Optimal Marketing
Entry Timing for Successive Product/Service Generations
Optimal Marketing Entry Timing for Successive Product/Service Generations

Zhengrui Jiang*

Dipak C. Jain**

* Thome Professor in Business, Associate Professor of Information Systems, Department of Supply Chain and Information Systems at College of Business at College of Business, Iowa State University 2340 Gerdin Business Building Ames, IA 50011-1350, USA. Email: zjiang@iastate.edu

** INSEAD Chaired Professor of Marketing and Dean of INSEAD, Boulevard de Constance 77305 Fontainebleau Cedex, France. Email: dipak.c.jain@insead.edu

A Working Paper is the author’s intellectual property. It is intended as a means to promote research to interested readers. Its content should not be copied or hosted on any server without written permission from publications.fb@insead.edu

Find more INSEAD papers at http://www.insead.edu/facultyresearch/research/search_papers.cfm
Abstract

When to introduce successive generations of products or services to the market is an important decision for firms. In this study, we develop analytical models to help decide the optimal market entry timing for each generation. Our analysis is very comprehensive in that we consider two commonly observed revenue streams (i.e., revenue is generated from either *one-time sale* or *continuous service*) and two generation transition strategies (i.e., older generation are either gradually *phased out* or *totally replaced*). Pioneering research models by Wilson and Norton (1989) and Mahajan and Muller (1996) have considered the specific business scenario corresponding to one-time sale and phase-out transition. Results of these studies conclude that a second generation should be introduced *now or never* (Wilson and Norton 1989) or *now or at maturity* (Mahajan and Muller 1996). We find that under one-time sale and phase-out transition, the optimal entry timing is not limited to *now or never or now or at maturity*, but it can also lie between *now* and *maturity*. Interestingly, when continuous service (instead of one-time sale) is the source of revenue and the new generation is less profitable than the older per unit time’s service, the *now or never* pattern is observed. Lastly, we demonstrate that the proposed modeling framework can be easily used to derive market entry timing for more general business scenarios where a) revenue is generated from both one-time sale and continuous service, and b) there exist three product generations, which also extend previous research.

Keywords: *multi-generation diffusion, market entry timing, business revenue models*
Optimal Marketing Entry Timing for Successive Product/Service Generations

1. Introduction

Most of the products and services we consume today represent improved versions of earlier generations, and today’s products will eventually be substituted by even newer generations. The development of successive product generations is driven by both technological advancements and evolving customer needs and preferences. The continuous enhancement of existing technologies and/or the emergence of new technologies make product improvement possible. Further, in a competitive market, firms are often forced to constantly improve their products in order to stay competitive. Even for monopolist firms, facing a relatively constant population, they often rely on new product generations to generate cross-generation repeat purchases. Therefore, continuous product improvement in the form of successive release of product generations is frequently observed in the marketplace. Well known examples including Apple Inc.’s iPhone and iPad, and Microsoft’s Office and Windows lines of products.

Market entry timing is an important strategic decision for firms. The decision is even more critical when introducing successive generations over time. This is mainly because product diffusion dynamics are more complex when there are multiple generations. Older generations can help speed up the diffusion of a new generation because existing adopters may pay more attention to the new generation and consider switching to it. For instance, it is reported that 70% of the first-day buyers of iPhone 4 are users of earlier generations of iPhone (Hughes 2010). On the contrary, the introduction of a new generation can cannibalize the sale of older generations. This can potentially affect the firm’s total revenue, because it decreases the number of cross-generation repeat purchases. Given that a firm’s goal is to maximize total profit from all product

---

1 Our analyses are applicable to physical products, digital products, and services. For brevity, we often use the term product to represent multiple forms of products and/or services.
generations, the cross-generation diffusion dynamics need to be adequately considered when making important decisions such as market entry timing.

Market entry strategy for new products has been studied in the marketing literature. Some prior studies have only focused on one generation and analyzed the tradeoffs between early market entry and product maturity. For instance, Kalish and Lilien (1986) develop a model that incorporates the impact of product quality on the probability of adoption, and then use it to help decide the optimal introduction time for a new product. Since these prior studies consider only one product generation, their research focus and findings are different from ours.

Another stream of research deals with two product versions, one with a high profit margin and the other with a low margin (Moorthy & Png 1992; Prasad, Bronnenberg, & Mahajan 2004). These two prior studies have developed analytical models to help determine the market entry timing for the low-margin version. There are some important differences between the current research and these prior studies. First, both studies assume that the first product version is of high quality and the second one has a low quality (e.g., hardcover and paperback textbooks); while in the present study, later versions offer improved quality compared to earlier versions (e.g., standard and high definition TVs). Second, Prasad at el.’s model is specifically designed for the movie industry, and it considers a sales function that decreases monotonically over time for the primary product (i.e., theatrical release); whereas the focus of the present research is on products or services that exhibit a bell-shaped diffusion pattern (e.g., standard definition color TVs). Finally, neither of the prior research considers the diffusion dynamics across product versions (e.g., how adopters skip an old generation to adopt a new generation or make an upgrade to the new generation) and over time (i.e., how the rates of different adoptions change over time), while such diffusion patterns are an integral part of this research.
To derive the optimal market entry timing for successive product generations, we need to understand how revenues are generated from the entire product lines. We consider two practices observed in the marketplace. First, some firms charge a one-time price when transferring the ownership of a product to a customer and no additional revenue is generated from the same customer afterwards. Many consumer products fall under this scenario. Examples include TV sets, computers, and packaged software such as Microsoft Office. We refer to this as the one-time sale (OTS) revenue model. The second observed practice is where a firm generates revenue by providing ongoing service to its customers, where the charge depends on the duration/frequency of the service. For instance, cellular phone networks collect subscription fees from their customers, and such fees are charged on a per-time basis as long as a customer is using its service. This is referred to as the continuous service (CNS) revenue model.

Furthermore, after a new generation is introduced, a firm adopts one of two possible transition strategies, i.e., phase-out transition or total replacement, to terminate the supply of early product generations. For phase-out transition, older generations are allowed to run their course on the market — they continue to be sold as long as there is sufficient demand. For instance, digital cellular phone service was introduced in the early 1990s, but analog service continued to be provided until early 2000 in the U.S. Similarly, standard definition TV sets are still being sold in a market that is increasingly dominated by high-definition TVs. Under total replacement, firms discontinue the production and/or sale of the old generation as soon as a new generation is introduced. For instance, Microsoft stops selling older software versions as soon as a new version is released. In this study, we analyze both phase-out transition and total replacement.
As shown in Table 1, we examine four business scenarios (resulting from two revenue models and two generation transition strategies as discussed above) in the present research.

Table 1. Business Scenarios based on Revenue Models and Transition Strategies

<table>
<thead>
<tr>
<th>Phase-out Transition</th>
<th>One-time Sale</th>
<th>Continuous Service</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Scenario I (e.g., standard and high-definition TV sets)</td>
<td>Scenario II (e.g., analog and digital cellular service)</td>
</tr>
<tr>
<td>Total Replacement</td>
<td>Scenario III (e.g., previous and current versions of Microsoft Office)</td>
<td>Scenario IV (e.g., previous and current versions of SAS software license)</td>
</tr>
</tbody>
</table>

There are two pioneering studies (Wilson and Norton 1989; Mahajan and Muller 1996) that are the closest to the present research.² Wilson and Norton (1989) develop a model that captures the relationship between the adoptions of a new product and its product line extension. Based on that model, the authors find that in most cases a product line extension should be introduced either at the same time as the main product (now) or not to introduce at all (never).

Mahajan and Muller (1996) also study market entry timing for a two-generation case. The general conclusion of their study is that the second generation should be introduced together with the first generation (now) or when the diffusion of the first generation has reached its maturity stage (at maturity). The authors do not formally define maturity. Informally, the term seems to refer to the stage after the time of peak adoption.

A major difference between the two pioneering research and the present research is that Wilson and Norton (1989) and Mahajan and Muller (1996) seem to consider only Scenario I (one-time sale with phase-out transition), whereas we analyze all four business scenarios shown in Table 1. Even for Scenario I considered by both the present research and the two pioneering

² Besides the models by Wilson and Norton (1989) and Mahajan and Muller (1996), several other multigeneration models have been proposed (Norton and Bass 1987, Speece and MacLachlan 1995, Jun and Park 1999, Kim et al. 2000, Danaher et al. 2001, Jiang and Jain 2012), but none of them consider the issue of market entry timing.
studies, some of the key findings are different. For instance, we find that the optimal entry timing is not limited to *now, never, or at maturity*, but it can also lie between *now* and *maturity*.

