) = \min \left(T \mid T > 0, c = (m - D_1) \left(p + q \frac{D_1}{m} + \frac{qc}{m} T \right) \times \exp \left[- \left(\left(p + q \frac{D_1}{m} \right) T + \frac{qc}{m} \frac{T^2}{2} \right) \right] \right). \quad (42)$$

This result implies that the timings and amplitudes of the demand and sales peaks remain the same as in the case of $l = 0$, and the results presented in Table 2 are fully applicable to the case of impatient customers.

The total fraction of customers lost due to waiting may serve as an important measure of customer service:

PROPOSITION 4. *The fraction of customers lost is given by*

$$f(l) = \frac{(m - D_1) \left(1 - \exp \left[- \left(p T_c(l) + \frac{qc}{m} \frac{T_c^2(l)}{2} \right) \right] \right) - c T_c(l)}{m}, \quad (43)$$

where D_1 is defined in the Proposition 2. $f(l)$ is an increasing function of l :

$$\frac{\partial f}{\partial l} > 0. \quad (44)$$

For the case of infinitely impatient customers, we have

COROLLARY. *The fraction of customers lost for $l \rightarrow \infty$ can be expressed as*

$$f(\infty) = 1 - \frac{D_1}{m} - \frac{c T_c(\infty)}{m} - \frac{c}{pm + qD_1 + qcT_c(\infty)}. \quad (45)$$

Given that in most managerial situations the loss parameter l is not readily available, the expression provided by this corollary may be used as an upper bound estimate on the fraction of customers lost.

6. Optimal Supply Decisions

The above characterizations of demand and sales dynamics allow us to determine the optimal capacity and time to market. We first use these characterizations to develop expressions for the life-cycle profits for given values of capacity and time to market. We then use these expressions for computing the optimal capacity and time to market.

6.1. Life-Cycle Profits

We first turn to expression (6) for the life-cycle profits. For analytical tractability, we consider the case of constant profit margin: $a(t) = a$. For given values of production capacity c and launch time t_l , let

$$\tau_1 = \min \left(\tau \mid c(\tau + t_l) = m \left(1 - \frac{q + p}{q + p \exp((p + q)\tau)} \right) \right), \quad (46)$$

$$D_1 = c(\tau_1 + t_l), \quad (47)$$

and

$$\tau_2 = \tau_1 + T_c \quad (48)$$

where T_c is the duration of the constrained phase, given by the smallest positive solution to (40). Define

$$D_2^* = (m - D_1) \exp \left(- \left(\left(p + \frac{q}{m} D_1 \right) T_c + \frac{qc}{2m} T_c^2 \right) \right) \quad (49)$$

and

$$I(x, y, \theta, p, q, m) = \int_x^y dt \exp(-\theta t) \left(m \left(1 - \frac{q + p}{q + p \exp((p + q)t)} \right) \right). \quad (50)$$

The following result characterizes the life-cycle profits in terms of c and t_l .

PROPOSITION 5. *The life-cycle profits $P(c, t_l)$ can be expressed as*

$$P(c, t_l) = \int_0^{+\infty} (as(t) - hI(t)) e^{-\theta t} dt = (a\theta + h)I(0, \tau_1, \theta, p, q, m) + ac(\tau_1 + t_l) \exp(-\theta \tau_1)$$

$$\begin{aligned}
 & + \frac{ac}{\theta} (\exp(-\theta\tau_1) - \exp(-\theta\tau_2)) \\
 & - \frac{hc(t_i + \frac{1}{\theta})}{\theta} (1 - \exp(-\theta\tau_1)) \\
 & + \frac{hc\tau_1}{\theta} \exp(-\theta\tau_1) + a\theta \exp(-\theta\tau_2) \\
 & \times I\left(0, +\infty, \theta, p + \frac{q}{m}c(\tau_2 + t_i), \frac{q}{m}D_2^*, D_2^*\right). \quad (51)
 \end{aligned}$$

We observe that in spite of complex appearance, the computation of life-cycle profits reduces to evaluating several expressions (including two easily computable one-dimensional integrals) containing switching times τ_1 and τ_2 . Both of these switching times are expressed through the solutions to simple transcendental equations. Their values are easily computed numerically. Below we present the results of a numerical study focused on computing the optimal values of capacity c and time to market t_l .