The remainder of the paper is organized as follows. In Section 2, we show how profit can be projected under different business scenarios. In Section 3, we derive the market entry timing for the two-generation case under phase-out transition. The decision problem under total replacement is analyzed in Section 4. In Section 5, we extend the proposed model to derive the optimal market entry timing when (a) revenue is generated from both one-time sale and continuous service, and (b) there are three successive product generations. Finally, we discuss in Section 6 the managerial implications, research contributions, and future research directions of the present research.

2. **Profit Projection for Multiple Generations**

To obtain the optimal market entry timing for successive product generations, we need a multigeneration model to capture the across-generation diffusion dynamics. We review the prior literature on multigeneration diffusion models and find that a viable approach is to build on the Generalized Norton-Bass (GNB) model proposed by Jiang and Jain (2012). The GNB model is mathematically consistent with the Norton and Bass model (Norton and Bass 1987), but offers more functionality and flexibility and strong empirical support.

In the presence of successive product generations, potential (existing) adopters of an older generation can leapfrog (switch) to a newer generation. Specifically, *leapfrogging* represents the behavior of potential adopters skipping previous generation(s) and directly adopting a newer generation, and *switching* represents the behavior of existing adopters of the immediate previous generation making an upgrade to a new generation.
The rates of leapfrogging and switching adoptions depend on the generation transition strategy for a product line. Under total replacement, all potential adopters who would have adopted the older generation will leapfrog to the new generation after it is introduced. Under phase-out transition, since the old generation and new generation coexist in the market, only a portion of the potential adopters of the old generation will leapfrog to the new generation. Because the GNB model considers only phase-out transition, we first derive the total profit under phase-out transition.

2.1. Profit Projection under Phase-out Transition

The GNB model differentiates leapfrogging from switching, and provides closed-form expressions for both the number of units-in-use and the instantaneous adoption rate. Figures 1a and 1b illustrate the key differences between the number of units-in-use and the adoption rate for a two-generation case.

![Figure 1a. Adoption Rate under Phase-out Transition](image1)

![Figure 1b. Units-in-Use under Phase-out Transition](image2)

Suppose generation 1 (G1) is introduced at time $\tau_1$ generation 2 (G2) at time $\tau_2 \geq \tau_1$. Before $\tau_2$, the adoption rate of G1 follows the noncumulative Bass diffusion curve, while the number of units-in-use of G1 represents the cumulative number of adoptions until a given time. Therefore, the adoption rate of G1 could decrease before $\tau_2$, whereas the number of units-in-use of G1 is always increasing before $\tau_2$. After $\tau_2$, the adoption rate of G2 typically exhibits a bell-shaped

---

3 Without loss of generality, we let $\tau_1 = 0$. 
curve, while the units-in-use of G2 will be monotonically increasing. At some point during the second time period, the units-in-use curve for G1 will start to decline, because a large number of existing adopters of G1 will switch to G2.

We take the cellular phone market as an example. The adoption rate curve represents the rate of initial adoptions of a cellular service (e.g., analog, digital), while the units-in-use curve captures the number of active subscribers of the service. Although the cumulative number of adoptions of the analog service cannot decrease with time, the number of units-in-use for analog service can decrease as a result of switching.

Based on the GNB model, the number of units-in-use for the two successive generations can be represented by the following equations:

\[
S_1(t) = m_1 f_1(t) - m_1 f_1(t) F_2(t - \tau_2) = m_1 f_1(t) [1 - F_2(t - \tau_2)], \quad (1)
\]

\[
S_2(t) = m_2 F_2(t - \tau_2) + m_1 f_1(t) F_2(t - \tau_2) = [m_2 + m_1 f_1(t)] F_2(t - \tau_2). \quad (2)
\]

The instantaneous adoption rates for the two generations are

\[
y_1(t) = m_1 f_1(t) [1 - F_2(t - \tau_2)], \quad (3)
\]

\[
y_2(t) = [m_2 + m_1 f_1(t)] f_2(t - \tau_2) + m_1 f_1(t) F_2(t - \tau_2), \quad (4)
\]

and the cumulative numbers of adoptions for the two generations can be expressed as

\[
Y_1(t) = m_1 F_1(t) - m_1 \int_{\tau_2}^{t} f_1(\theta) F_2(\theta - \tau_2) d\theta, \quad (5)
\]

\[
Y_2(t) = [m_2 + m_1 F_1(t)] F_2(t - \tau_2). \quad (6)
\]

In Equations (1)-(6), \(m_1\) represents the market potential for generation 1, and \(m_2\) denotes the market potential specific to generation 2, i.e., potential adopters who are only interested in generation 2. Based on the GNB model, all potential (existing) adopters of generation 1 can leapfrog (switch) to generation 2, hence the total market potential for generation 2 is \(m_1 + m_2\). \(F_G(t)\) \((G = 1, 2)\) is the integral of \(f_G(t)\) and takes the following form:
As is common in the diffusion literature, we refer to $p_G$ and $q_G$ as the coefficient of innovation and coefficient of imitation, respectively, for generation $G$.

To project the total profit, we also need to know (i) the quantity of products sold or service provided, and (ii) the time-discounted profit per sale or per unit time’s service. The GNB model can be used to estimate the quantity of products sold or service provided, because it provides closed-form expressions for both the adoption rate and the number of units-in-use. Specifically, if the revenue model is one-time sale (OTS), the profit at any given time is proportional to the adoption rate at that time. If continuous service (CNS) is the underlying revenue model, the profit at any given time is proportional to the number of units-in-use at that time. We next discuss how the time-discounted profit per sale or per unit time’s service can be estimated.

When making product development decisions, a firm typically considers only cash flows generated during a certain planning horizon (e.g., 20 years), because projections beyond the planning horizon are often deemed too unreliable to be worth considering. We denote the length of a planning horizon by $D$. During a long planning horizon, the cost of providing product/service is expected to rise due to inflation or increased cost of capital. We denote the cost at time zero by $C_0$ and the rate of increase in cost by $r$; then the cost at time $t$ equals $C(t) = C_0 e^{rt}$. To cope with the rising costs, it is a common practice for firms to adjust its product/service price at approximately the same pace as the change in costs. If we denote the
price at time 0 by \( P_0 \), the price at time \( t \) becomes \( P(t) = P_0 e^{rt} \). Suppose the discount rate is also \( r \), then the present value of the profit derived from product/service provided at time \( t \) equals

\[
PV(P(t) - C(t)) = [P_0 e^{rt} - C_0 e^{rt}] e^{-rt} = P_0 - C_0.
\] (8)

Therefore, we treat the time-discounted profit per unit sale or per unit time’s service as constant during the planning horizon. For expositional convenience, we define unit contribution margin (for sale) as the present value of the profit resulting from selling one unit of a product, and denote the unit contribution margin for selling generation \( G \) by \( \pi_G \). Similarly, we define unit contribution margin (for service) as the present value of the profit resulting from providing one unit time’s service for one customer, and denote the unit contribution margin for providing \( G \) by \( \varphi_G \). We assume that all profit margins are positive, i.e., \( \pi_G > 0, \varphi_G > 0 \).

We denote the introduction time of generation \( G \) by \( \tau_G \) and the vector of introduction times for all later generations except \( G_1 \) by \( \bar{\tau} \). Under the one-time sale (OTS) revenue model, the total time-discounted profit for \( N \) product generations during the entire planning horizon (from time 0 to \( D \)) equals

\[
\pi(\bar{\tau}) = \sum_{G=1}^{N} \int_{\tau_G}^{D} \pi_G Y_G(\theta) d\theta = \sum_{G=1}^{N} \pi_G Y_G(D).
\] (9)

When continuous service (CNS) is the underlying revenue model, the total profit is

\[
\pi(\bar{\tau}) = \sum_{G=1}^{N} \pi_G \int_{\tau_G}^{D} S_G(\theta) d\theta.
\] (10)

Equations (9) and (10) both assume that the fixed cost of introducing a new generation is insignificant when compared to the variable cost and the total revenue generated from product sale or service, hence the fixed cost is not considered in our analysis. This assumption also allows us to have a better comparison between the findings of this research and those of Wilson.

\[^4\text{We have tried discount rates slightly higher than the rate of increase in costs, and found that the qualitative findings of this research remain valid.}\]
and Norton (1989) and Mahajan and Muller (1996), since the same assumption is also implicitly adopted by the two prior studies.

2.2. Profit Projection under Total Replacement

As stated earlier, the GNB model considers only phase-out transition; hence we need to extend it for profit projection under total replacement. We provide below a modeling framework for the two-generation case.

In Equations (1)-(4), the term $F_2(t - \tau_2)$ represents the *leapfrogging multiplier*, i.e., the proportion of potential adopters who leap to G2 instead of adopting G1. Under total replacement, since G1 is discontinued once G2 enters the market, we assume that all potential adopters who would have adopted G1 will leapfrog to G2, i.e., the effective leapfrogging multiplier is 1. Therefore, the adoption rate for G1 drops to 0 after $\tau_2$, and the original adopt rate of G1 is added to the rate of G2. Hence, the adoption rates for G1 and G2 become

$$
\hat{y}_1(t) = \begin{cases} 
    m_1 f_1(t), & t < \tau_2, \\
    0, & t \geq \tau_2,
\end{cases}
$$

(11)

$$
\hat{y}_2(t) = [m_2 + m_1 F_1(\tau_2)] f_2(t - \tau_2) + m_1 f_1(t), \quad t \geq \tau_2.
$$

(12)

From the adoption rates, we obtain the cumulative number of adoptions for G1 and G2:

$$
\hat{Y}_1(t) = \begin{cases} 
    m_1 F_1(t), & t < \tau_2, \\
    m_1 F_1(\tau_2), & t \geq \tau_2,
\end{cases}
$$

(13)

$$
\hat{Y}_2(t) = [m_2 + m_1 F_1(\tau_2)] F_2(t - \tau_2) + m_1 [F_1(t) - F_1(\tau_2)], \quad t \geq \tau_2.
$$

(14)

We next derive the number of units-in-use for the two generations. Here, we consider the scenario where existing adopters of G1 can continue to use the old generation until they decide to switch to G2, and the probability of switching at any given time is the same as that in the phase-out transition case. An example is that cellular phone users who have adopted analog service before the introduction of digital services are allowed to keep their analog service until they
voluntarily switch to digital service. Therefore, before $\tau_2$, the number of units-in-use of G1 is the same as the cumulative number of adoptions of G1. After $\tau_2$, the number of units-in-use of G1 equals the cumulative number of adoptions of G1 minus the cumulative number of switchings from G1 to G2. On the other hand, since G2 is the newest generation, the number of units-in-use of G2 always equals the cumulative number of adoptions of G2. Hence,

$$\dot{S}_1(t) = \begin{cases} m_1F_1(t), & t < \tau_2, \\ m_1F_1(\tau_2)[1 - F_2(t - \tau_2)], & t \geq \tau_2, \end{cases} \quad (15)$$

$$\dot{S}_2(t) = \dot{Y}_2(t) = [m_2 + m_1F_1(\tau_2)]F_2(t - \tau_2) + m_1[F_1(t) - F_1(\tau_2)], \quad t \geq \tau_2. \quad (16)$$

Equations (11)-(16) can be used for profit projection under total replacement. Similar to Equations (9) and (10), the total profits under one-time sale (OTS) and continuous service (CNS) are

$$\pi(\bar{\tau}) = \sum_{G=1}^{N} \pi_G \bar{Y}_G(D) = \sum_{G=1}^{N-1} \pi_G \bar{Y}_G(\tau_{G+1}) + \pi_N \bar{Y}_N(D), \quad \text{and}$$

$$\pi(\bar{\tau}) = \sum_{G=1}^{N} \pi_G \int_{\tau_G}^{D} \dot{S}_G(\theta) d\theta, \quad (17)$$

respectively.

For better comparison with phase-out transition, we show the adoption rate and units-in-use curves under total replacement in Figures 2a and 2b.

![Adoption Rate under Total Replacement](image1)

**Figure 2a. Adoption Rate under Total Replacement**

![Units-in-Use under Total Replacement](image2)

**Figure 2b. Units-in-Use under Total Replacement**

By comparing these two figures with Figures 1a and 1b for phase-out transition, we can see a number of important differences: (i) Under total replacement, the adoption rate of G1 drops to
zero after G2 is introduced at time $\tau_2$. (ii) Owing to the “forced” leapfrogging, the adoption rate of G2 starts at a higher level than the adoption rate of G1 just before $\tau_2$. (iii) Since there are no additional adoptions of G1 after the introduction of G2, the number of units-in-use of G1 decreases monotonically after $\tau_2$.

3. Market Entry Timing under Phase-out Transition

In this section, we derive the optimal market entry timing for the two-generation case under phase-out transition. Suppose the first generation (G1) is introduced at time zero; our goal is to find the optimal entry time for the second generation (G2) that leads to the highest profit. As discussed earlier, a firm’s revenue can be generated through either one-time sale (OTS) or continuous service (CNS). We examine both revenue models under phase-out transition.

3.1. Scenario I: One-Time Sale (OTS)

The one-time sale (OTS) revenue model with phase-out transition corresponds to Scenario I in Table 1. Many consumer products (e.g., computers, TVs) fall under this business scenario. The total profit for such products is given in Equation (9); hence the decision problem for deciding the optimal market entry time for G2 is formulated as

$$\max_{a_2 \leq \tau_2 \leq D} \pi(\tau_2) = \pi_1 Y_1(D) + \pi_2 Y_2(D),$$

(19)

where $Y_1(D)$ and $Y_2(D)$ are given in Equations (5) and (6), respectively.

Regarding the uniformity of the coefficients of innovation and imitation across generations, the prior literature has different assumptions and empirical findings. Studies such as Norton and Bass (1987), Wilson and Norton (1989), and Mahajan and Muller (1996) assume that these coefficients remain constant across generations. Other studies (e.g. Islam and Meade 1997, Danaher et al. 2001) find that allowing the coefficient of imitation to change across generations can lead to a better model fit. In a more recent study, based on data for 39 product generations in
twelve product markets, Stremersch et al. (2010) find that the acceleration of diffusion from
generation to generation is primarily a result of the shorter time to takeoff instead of generational
shifts. The changes in the coefficients of innovation and imitation across generations are found to
be insignificant for all but one product categories. We therefore assume that the coefficients of
innovation and imitation both remain constant across the two generations.

Denoting the constant coefficients as $p$ and $q$, i.e., $p = p_1 = p_2$ and $q = q_1 = q_2$, we have
$F(t) = F_1(t) = F_2(t)$ and $f(t) = f_1(t) = f_2(t)$, $\forall t$. Formulation (19) then becomes

$$\text{MAX} \pi(\tau_2) = \pi_1 \left[ m_1 F(D) - m_1 \int_{\tau_2}^{D} f(\theta) F(\theta - \tau_2) d\theta \right] + \pi_2 \left[ m_2 + m_1 F(D) \right] F(D - \tau_2).$$

(20)

Based on the GNB model, we find that delaying the introduction of G2 allows G1 to reach
a larger portion of its potential adopters (represented by $m_1$), which leads to less leapfrogging and
more switching to G2. This is beneficial to a firm since switching implies across-generation
repeat purchases while leapfrogging does not. On the other hand, delaying the market entry of
G2 results in fewer adoptions by those who are only interested in G2 (counted in $m_2$), because a
larger portion of the planning horizon will lapse when G2 enters the market.

We present below Propositions 1 and 2 regarding the optimal market entry timing under
Scenario I.\footnote{All proofs are provided in the Appendix.

**Proposition 1.** Under the one-time sale revenue model and phase-out transition, it is always
optimal to introduce the second generation sometime before the end of the planning horizon ($D$).
Specifically, there exists a positive $\delta$ such that

$$\begin{cases} 
\tau_2^* < D - \delta, & \text{if } \delta < D - \alpha_2, \\
\tau_2^* = \alpha_2, & \text{if } \delta \geq D - \alpha_2,
\end{cases}$$

where $\alpha_2$ represents the earliest possible introduction time for the second generation.
Proposition 1 can be illustrated by Figure 3. If $\delta$ (expression given in the proof of Proposition 1 in the Appendix) is smaller than $(D - \alpha_2)$, G2 should be introduced to the market at a time between $\alpha_2$ (inclusive) and $(D - \delta)$. If $\delta$ is larger than or equal to $(D - \alpha_2)$, G2 should be introduced as early as possible, i.e., $\tau^*_2 = \alpha_2$.

![Figure 3. Illustration of Proposition 1](image)

The total profit generated from both product generations depends on the numbers of adoptions of G1 and G2 and their relative unit contribution margins. With a close examination, we find that, unless the unit contribution margin for G2 is less than that for G1, introducing G2 at any given time $\tau_2$ during the planning horizon (even if $\tau_2$ is not the optimal time) is always better than not introducing G2 at all. This is because with $\pi_1 \leq \pi_2$, leapfroggings from G1 to G2 do not decrease the profit, while switchings and the adoptions by the G2-specific adopters strictly increase the profit.