6.2. A Numerical Study

We conduct a numerical study to compute the optimal time to market for a given value of capacity c . This analysis is particularly relevant for situations where capacity can only be increased in big chunks (e.g., building an additional production facility). We substitute Equation (51) into Equation (9) and use the resulting expression to find the optimal time to market t_l .

Optimal Time to Market. We define the relative innovation factor of a diffusion as the ratio of its coefficient of innovation (p) and the average coefficient of innovation reported in Bass (1969) ($p_{ave} = 0.01632$). Similarly, we define the relative imitation factor of a diffusion as the ratio of its coefficient of imitation with respect to its average value ($q_{ave} = 0.3250$). Figures 5a–5b show how the optimal time to market, t_l , varies with the innovation and imitation factors for three different levels of capacity: $c = 25\%, 50\%, 75\%$ of $c_s^*(p, q, m)$. The discounting factor θ and the loss parameter l were set at 0.01 and 0.1 respectively. We observe that for a fixed value of capacity, the optimal time to market increases with both the innovation and imitation factors. This increase is more dramatic for lower levels of capacity.

Figure 5a The Optimal Production Delay t_l for Fixed Production Capacity As a Function of Innovation Parameter p ($p_{ave} = 0.0163221, q = 0.325044, m = 4.12984 \times 10^7, \theta = 0.01, l = 0.1, h = 0.001$)

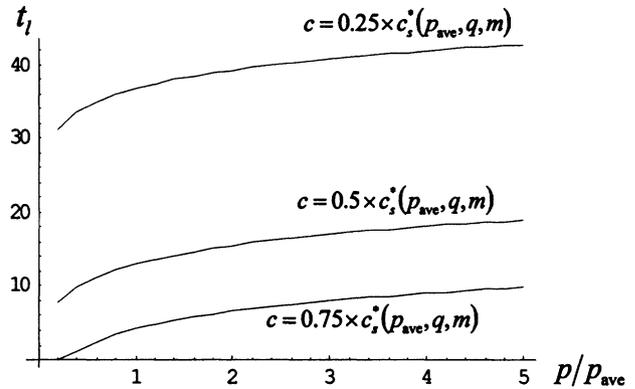
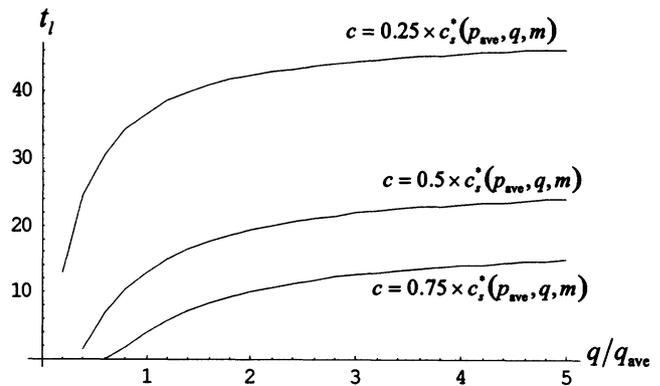


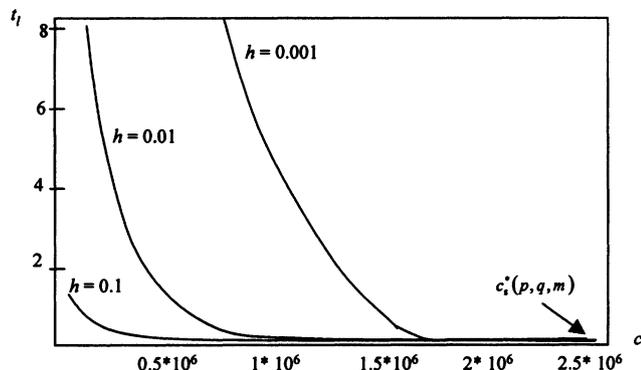
Figure 5b The Optimal Production Delay t_l for Fixed Production Capacity As a Function of Imitation Parameter q ($p = 0.0163221, q_{ave} = 0.325044, m = 4.12984 \times 10^7, \theta = 0.01, l = 0.1, h = 0.001$)



A comparison of Figures 5a and 5b reveals that the optimal time to market is more sensitive to imitation than innovation factors. We believe this is due to the nonlinear effect of imitation on the sales process. This result implies that it is more important to obtain a precise estimate for q than for p . Since prior research suggests that q is more seriously biased by ill-conditioned data than p (Van den Bulte and Lilian 1997, Van den Bulte 2000), the importance of obtaining a precise estimate for q cannot be overemphasized.