In case $\pi_1 > \pi_2$, each leapfrogging adoption reduces the profit by $(\pi_1 - \pi_2)$, while each switching adoption or each initial adoption by a G2-specific adopter increases the profit by $\pi_2$. Even if the benefit is less than the cost when G2 is introduced early in the planning horizon, the benefit/cost ratio will increase as the introduction time moves closer to the end of the planning horizon. This is because as the introduction time ($\tau_2$) of G2 is delayed, more potential adopters would have adopted G1 by time $\tau_2$, hence the rate of leapfrogging after $\tau_2$ will strictly decrease, while rate of switching will strictly increase. Proposition 1 implies that the time point where the
total benefit exceeds the total cost of introducing G2 always occurs before the end of the planning horizon.

From Proposition 1, we conclude that regardless of the length of the planning horizon, the relative unit contribution margin, the relative market potential, and the projected diffusion curves of the two generations, it is always optimal to introduce the second generation sometime during the planning horizon.

This is different from the conclusion reached by Wilson and Norton (1989) which states that under certain conditions a product line extension should never be introduced. Furthermore, Wilson and Norton find that if it is beneficial to introduce a line extension, it should be introduced now, while we find that even if the new generation is available now, it may be optimal to release it at a later time.

We believe that our finding is consistent with what we observe in the marketplace. First, it is frequently observed that firms introduce a new generation a few years after the older generation is introduced. Furthermore, there are many reported examples that firms intentionally delayed the introduction of a new innovation to achieve a higher benefit. Examples include Microsoft Longhorn operating system, Intel Camino chipset, and other innovations such as DVD, MP3, and 3G Cellular networks (Wang and Hui 2005, 2010).

**Proposition 2.** Under the one-time sale revenue model and phase-out transition, if the unit contribution margin for the second generation ($\pi_2$) is equal to or greater than that for the first generation ($\pi_1$), and the planning horizon ($D$) is shorter than or equal to the time of peak diffusion ($T^*$) for the first generation (in the absence of leapfrogging), then it is optimal to introduce the second generation as early as possible. Formally,

$$\tau_2^* = \alpha_2 \text{ if } \pi_1 \leq \pi_2 \text{ and } D \leq T^* = \ln(q/p)/(p + q).$$

Proposition 2 provides a sufficient (but not necessary) condition for introducing the new generation as early as possible. According to Bass (1969), the time of peak diffusion for G1
(assuming that leapfrogging does not occur) is reached at $T^* = \ln(q/p)/(p + q)$. For certain new products, it may take a long time before the peak diffusion can be reached. In such cases, especially in a volatile market, a firm may face pressure to deliver quicker performance results. Therefore, decision-makers may consider a relatively short planning horizon. From Proposition 2, we conclude that if the planning horizon is shorter than $T^*$ and the unit contribution margin for the second generation is at least as high as that for the first generation, the second generation should be introduced as early as possible.

Proposition 2 can be compared with a related finding by Mahajan and Muller (1996). In their study, Mahajan and Muller consider an infinite planning horizon, and find that G2 should be introduced now if (i) the market potential for G2 is larger than that for G1, or (ii) the (current and future) profit per sale for G2 is larger than that for G1; otherwise G2 should be introduced when G1 has reached its maturity stage.

There are clear differences between the sufficient conditions for introducing G2 at its earliest possible date (i.e., now). Specifically, Mahajan and Muller conclude that either of the two aforementioned condition assures that it is optimal to introduce G2 as soon as possible, whereas we find that the comparison of market potential and/or contribution margins do not constitute a sufficient condition for introducing G2 as soon as possible. Instead, the duration of the finite planning window is also needed in order to determine whether introducing G2 now is the optimal solution. Furthermore, based on our model formulation, we find that under certain conditions even if G2 has a larger market potential and a higher profit margin, introducing it at a later date may still be optimal.

---

6 Proposition 2 cannot be directly compared with the findings presented by Wilson and Norton (1989), because $\pi_1 > \pi_2$ is assumed in that study.
For illustration purpose, we use a US cellular phone subscribers dataset\(^7\) and adopt its parameter values, i.e., \(p = 0.0158\), \(q = 0.279\), \(m_1 = 38.68\) million, and \(m_2 = 318.06\) million. Note that the time of peak is \(T^* = 9.74\) years based on the adopted \(p\) and \(q\) values. The unit contribution margin for selling cellular phones is assumed to be \(\pi_1 = \pi_2 = 20\). In order to better understand how sensitive the optimal solution is to different parameter values, we assume that \(G_2\) is available for market introduction at time zero \((\tau_2 = 0)\). Numerically, we find that when planning horizon is set to \(D = 30\) years, the highest profit \(\pi^* = 7.63\) billion is achieved at \(\tau_2^* = 5.8\) years. With a shorter planning horizon \(D = 25\) years, the optimal entry timing is \(\tau_2^* = 1.8\) years, leading to a total profit of \(\pi^* = 7.45\) billion. When \(D = 20\) years, it is optimal to introduce \(G_2\) at the same time as \(G_1\), i.e., \(\tau_2^* = 0\), and the total profit is \(\pi^* = 7.13\) billion.

These results confirm another key difference in findings between this research and Mahajan and Muller (1996). Specifically, the solution \(\tau_2^* = 5.8\) years (when \(D = 30\)) and \(\tau_2^* = 1.8\) years (when \(D = 25\)) show that it may be optimal to introduce a new generation after its earliest possible introduction date, but before the maturity date (i.e., time of peak \(T^* = 9.74\) years) of the previous generation. This practice is observed in the real-world. For instance, Apple Inc. launched its new iPad 3 in March 2012, even though the sales of iPad 2 (launched in March 2011) were still on a strong upward trajectory (Olanoff 2012), implying that its maturity stage has not arrived yet.

### 3.2. Scenario II: Continuous Service (CNS)

Under the continuous service (CNS) model, revenue is generated by providing ongoing service to customers. Cellular network service providers such as Verizon are good examples of this business scenario. Unlike the OTS model, cellular phone customers do not pay a one-time fee to

\(^7\) Detailed information is available from http://www.itu.int/ITU-D/ict/publications/world/world.html.
use the service; instead, a monthly subscription fee is charged as long as the customer is using the service.

The total profit under continuous service and phase-out transition can be estimated from Equation (10). Therefore, the optimal market entry time for G2 can be obtained by solving

$$\max_{a_2 \leq \tau_2 \leq D} \pi(\tau_2) = \varphi_1 \int_0^D S_1(\theta) d\theta + \varphi_2 \int_{\tau_2}^D S_2(\theta) d\theta,$$

where $S_1(t)$ and $S_2(t)$ are defined in Equations (1) and (2), respectively.

We again assume $p = p_1 = p_2$ and $q = q_1 = q_2$. Formulation (21) then becomes

$$\max_{0 \leq \tau_2 \leq D} \pi(\tau_2) = \varphi_1 \int_0^D m_1 F(\theta)[1 - F(\theta - \tau_2)] d\theta + \varphi_2 \int_{\tau_2}^D [m_2 + m_1 F(\theta)] F(\theta - \tau_2) d\theta.$$  (22)

There are two important differences between this CNS model and the OTS model. First, under CNS, since the revenue is not generated through one-time sale, whether a customer adopts G2 through leapfrogging or switching does not affect the firm’s revenue from G2; whereas switching is more beneficial than leapfrogging under OTS. Second, under CNS how long a service is being consumed by users directly affects a firm’s revenue, while under OTS the duration of usage has no direct effect on revenue.

For most service types, we expect that the new generation is at least as the old generation per unit time’s service ($\varphi_1 \leq \varphi_2$). Under this condition, we have the following finding:

**Proposition 3.** Under the continuous service revenue model and phase-out transition, if the unit contribution margin for the second generation is equal to or greater than that for the first generation, it is always optimal to introduce the second generation as early as possible, i.e.,

$$\tau_2^* = a_2 \text{ if } \varphi_1 \leq \varphi_2.$$

Proposition 3 can be explained as follows. If G2 is introduced earlier, although the number of switchings during the planning horizon may either increase or decrease, we can tell from Equations (1) and (2) that the sum of the numbers of leapfroggings or switchings can only

18
increase. Since G2 is at least as profitable as G1, more leapfroggings or switchings from G1 to G2 can never hurt the revenue. Even if the number of switchings decreases, it does not affect a firm’s revenue because no revenue is generated from initial adoptions anyway. Furthermore, an earlier market entry time allows G2 to be used longer during the planning horizon, and more G2-specific potential adopters (represented by $m_2$) can adopt G2 by the end of the planning horizon, thus leading to higher revenue for the firm. Since all factors favor an earlier introduction of G2, we conclude that when the unit contribution margin for the second generation is at least as high as that for the first generation, the second generation should be introduced to the market at its earliest possible date.

We take the cellular phone service again as an example. If the unit contribution margin for 4G service is at least as high as that for 3G, then 4G service should be introduced as soon as it becomes available. This is because customers’ leapfrogging or switching from 3G to 4G service cannot decrease the profit; and potential customers who are waiting for 4G service can start adopting the service earlier, thus increasing the total profit during the planning horizon.

In case the new generation is not as profitable as the old generation per unit time’s service ($\varphi_1 > \varphi_2$), we have the following finding:

**Proposition 4.** Under the continuous service revenue model and phase-out transition, even if the unit contribution margin for the second generation is less than that for the first generation, it is still optimal to introduce the second generation as early as possible if the market potential specific to the second generation is sufficiently large. Specifically,

$$
\tau^*_2 = \alpha_2, \text{ if } m_2 \geq \frac{(\varphi_1 - \varphi_2)F(D)}{\varphi_2}m_1.
$$

This proposition can also be explained. If G2 is not as profitable as G1, then whenever a customer shifts to G2 either through leapfrogging or switching, the firm suffers a decrease in profit. However, if the number of potential adopters who are only interested in G2 (i.e. the
market potential specific to G2) is sufficiently large, the loss can be compensated by the additional revenue generated from such adopters. Therefore, even if the unit contribution margin for the second generation is less than that for the first generation, introducing G2 at the earliest date may still be optimal if the market potential for G2 is sufficiently large. If the market potential for G2 is not sufficiently large, we find that a firm should either introduce it late in the planning horizon or not introduce it at all.

We set the unit contribution margins for one year’s cellular service to \( \varphi_1 = \varphi_2 = $200 \). Consistent with Proposition 3, we find numerically that regardless of the duration of planning horizon, the optimal introduction time for G2 is always \( \tau_2^* = 0 \). To understand the solution under the less likely scenario with G2 less profitable than G1 \( (\varphi_1 > \varphi_2) \), we let \( \varphi_2 = $150 \) and vary the value of \( m_2 \). The planning horizon is fixed at \( D = 30 \) years. Consistent with Proposition 4, we find that \( \tau_2^* = 0 \) if \( m_2 \geq \frac{(\varphi_1-\varphi_2)F(D)}{\varphi_2} m_1 \) holds. With \( m_2 = 20 \) million, the total profit decreases monotonically as the introduction of G2 is delayed. Hence, it is optimal to introduce G2 as early as possible. On the other hand, when \( m_2 = 10 \) million, the impact of G2’s introduction time on the total profit is non-monotonic, and the highest profit is achieved at \( \tau_2^* = 30 \) years, implying that G2 should not be introduced during the planning horizon.

We further vary the value of \( m_2 \), and find that the optimal solution exhibits an interesting now or never pattern, similar to the finding reported in the pioneering study by Wilson and Norton (1989). Specifically, we find that \( \tau_2^* = 0 \) when \( m_2 \geq 10.93 \) million, and it jumps to \( \tau_2^* = 30 \) years when \( m_2 \leq 10.92 \) million. There seems to exist a threshold, around which a very small change in the market potential for G2 can change its optimal entry timing from as early as possible to not to introduce at all. We would like to point out that there are similarities and differences between this finding and the now or never conclusion arrived at by Wilson and
Norton (1989). Both findings are under the condition that the unit contribution margin for G2 is lower than that for G1. The difference is that Wilson and Norton (1989) derive the finding for the OTS model (the business scenario examined in their study is the closest to Scenario I shown in Table 1), whereas our finding is only valid for the CNS model. It is very interesting to observe that conclusion of the pioneering study remains valid, although under a different business scenario.

4. Market Entry Timing under Total Replacement

We now examine the two-generation case under total replacement. Since G1 is discontinued once G2 is released, all potential adopters who would have adopted G1 at time $t$ will leapfrog to G2. This leads to adoption rates and units-in-use numbers that are different from the phase-out transition case. To the best of our knowledge, market entry timing under total replacement has not been formally studied in prior literature.

4.1. Scenario III: One-Time Sale (OTS)

The total profit under one-time sale and total replacement can be obtained based on Equation (17). Therefore, the decision problem for deciding the profit-maximizing market entry timing for G2 is formulated as:

$$
\max_{a_2 \leq \tau_2 \leq D} \pi(\tau_2) = \pi_1 \hat{Y}_1(\tau_2) + \pi_2 \hat{Y}_2(D),
$$

where $\hat{Y}_1(t)$ and $\hat{Y}_2(t)$ are defined in Equations (13) and (14), respectively.

We still assume that the coefficients of innovation and imitation remain the same across generations, implying $F(t) = F_1(t) = F_2(t)$ and $f(t) = f_1(t) = f_2(t), \forall t$. Then, (23) becomes

$$
\max_{a_2 \leq \tau_2 \leq D} \pi(\tau_2) = \pi_1 m_1 F(\tau_2) + \pi_2 [m_2 + m_1 F(\tau_2)] F(D - \tau_2) + \pi_2 m_1 [F(D) - F(\tau_2)].
$$

The objective function of (24) can be reorganized as

$$
\pi(\tau_2) = \pi_2 m_1 F(D) + (\pi_1 - \pi_2) m_1 F(\tau_2) + \pi_2 [m_2 + m_1 F(\tau_2)] F(D - \tau_2).
$$
Taking the derivative of (25) with respect to \( \tau_2 \) yields

\[
\frac{\partial \pi(\tau_2)}{\partial \tau_2} = (\pi_1 - \pi_2)m_1 f(\tau_2) - \pi_2 m_2 f(D - \tau_2) + \pi_2 m_1 f(\tau_2) F(D - \tau_2) - \pi_2 m_1 F(\tau_2) f(D - \tau_2). 
\]

After substituting \( F(\cdot) \) and \( f(\cdot) \), letting \( x = e^{(p+q)\tau_2} \) and \( \delta = e^{-(p+q)D} \), and some additional algebraic rearrangement (details available from the authors), the above derivative can be expressed as

\[
\frac{d\pi(\tau_2)}{d\tau_2} = H(x)(ax^2 + bx + c), \tag{27}
\]

where

\[
H(x) = -\frac{(p+q)^2 x}{p[(q/p)+x][(q/p)\delta x+1]^2},
\]

\[
a = (\pi_2 - \pi_1)m_1 \delta^2 \left( \frac{q}{p} \right)^2 + \pi_2 m_2 \delta + \pi_2 m_1 \delta \frac{q}{p} + \pi_2 m_1 \delta,
\]

\[
b = 2(\pi_2 - \pi_1)m_1 \delta \frac{q}{p} + 2\pi_2 m_2 \delta \frac{q}{p}, \text{ and}
\]

\[
c = \pi_2 m_2 \delta \left( \frac{q}{p} \right)^2 - \pi_1 m_1 - \pi_2 m_1 \delta \frac{q}{p}.
\]

We next take a closer look at the terms in (27). Since \( x = e^{(p+q)\tau_2} \geq 1 \), we must have

\[
H(x) < 0.
\]

Furthermore, for most products, we expect the unit contribution margin for G2 to be at least as high as that for G1 (i.e., \( \pi_2 \geq \pi_1 \)), hence \( a > 0 \) and \( b > 0 \) must hold. Even if \( \pi_2 \) is slightly lower than \( \pi_1 \), numerical analysis based on common parameter values show both \( a \) and \( b \) are likely to remain positive. The parameter \( c \), on other hand, can easily fall into either positive or negative ranges.

Regardless of the sign of the parameter values, the optimal market entry timing can be analytically derived based on (27) and the corresponding first-order condition. We present below the solution for the most likely scenario, i.e., \( a > 0 \) and \( b > 0 \):

**Proposition 5:** Under the one-time sale revenue model and total replacement, when \( a > 0 \) and \( b > 0 \), it is optimal to introduce the second generation at
\[
\tau_2^* = \begin{cases} 
\ln \left( \frac{-b + \sqrt{b^2 - 4ac}}{2a} \right) / (p + q), & \text{if } c < -a e^{2(p+q)a_2} - b e^{(p+q)a_2}, \\
\alpha_2, & \text{otherwise.}
\end{cases}
\]

To illustrate, we again assume \( \pi_1 = \pi_2 = 20 \) and \( D = 30 \) years. Using the earlier parameter values, we obtain the values for derived parameters as \( \delta = 0.000146 \), \( a = 1.04 \), \( b = 32.65 \), and \( c = -488.37 \). Based on Proposition 2, it is optimal to release \( G_2 \) at \( \tau_2^* = 8.16 \) years, and the total profit is \( \pi^* = 7.21 \) billion. We also try shorter planning horizons of 20 and 25 years. For both cases, we find \( \tau_2^* = 0 \), implying that it is optimal to introduce \( G_2 \) as early as possible.