Figure 6 plots $t_l^{opt}(c)$ for three different values of inventory holding cost: $h = 0.001, 0.01, 0.1$. The discounting factor θ and the loss parameter l were set

Figure 6 Optimal Values of the Production Delay t As a Function of the Production Capacity c for Different Values of the Inventory Holding Cost h ($p = 0.0163221, q = 0.325044, m = 4.12984 \times 10^7, \theta = 0.001, l = 0.001$)



at 0.001 and 0.001, respectively. We observe that for a fixed value of inventory holding cost, the optimal time to market shortens as the production capacity is increased. Thus, prelaunch inventory and production capacity play the roles of substitutes in constrained new product diffusion. A comparison of the $t_i^{opt}(c)$ curves for different values of h shows that as the value of the inventory holding cost increases, the optimal time to market decreases for the same level of production capacity, resulting in lower inventory.

Our results suggest that firms may want to substitute capacity with preproduction by delaying product launch. This is particularly relevant if the capacity is costly to acquire and if the word-of-mouth effect is dominant, leading to a high demand peak. Industry examples where word-of-mouth effect is dominant include high-technology products with network externalities as well as products with high fashion contents (Van den Bulte 2000). The impact of insufficient preproduction can be dramatic, as illustrated by the recent introduction of the Sega Dreamcast video game console (Thomke 1999). Due to failure to use preproduction to meet initial demand (which led to a slow diffusion of the new product), Sega was forced to withdraw the product prematurely.

Optimal Capacity Size. If the firm does not want to incur any supply shortage, the minimal level of capacity without preproduction is $c_s^*(p, q, m)$. This value can be used as the upper bound for the capacity investment under constrained new product diffu-

sion. Once the optimal time to market is established, (10) can be used to determine the optimal production capacity level c^{opt} . In this numerical study, the values of c^{opt} were computed through a one-dimensional search on a capacity interval $[0, c_s^*(p, q, m)]$.

Figure 7 shows how c^{opt} varies with the innovation and imitation factors. As shown, the optimal capacity increases with both the innovation and imitation factors. Interestingly, the optimal capacity exhibits a clear saturation effect as the speed of diffusion increases.

Figure 8 plots c^{opt} as a function of capacity cost H for three different values of inventory holding cost: $h = 0.001, 0.05, 0.5$. As expected, c^{opt} is a decreasing function of H . In particular, high cost of capacity

Figure 7a Optimal Production Capacity As a Function of Innovation Parameter p ($p_{ave} = 0.0163221, q = 0.325044, m = 4.12984 \times 10^7, \theta = 0.05, l = 0.1, h = 0.001, H = 4$)

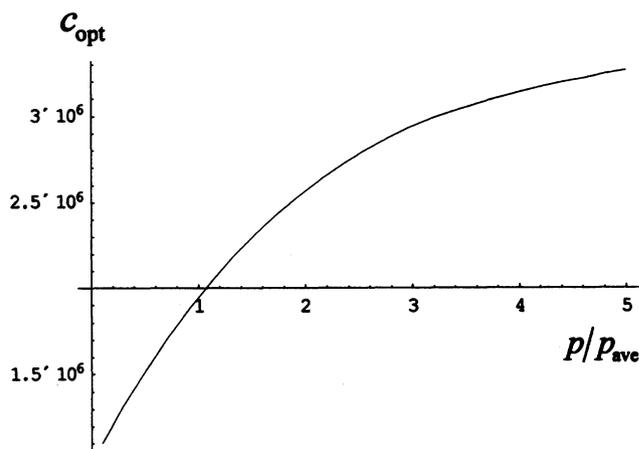


Figure 7b Optimal Production Capacity As a Function of Imitation Parameter q ($p = 0.0163221, q_{ave} = 0.325044, m = 4.12984 \times 10^7, \theta = 0.05, l = 0.1, h = 0.001, H = 4$)

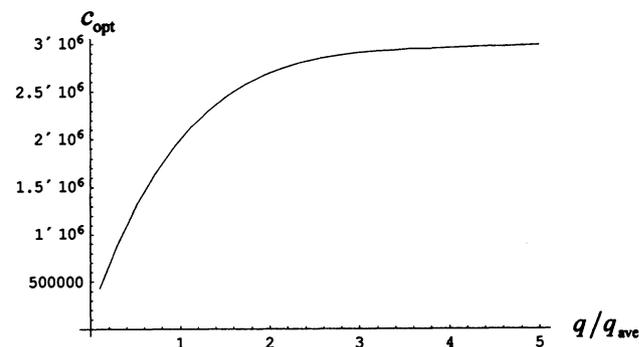
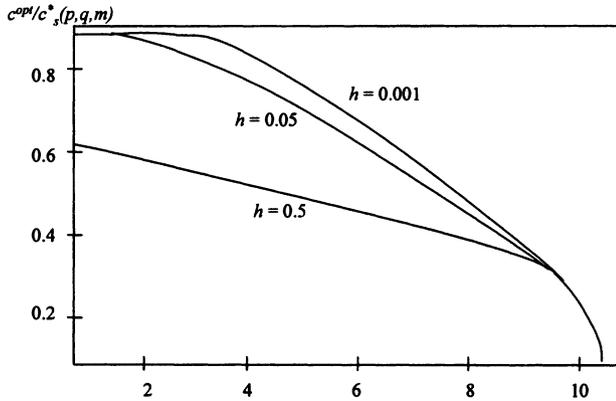


Figure 8 Optimal Production Capacity Values c^{opt} As a Function of the Capacity Holding Cost H for Different Values of the Inventory Holding Cost h ($\rho = 0.0163221$, $q = 0.325044$, $m = 4.12984 \times 10^7$, $\theta = 0.1$, $l = 0.1$)



forces the system to operate in the low production capacity regime, resulting in low profit values. Also, higher inventory costs push the optimal inventory levels down and result in lower optimal production capacities for the same level of capacity cost. When H is negligibly small, high inventory cost will result in an optimal production level that is much lower than $c_s^*(p, q, m)$.

Value of Endogenizing Demand. We can determine the value of endogenizing demand by comparing the optimal profits with the profits obtained under the assumption that the demand dynamics follow the

original Bass dynamics. This latter assumption we will label as “Bass heuristic.” Under the Bass heuristic, the life-cycle profits will still be expressed by (51); however, the values of τ_2 and D_2^* should be computed differently.

LEMMA 4. Let

$$D_{\text{Bass}}(t) = m(1 - (q + p)/(q + p \exp((p + q)t))).$$

Then, under the Bass heuristic, the value of the “switching” time τ_2 is the smallest solution to

$$D_{\text{Bass}}(\tau_2) = \exp(-l(\tau_2 - \tau_1))D_{\text{Bass}}(\tau_1) + \left(m + \frac{c}{l}\right)(1 - \exp(-l(\tau_2 - \tau_1))) - lm(p + q) \int_{\tau_1}^{\tau_2} \frac{\exp(l(u - \tau_2)) du}{q + p \exp((p + q)u)}, \quad (52)$$

such that $\tau_2 > \tau_1$, where τ_1 is given by (46). Also, $D_2^* = D_{\text{Bass}}(\tau_2)$.

Using (52) and (51), we can compute the overall profits under the Bass heuristic for any value of production capacity and establish the value of production capacity $c_{\text{Bass}}^{\text{opt}}$, which maximizes (10) computed under the Bass heuristic.

We can study the value of endogeneity as a function of the diffusion characteristics. Figures 9a–9b show the corresponding results. First, the value gained by endogenizing demand can be significant. In our

Figure 9a The Relative Performance Gap Between the Optimal Profits and Profits from the Bass Heuristic As a Function of Innovation Parameter ρ ($\rho_{\text{ave}} = 0.0163221$, $q = 0.325044$, $m = 4.12984 \times 10^7$, $\theta = 0.05$, $l = 0.1$, $h = 0.001$, $H = 8$)