4.2. Scenario IV: Continuous Service (CNS)

Under continuous service and total replacement, the total profit can be estimated from Equation (18). The optimal market entry timing under this business scenario can be obtained based on the following formulation:

\[
\max_{\tau_2 \leq \tau_2^D} \pi(\tau_2) = \varphi_1 \int_0^D \dot{S}_1(\theta) d\theta + \varphi_2 \int_{\tau_2}^D \dot{S}_2(\theta) d\theta,
\]

where \( \dot{S}_1(t) \) and \( \dot{S}_2(t) \) are defined in Equations (15) and (16), respectively.

With constant coefficients of innovation and imitation across generations, (28) becomes

\[
\max_{\tau_2 \leq \tau_2^D} \pi(\tau_2) = \varphi_1 m_1 \int_0^\tau_2 F(\theta) d\theta + \varphi_1 \int_{\tau_2}^D \dot{S}_1(\theta) d\theta + \varphi_2 \int_{\tau_2}^D \dot{S}_2(\theta) d\theta,
\]

Under the typical condition that the second generation is at least as profitable as the first generation per unit time’s service, we have the following conclusion:

**Proposition 6.** Under the continuous service revenue model and total replacement, if the unit contribution margin for the second generation (\( \varphi_2 \)) is equal to or greater than that for the first generation (\( \varphi_1 \)), it is always optimal to introduce the second generation as early as possible, i.e.,

\[ \tau_2^* = \alpha_2 \text{ if } \varphi_1 \leq \varphi_2. \]

Similar to Proposition 3, Proposition 6 holds for two main reasons. First, if \( G_2 \) is at least as profitable as \( G_1 \), then leapfrogging or switching between the two generations cannot hurt the
total profit. Second, introducing G2 earlier allows it to reach more customers and be used for a longer duration. Therefore, the second generation should be introduced as early as possible if it is at least as profitable as the first generation per unit time’s service.

In case the second generation is less profitable than the first generation per unit time’s service, we find conditions under which the total profit will be non-decreasing/non-increasing with the introduction time of the second generation, as presented below:

**Proposition 7.** Under the continuous service revenue model and total replacement, if the unit contribution margin for the second generation is less than that for the first generation, there exist two positive threshold values such that: (i) if the market potential for the second generation is less than the lower threshold value, the profit is non-decreasing with the introduction time of the second generation, and (ii) if the market potential for the second generation is greater than the higher threshold value, the profit is non-increasing with its introduction time.8

Based on this proposition, in case the market potential for the second generation is less than the lower threshold value, the second generation should not be introduced during the planning horizon (i.e., never). In case the market potential for the second generation is greater than a higher threshold value, it should be introduced as early as possible (i.e., now). When the market potential for the second generation is between the two threshold values, we find numerically that the optimal introduction time is still very likely to be either now or never, but can also be after the introduction of G1 but some time before the end of the planning horizon.

For numerical analysis, we examine only the less likely scenario with \( \varphi_1 > \varphi_2 \). We let \( \varphi_1 = $200, \varphi_2 = $150, \) and \( D = 30 \) years. We then gradually decrease the value of \( m_2 \) and search for the optimal solution. Similar to the interesting result obtained for the CNS model under phase-out transition, the optimal solution exhibits the now or never pattern first reported by Wilson and Norton (1989). The threshold lies close to 12.9 million. If \( m_2 \geq 12.9 \) million, G2

---

8 The expressions for the lower and higher threshold values are provided in the proof of Proposition 7 in the Appendix.
should be introduced as early as possible. If \( m_2 < 12.89 \) million, G2 should never be introduced. Once again, it is worth noting that although this *now or never* rule was first discovered by Wilson and Norton (1989), we find that the rule is valid only for a different business scenario.

### 5. Model Extensions

In the previous sections, we consider only the two-generation case and assume that a firm’s source of revenue is either one-time sale (OTS) or continuous service (CNS), but not both. In this section, we extend our model to analyze two more generation business scenarios: a) Firms’ revenue is generated from both one-time sale and continuous service (OTCNS), and b) A product line includes three successive generations. We consider only phase-out transition in both extensions.

#### 5.1. Both One-Time Sale and Continuous Service (OTCNS)

For some product lines, revenue is generated from both one-time product sale and after-sale service (OTCNS). An example of this business scenario is that manufacturers of GPS units can benefit from both one-time sales of GPS units and subsequent map updates. Similarly, producers of printers can collect revenue from initial sales of printers and future sales of cartridges.

The problem formulation for OTCNS can be obtained by combining the profits generated from both one-time sale and continuous service. We again assume that the coefficients of innovation and imitation both remain constant across generations. The optimal entry timing for the second generation can be obtained by solving

\[
\text{MAX}_{\tau_2 \leq D} \pi(\tau_2) = \pi_1 Y_1(D) + \pi_2 Y_2(D) + \phi_1 \int_0^D S_1(\theta) d\theta + \phi_2 \int_{\tau_2}^D S_2(\theta) d\theta, \tag{30}
\]

s.t. 
\[
Y_1(D) = m_1 F(D) - m_1 \int_{\tau_2}^D f(\theta) F(\theta - \tau_2) d\theta,
\]
\[
Y_2(D) = [m_2 + m_1 F(D)] F(D - \tau_2),
\]
\[
S_1(t) = m_1 F(t)[1 - F(t - \tau_2)],
\]
\[
S_2(t) = [m_2 + m_1 F(t)] F(t - \tau_2).
\]
From (30), we conclude the total profit under the OTCNS model can be interpreted as the weighted average of the profits under the OTS and CNS models. For example, if we set the unit contribution margin for service to $\varphi_1 = \varphi_2 = 0$, the formulation reduces to the OTS model (20). Similarly, if the unit contribution margin for one-time sale is $\pi_1 = \pi_2 = 0$, the simplified formulation represents the CNS model (22).

Therefore, whether the optimal solution is in conformity with the findings for the OTS model or the CNS model depends on the ratio between the unit contribution margin for sale and that for after-sale service.

We perform numerical analyses to examine the solutions for two different cases. In the first case, the profit resulting from continuous service is significantly higher than the profit from initial product sale. An example of this case is that profits are generated from both cellular phone sales and cellular network service. We set the unit contribution margin for cellular phone sale to $\pi_1 = \pi_2 = 20$, and the unit contribution margin for cellular network service to $\varphi_1 = \varphi_2 = 200$. Three different planning horizons (20, 25, and 30 years) are tried, and the solutions all suggest that it is optimal to introduce G2 as early as possible, i.e., $\tau_2^* = 0$. This is expected because the profit from continuous service dominates the profit from one-time phone sale and, based on Proposition 3, the service profit is maximized at $\tau_2^* = 0$.

In the second case, a firm derives revenue primarily from one-time product sale instead of after-sale service. An example is that refrigerator manufacturers profit mainly from selling refrigerators instead of water filters. For illustration, we change the unit contribution margins to $\pi_1 = \pi_2 = 200$ and $\varphi_1 = \varphi_2 = 1$, and set the planning horizon to 30 years. The optimal introduction time for G2 is found to be $\tau_2^* = 1.7$ years. In this case, since the unit contribution
margin for one-time sale is significantly higher than that for service, slightly delaying the introduction of G2 can lead to more cross-generation repeat purchases and a higher profit.

5.2. Entry Timing for Three Generations

We now demonstrate how the optimal market entry timing can be determined for a three-generation case. We consider only the one-time sale (OTS) revenue model under phase-out transition for illustration purpose.

The GNB model (Jiang and Jain 2012) is again used to project the adoption rates for all three generations. The optimal market entry timing for G2 and G3 can be determined based on:

\[
\text{MAX}_{\alpha_2 \leq \tau_2 \leq \tau_3 \leq D} \pi(\tau_2, \tau_3) = \pi_1 Y_1(D) + \pi_2 Y_2(D) + \pi_3 Y_3(D) \quad (31)
\]

s.t. 
\[
Y_1(D) = m_1 F(D) - m_1 \int_{\tau_2}^{D} f(\theta) F(\theta - \tau_2) d\theta,
\]
\[
Y_2(D) = [m_2 + m_1 F(D)] F(D - \tau_2) - \int_{\tau_3}^{D} [(m_2 + m_1 F(\theta)) f_2(\theta - \tau_2) + m_1 f(\theta) F(\theta - \tau_2)] F(\theta - \tau_3) d\theta,
\]
\[
Y_3(D) = \{m_3 + [m_2 + m_1 F(D)] F(D - \tau_2)\} F(D - \tau_3).
\]

Since the US cellular subscribers dataset include only two generations, we use the average parameter values reported in prior research (Sultan et al. 1990), i.e., \( p = 0.03, \quad q = 0.38 \), in the numerical analysis. The market potentials for the three generations are set to \( m_1 = 100 \) million, \( m_2 = 200 \) million, and \( m_3 = 300 \) million. The unit contribution margins are assumed to be equal across generations: \( \pi_1 = \pi_2 = \pi_3 = 20 \). The planning horizon is first set to \( D = 30 \) years, and the profit-maximizing introduction times are found to be \( \tau_2^* = 5.0 \) years for G2 and \( \tau_3^* = 14.1 \) years for G3. If the planning horizon is reduced to \( D = 25 \) years, it is optimal to introduce both generations a few years earlier: G2 at \( \tau_2^* = 2.4 \) years for and G3 at \( \tau_3^* = 10.3 \) years. With an even
shorter planning horizon $D = 20$ years, the optimal solution is $\tau_2^* = 0.0$ years and $\tau_3^* = 6.5$ years. Therefore, the shorter is the planning horizon, the earlier should G2 and G3 be introduced.

6. Conclusions and Future Research Directions

In today’s business environment, continuous product improvement in the form of successive release of product generations is critical for market success. Successful management of product generations not only helps firms achieve their competitive advantage, but also ensures that a firm can sustain its revenue stream from a relatively constant customer base. Despite its importance, studies pertaining to market entry timing for successive generations have been limited. The two pioneering studies (Wilson and Norton 1989, Mahajan and Muller 1996) have proposed analytical models to address this important problem and derived some interesting findings regarding the entry timing for a new product generation or a product line extension. In this research, we build on these pioneering studies and conduct a more comprehensive analysis of market entry timing for four business scenarios observed in the marketplace today, corresponding to two possible revenue models (one-time sale, continuous service) and two generation transition strategies (phase-out transition, total replacement). The models developed in this research can help firms make informed decisions when managing the introduction of successive product generations.

In addition to the practical implications, the present research makes important theoretical contributions to the existing literature on market entry timing. To the best of our knowledge, the present study is the only one that identifies and models multiple revenue models and generation transition strategies. To obtain the optimal market entry timing, we adopt the Generalized Norton-Bass model (Jiang and Jain 2012) to estimate the total profit under phase-out transition. For total replacement, we extend the GNB model for profit projection. Based on these proposed
models, we derive some new and interesting findings. For instance, when one-time sale is the only source of revenue under phase-out transition, we find that it is always beneficial to introduce a second generation sometime during the planning horizon. Moreover, the optimal introduction time is not limited to *now* or *at maturity* (i.e., after the time of peak diffusion); instead, it can lie between now and maturity.

When continuous service is the only source of revenue, we find that regardless of the generation transition strategies, if the unit contribution margin for the second generation is at least as high as that for the first generation, it is always optimal to introduce the second generation as early as possible. Furthermore, when continuous service is the only source of revenue and the unit contribution margin for the second generation is less than that for the first generation, the optimal market entry timing for the second generation is typically either *now* (i.e., to introduce as early as possible) or *never* (i.e., not to introduce during the planning horizon). Although this *now* or *never* rule was first discovered by Wilson and Norton (1989), we find that the pattern is valid only for a different business scenario.

In addition, we extend our modeling framework to derive market entry timing for two more generation business scenarios where: a) revenue is generated from both one-time sale and continuous service, and b) a product line includes three successive generations. Both extensions represent new contributions to the existing literature on market entry timing.

There exist several other interesting future research directions. First, we currently assume that the earliest possible introduction time for a future generation and the amount of technological improvements in a new generation are exogenously given. In a future study, we could endogenize the availability and maturity of technological advancements, and develop a
more comprehensive model to determine the entry timing along with the amount of improvements in the new generation.

Second, marketing mix variables are not considered in this study (Bass, Jain, Krishnan, 2000). It would be interesting to analyze how different pricing and advertising policies affect a firm’s market entry timing decision (Krishnan, Bass & Jain 1999; Krishnan & Jain 2006). There might exist a global optimal solution that offers the best combination of pricing or advertising policy and entry timing for multiple product generations.

Finally, it would be interesting to adapt the proposed models to help decide the optimal timing for a new textbook revision, a problem important to the publishing industry.
References


Hughes, N. 2010. Apple's recurring revenue stream: 77% of iPhone 4 sales were upgrades. *Apple Insider*, June 25. Available at http://www.appleinsider.com/articles/10/06/25/apples_recurring_revenue_stream_77_of_iphone_4_sales_were_upgrades.html.


Appendix

Proof of Proposition 1

We first obtain the derivative of the objective function of (20) with respect to $\tau_2$:

$$
\frac{d\pi(\tau_2)}{d\tau_2} = \pi_1 m_1 f(\tau_2)F(\tau_2 - \tau_2) + \pi_1 m_1 \int_{\tau_2}^{D} f(\theta)f(\theta - \tau_2)d\theta - \pi_2 [m_2 + m_1 F(D)] f(D - \tau_2)
$$

$$
= \pi_1 m_1 \int_{\tau_2}^{D} f(\theta)f(\theta - \tau_2)d\theta - \pi_2 [m_2 + m_1 F(D)] f(D - \tau_2).
$$

(A1)

From (A1), we have

$$
\frac{d\pi(\tau_2)}{d\tau_2} < \pi_1 m_1 \left[ \max_{0 \leq \theta \leq D} f(\theta) \right]^2 (D - \tau_2) - \pi_2 [m_2 + m_1 F(D)] f(D - \tau_2)
$$

$$
< \pi_1 m_1 \left[ \max_{0 \leq \theta \leq D} f(\theta) \right]^2 (D - \tau_2) - \pi_2 [m_2 + m_1 F(D)] \min_{0 \leq \theta \leq D} f(\theta)
$$

$$
= \pi_1 m_1 [\max_{0 \leq \theta \leq D} f(\theta)]^2 \left[ (D - \tau_2) - \frac{\pi_2 [m_2 + m_1 F(D)] \min_{0 \leq \theta \leq D} f(\theta)}{\pi_1 m_1 [\max_{0 \leq \theta \leq D} f(\theta)]^2} \right].
$$

Let

$$
\delta = \frac{\pi_2 [m_2 + m_1 F(D)] \min_{0 \leq \theta \leq D} f(\theta)}{\pi_1 m_1 [\max_{0 \leq \theta \leq D} f(\theta)]^2},
$$

we have

$$
\frac{d\pi(\tau_2)}{d\tau_2} < \pi_1 m_1 [\max_{0 \leq \theta \leq D} f(\theta)]^2 [(D - \delta) - \tau_2].
$$

Therefore,

$$
\frac{d\pi(\tau_2)}{d\tau_2} < 0, \text{ if } (D - \delta) \leq \tau_2 \leq D.
$$

Given that $\pi(\tau_2)$ is monotonically decreasing with $\tau_2 \in [D - \delta, D]$, we conclude that

$$
\tau_2^* < (D - \delta), \text{ if } (D - \delta) > \alpha_2 \text{ or equivalently, } \delta < D - \alpha_2, \text{ and}
$$

$$
\tau_2^* = \alpha_2, \text{ if } (D - \delta) \leq \alpha_2 \text{ or equivalently, } \delta \geq D - \alpha_2.
$$

Proof of Proposition 2

To prove that G2 should be introduced as early as possible, we only need to show

$$
\frac{d\pi(\tau_2)}{d\tau_2} < 0, \forall \tau_2 \in [0, D],
$$

where $\frac{d\pi(\tau_2)}{d\tau_2}$ has the same expression shown in (A1). First, with $\theta \in [\tau_2, D]$, we have
0 \leq (\theta - \tau_2) \leq (D - \tau_2) \leq D \leq T^* = \ln(q/p)/(p + q).