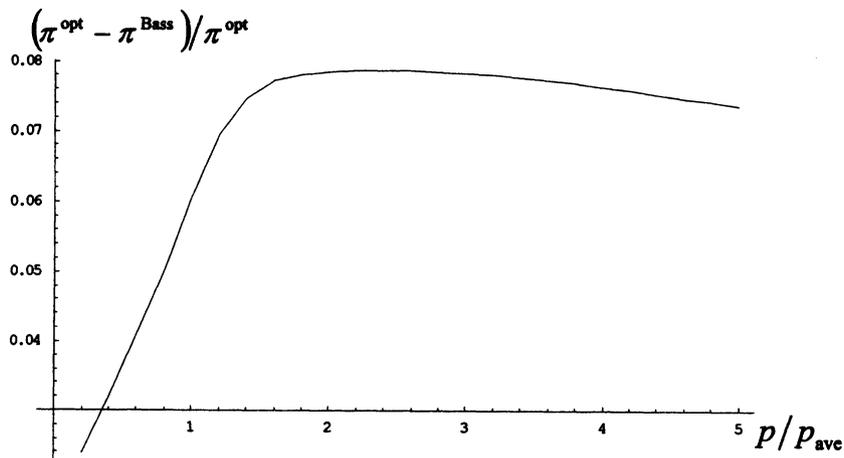
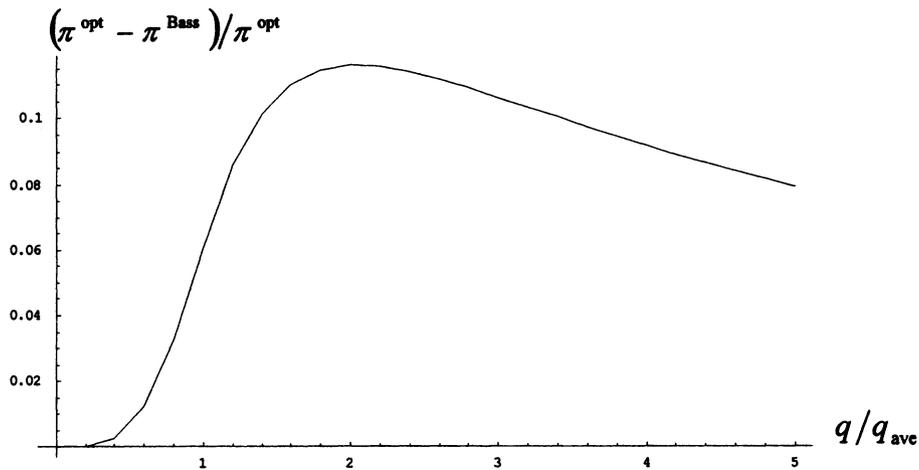


Figure 9b The Relative Performance Gap Between the Optimal Profits and Profits from the Bass Heuristic As a Function of Imitation Parameter q ($\rho = 0.0163221, q_{ave} = 0.325044, m = 4.12984 \times 10^7, \theta = 0.05, l = 0.1, h = 0.001, H = 8$)

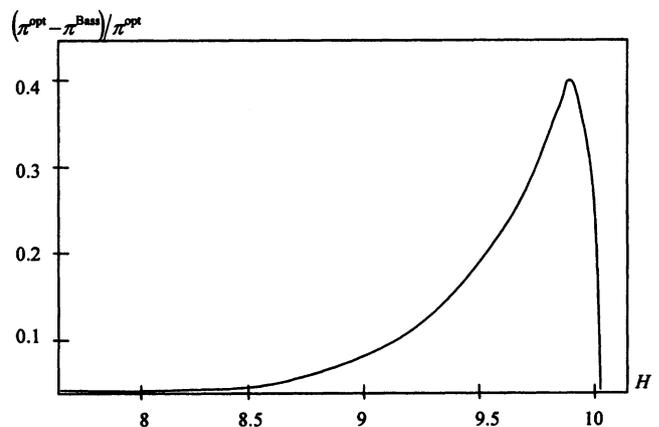


numerical example, the saving is 6% if the innovation and imitation factors are both equal to one. Second, the figures reveal an interesting qualitative result. The value of endogeneity first increases and reaches a peak and then decreases for both the innovation and imitation factors. For a slow rate of diffusion (small innovation and imitation factor), the optimal demand dynamics are less likely to be constrained (they are more like original Bass dynamics), so the value of endogeneity is small. When the rate of diffusion is large, the product life cycle is compressed and the optimal demand dynamics are heavily constrained. In such cases, a large fraction of customers will be lost. Put differently, there is no useful information to be gained in the slow diffusion rate and it is too expensive to act on the useful information when the diffusion rate is high.

Figure 10 graphs the relative difference in profits (10) computed at c^{opt} and c_{Bass}^{opt} as a function of capacity maintenance cost H for $\theta = l = 0.1, h = 0.001$ (profit values π^{Bass} under the Bass heuristic were computed by using the life-cycle profit expression (51) with $c = c_{Bass}^{opt}$ and τ_2 and D_2^* given by (48) and (49), respectively).

We observe that the fraction of profit lost due to the use of the exogenous model of demand dynamics can be rather high for intermediate and high values of capacity costs. In these high cost scenarios, the diffusion occurs in the regime where the capacity is severely constrained. As cost of capacity

Figure 10 The Relative Performance Gap Between the Optimal Profits and Profits from the Bass Heuristic As a Function of the Capacity Holding Cost H ($\rho = 0.0163221, q = 0.325044, m = 4.12984 \times 10^7, \theta = 0.1, l = 0.1, h = 0.001$)



decreases, the optimal capacity increases so that the degree of capacity constraint diminishes. As a result, both exogenous and endogenous models of demand dynamics result in similar optimal capacity levels.

7. Discussion

In this paper we provide a joint analysis of demand and sales dynamics in a constrained new product diffusion. Our analysis generalizes the Bass model to include backordering and customer losses. In addition, we determine the diffusion dynamics when the firm actively chooses supply-related decisions to

influence the diffusion process. We derive closed-form expressions for the optimal diffusion dynamics (both sales and demand) and show how the timing and the amplitude of the peak demand rate differ from that of the Bass model.

Our results suggest that it is important to include supply constraints in the estimation of diffusion parameters such as the coefficients of innovation (p) and imitation (q). An estimation which assumes the Bass model, despite the occurrence of supply shortages during life-cycle, is likely to lead to biased estimates of parameters. Consequently, demand forecasts based on these estimated parameters could be systematically biased as well.

In addition to characterizing the resulting diffusion dynamics in the presence of supply constraint, we investigate how supply-related decisions such as capacity sizing and time to market may interact. We show that an increase in the amount of preproduction (by delaying the product launch) can act as a substitute for capacity. This substitution strategy can be particularly relevant when incremental changes in capacity are prohibitively expensive.

We also analyze how optimal time to market and capacity vary with the diffusion parameters. We show that both the timing and capacity are more sensitive to the coefficient of imitation q than to the coefficient of innovation p , suggesting a need for a precise estimate for the former. In addition, the optimal capacity exhibits a saturation effect as the speed of the diffusion increases.

Finally, we show that the value of endogenizing demand in determining supply-related decisions can be substantial. This is so because the diffusion process depends on the amount of capacity in place. The link between capacity and diffusion dynamics is particularly important when word-of-mouth effects create a causal link between the past and the future sales. Thus, using an exogenous characterization of demand to determine capacity can be suboptimal in such situations.

Our model allows managers to improve their operations decision making in three ways. First, our characterizations of the constrained new product diffusion dynamics can be used to develop more accurate forecasts of demand. This improved accuracy will

lead to more informed decisions, resulting in higher profits. Second, this paper highlights the importance and benefits of endogenizing demand. This observation challenges the standard assumption that demand forecasts merely serve as inputs to operations planning processes and are not affected by supply decisions. Third, our results suggest it is optimal to preproduce and have an initial inventory serve as a substitute for capacity, if new product diffusion does not begin before product launch. This may explain why many high-tech firms choose to preproduce before product launch.

Our model of supply-constrained diffusion opens up several avenues for future research:

- *Estimation of diffusion parameters:* Our model suggests that estimation of diffusion parameters p , q , and m may be significantly biased if the supply to the diffusion process is constrained. The extent to which these diffusion parameters are biased can be easily studied by simulating sales and demand data from a constrained process and using usual estimation procedures to estimate them as if the process is unconstrained. Moreover, the expression for the fraction of customers lost over the life cycle, $f(l)$, can be used to estimate the total number of lost customers. We believe this will make product diffusion models more realistic and hence more applicable.

- *Using marketing mix variables to influence diffusion:* The firm can also use marketing mix variables such as price and advertising to influence the diffusion process. Prior studies have investigated these effects but without considering supply constraint (e.g., Kalish 1985). It will be interesting to investigate how the presence of supply constraint affects the determination of these marketing-mix variables.

- *Waiting time dynamics:* Our results can be used in future research related to customer service metrics, such as the average lead time a customer must wait before she receives the new product.

In conclusion, this paper enables a deeper understanding of the interaction between supply and demand in the adoption of new products and services. We hope our work will be the beginning of a larger stream of work that endogenizes new product demand in order to enhance operations management decisions.

Acknowledgments

This research is partially supported by a grant from the Wharton-SMU Research Center. The authors would like to thank Roger Bohn, Christophe Van den Bulte, and two anonymous reviewers for their helpful comments. They are also grateful for comments received from seminar participants at the INFORMS 99 meeting in Philadelphia, NYU, and INSEAD.

References

- Bass, F. M. 1969. A new product growth model for consumer durables. *Management Sci.* 15(5) 215–227.
- , T. Krishnan. 1999. Modeling the marketing mix influence on single product diffusion. Presented at New Product Diffusion Conference, The Wharton School, University of Pennsylvania, Philadelphia, PA.
- Dodson, J. A., E. Muller. 1978. Models of new product diffusion through advertising and word-of-mouth. *Management Sci.* 24(11) 1568–1578.
- Fine, C., L. Li. 1988. Technology choice, product lifecycles, and flexible automation. *J. Manufacturing Oper. Management* 1(4) 372–396.
- Ho, T.-H., S. Savin, C. Terwiesch. 2001. Managing demand and sales dynamics in new product diffusion under supply constraint. Working paper, The Wharton School, University of Pennsylvania, Philadelphia, PA. (<http://hops.wharton.upenn.edu/people/faculty/ho.html>).
- Jain, D., V. Mahajan, E. Muller. 1991. Innovation diffusion in presence of supply restrictions. *Marketing Sci.* 10(1) 100–113.
- Kalish, S. 1985. A new product adoption model with price, advertising and uncertainty. *Management Sci.* 31(12) 1569–1585.
- Kumar, S., J. Swaminathan. 2000. Diffusion of innovation under supply constraint. Working paper, Graduate School of Business, Stanford University, Stanford, CA.
- Kurawarwala, A., H. Matsuo. 1996. Forecasting and inventory management of short lifecycle products. *Oper. Res.* 44(1) 131–150.
- Luss, H. 1982. Operations research and capacity expansion problems: a survey. *Oper. Res.* 30(5) 907–945.
- Mahajan, V., E. Muller, F. M. Bass. 1990. New product diffusion models in marketing: a review and directions for research. *J. Marketing* 54 1–26.
- , —, J. Wind. 2000. *New Product Diffusion Models*. Sage, Thousand Oaks, CA.
- , S. Sharma, R. Buzzell. 1993. Assessing the impact of competitive entry on market expansion and incumbent sales. *J. Marketing* 57 39–52.
- Parker, P., H. Gatignon. 1994. Specifying competitive effects in diffusion models: an empirical analysis. *Internat. J. Res. Marketing* 11 17–39.
- Pisano, G. 1997. *The Development Factory*. HBS Press, Boston, MA.
- Sawhney, M., J. Eliashberg. 1996. A parsimonious model for forecasting gross box-office revenues of motion pictures. *Marketing Sci.* 15 113–131.
- Sethi, S. P., G. L. Thompson. 2000. *Optimal control theory: applications to management science and economics*. Kluwer Academic Publishers.
- Simon, H., K. H. Sebastian. 1987. Diffusion and advertising: the German telephone campaign. *Management Sci.* 33(4) 451–466.
- Thomke, S. 1999. Project dreamcast: serious play at Sega Enterprises. Harvard Business School Case 9-600-028, Cambridge, MA.
- Van den Bulte, C. 2000. New product acceleration: measurement and analysis. *Marketing Sci.* 19(4) 366–380.
- , G. Lilien. 1997. Bias and systematic change in the parameter estimates of macro-level diffusion models. *Marketing Sci.* 16(4) 338–353.

Accepted by Christopher S. Tang; received April 2000. This paper was with the authors 9 months for 3 revisions.