Since \( f(t) \) is a monotonically increasing function of \( t \) when \( t \leq T^* \), we have

\[
f(\theta - \tau_2) \leq f(D - \tau_2) \Rightarrow 
\int_{\tau_2}^{D} f(\theta)f(\theta - \tau_2)d\theta < \int_{\tau_2}^{D} f(\theta)f(D - \tau_2)d\theta = [F(D) - F(\tau_2)]f(D - \tau_2) \leq F(D)f(D - \tau_2).
\]

From \( \pi_1 \leq \pi_2 \), we further conclude

\[
\pi_1 m_1 \int_{\tau_2}^{D} f(\theta)f(\theta - \tau_2)d\theta < \pi_2 m_1 F(D)f(D - \tau_2),
\]

which leads to

\[
\frac{d\pi(\tau_2)}{d\tau_2} = \pi_1 m_1 \int_{\tau_2}^{D} f(\theta)f(\theta - \tau_2)d\theta - \pi_2 [m_2 + m_1 F(D)]f(D - \tau_2) < 0, \forall \tau_2 \in [0, D].
\]

**Proof of Proposition 3**

The objective function of (22) can be rearranged to

\[
\pi(\tau_2) = \varphi_1 m_1 \int_0^{\tau_2} F(\theta)d\theta + (\varphi_2 - \varphi_1)m_1 \int_{\tau_2}^{D} F(\theta)f(\theta - \tau_2)d\theta + \varphi_2 m_2 \int_{\tau_2}^{D} F(\theta - \tau_2)d\theta.
\]

Therefore,

\[
\frac{d\pi(\tau_2)}{d\tau_2} = (\varphi_2 - \varphi_1)m_1 \left[-F(\tau_2)F(\tau_2 - \tau_2) - \int_{\tau_2}^{D} F(\theta)f(\theta - \tau_2)d\theta \right]
+ \varphi_2 m_2 \left[-F(\tau_2 - \tau_2) - \int_{\tau_2}^{D} f(\theta - \tau_2)d\theta \right]

= -(\varphi_2 - \varphi_1)m_1 \int_{\tau_2}^{D} F(\theta)f(\theta - \tau_2)d\theta - \varphi_2 m_2 F(D - \tau_2). \quad (A2)
\]

If \( \varphi_1 \leq \varphi_2 \), we have

\[
\begin{align*}
\frac{d\pi(\tau_2)}{d\tau_2} &< 0, \text{ if } \tau_2 < D, \\
\frac{d\pi(\tau_2)}{d\tau_2} &> 0, \text{ if } \tau_2 = D.
\end{align*}
\]

Since \( \pi(\tau_2) \) decreases monotonically as \( \tau_2 \) increases, G2 should be introduced to the market as early as possible, i.e., \( \tau_2^* = \alpha_2 \).
Proof of Proposition 4

If \( \varphi_1 > \varphi_2 \), from (A2) we have

\[
\frac{d\pi(\tau_2)}{d\tau_2} = (\varphi_1 - \varphi_2)m_1 \int_{\tau_2}^{D} F(\theta) f(\theta - \tau_2) d\theta - \varphi_2 m_2 F(D - \tau_2)
\]

\[
< (\varphi_1 - \varphi_2)m_1 \int_{\tau_2}^{D} F(D) f(\theta - \tau_2) d\theta - \varphi_2 m_2 F(D - \tau_2)
\]

\[
= [((\varphi_1 - \varphi_2)m_1 F(D) - \varphi_2 m_2]F(D - \tau_2).
\]

Therefore,

\[
\frac{d\pi(\tau_2)}{d\tau_2} < 0 \text{ if } (\varphi_1 - \varphi_2)m_1 F(D) \leq \varphi_2 m_2,
\]

or equivalently,

\[
\frac{d\pi(\tau_2)}{d\tau_2} < 0 \text{ if } m_2 \geq \frac{(\varphi_1 - \varphi_2)F(D)}{\varphi_2} m_1.
\]

Thus

\[
\tau^*_2 = \alpha_2 \text{ if } m_2 \geq \frac{(\varphi_1 - \varphi_2)F(D)}{\varphi_2} m_1.
\]

Proof of Proposition 5

We first examine the first-order condition \( \frac{d\pi(\tau_2)}{d\tau_2} = 0 \). Since \( H(x) \) in (27) is always negative, we conclude that the first-order condition requires

\[
ax^2 + bx + c = 0. \quad (A3)
\]

Note that here \( x = e^{(p+q)\tau_2} \geq 1 \) for \( \forall \tau_2 \geq 0 \).

We examine three conditions, based on how the value of \( c \) compares with the other parameters.

(I) \( c > \frac{b^2}{4a} \). This condition implies \( b^2 - 4ac < 0 \). Under this scenario, (A3) has no real solution. From \( a > 0 \), we have

\[
ax^2 + bx + c = 0 > 0, \forall x.
\]

Since \( H(x) < 0 \), we conclude

\[
\frac{d\pi(\tau_2)}{d\tau_2} < 0, \forall \tau_2 \geq 0.
\]
Because the net profit decreases monotonically as the introduction of G2 is delayed, it is optimal to introduce generation 2 as early as possible, i.e.,

$$\tau_2^* = \alpha_2.$$  

(II) \(0 \leq c \leq b^2/(4a)\). Under this scenario, \(b^2 - 4ac \geq 0\), hence (A3) has a real solution:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$  

Since \(c \geq 0\), we must have \((b^2 - 4ac) \leq b^2\). Under this condition, the root(s) of (A3) are non-positive. Hence,

$$ax^2 + bx + c = 0 > 0, \, \forall x \geq 1.$$  

$$\Rightarrow \frac{d\pi(\tau_2)}{d\tau_2} < 0, \, \forall \tau_2 \geq 0.$$  

Therefore,

$$\tau_2^* = \alpha_2.$$  

Based on scenarios (I) and (II), we conclude

$$c \geq 0 \Rightarrow \tau_2^* = \alpha_2.$$  

(III) \(c < 0\). This condition leads to \(b^2 - 4ac \geq b^2\). In this case, the two roots of (A3) are

$$\begin{cases}  
x_1 = \frac{-b - \sqrt{b^2 - 4ac}}{2a} < 0, \\
x_2 = \frac{-b + \sqrt{b^2 - 4ac}}{2a} > 0. 
\end{cases}$$  

In this case, if \(x_2 \leq e^{(p+q)\alpha_2}\), or equivalently,

$$c \geq -ae^{(p+q)\alpha_2} - be^{(p+q)\alpha_2},$$  

we still have

$$\frac{d\pi(\tau_2)}{d\tau_2} \leq 0, \, \forall x \geq e^{(p+q)\alpha_2}, \text{ or } \forall \tau_2 \geq \alpha_2.$$  

Therefore,
\[ \tau_2^* = \alpha_2. \]

On the other hand, if \( x_2 > e^{(p+q)\alpha_2} \), or equivalently,
\[ c < -ae^{2(p+q)\alpha_2} - be^{(p+q)\alpha_2}, \]
we have
\[ \frac{d\pi(x_2)}{d\tau_2} \geq 0, \forall x \leq x_2, \text{ and} \]
\[ \frac{d\pi(x_2)}{d\tau_2} < 0, \forall x > x_2. \]

Hence,
\[ \tau_2^* = \frac{\ln(x_2)}{p+q}, \text{ or } \tau_2^* = \frac{\ln(-b + \sqrt{b^2 - 4ac})}{2a} / (p + q). \]

Taking into consideration all scenarios, the solution is
\[ \tau_2^* = \begin{cases} \frac{\ln(-b + \sqrt{b^2 - 4ac})}{2a} / (p + q), & \text{if } c < -ae^{2(p+q)\alpha_2} - be^{(p+q)\alpha_2}, \\ \alpha_2, & \text{otherwise.} \end{cases} \]

**Proof of Proposition 6**

The objective function of (29) can be rearranged to
\[ \pi(x_2) = \varphi_1 m_1 \int_0^{x_2} F(\theta) d\theta + \varphi_1 \int_{\tau_2}^{D} m_1 F(\tau_2)[1 - F(\theta - \tau_2)]d\theta \]
\[ + \varphi_2 \int_{\tau_2}^{D} \{[m_2 + m_1 F(\tau_2)]F(\theta - \tau_2) + m_1 [F(\theta) - F(\tau_2)]\}d\theta \]
\[ = \varphi_1 m_1 \int_0^{D} F(\theta) d\theta - \varphi_1 m_1 \int_{\tau_2}^{D} F(\theta) d\theta + \int_{\tau_2}^{D} \{(\varphi_1 - \varphi_2) m_1 F(\tau_2) + \varphi_2 m_2 F(\theta - \tau_2) \}
\]
\[ +(\varphi_2 - \varphi_1) m_1 F(\tau_2) F(\theta - \tau_2) + \varphi_2 m_1 F(\theta) d\theta \]
\[ = \varphi_1 m_1 \int_0^{D} F(\theta) d\theta + \int_{\tau_2}^{D} \{(\varphi_1 - \varphi_2) m_1 F(\tau_2) + \varphi_2 m_2 F(\theta - \tau_2) \}
\]
\[ +(\varphi_2 - \varphi_1) m_1 F(\tau_2) F(\theta - \tau_2) + (\varphi_2 - \varphi_1) m_1 F(\theta) d\theta. \]
Then,
\[
\frac{d\pi(\tau_2)}{d\tau_2} = -((\varphi_1 - \varphi_2)m_1 F(\tau_2) + \varphi_2 m_2 F(\tau_2 - \tau_2) + (\varphi_2 - \varphi_1)m_1 F(\tau_2) F(\tau_2 - \tau_2)
\]
\[
+ (\varphi_2 - \varphi_1)m_1 F(\tau_2)) + \int_{\tau_2}^{D} \{((\varphi_1 - \varphi_2)m_1 f(\tau_2) - \varphi_2 m_2 f(\theta - \tau_2)
\]
\[-(\varphi_2 - \varphi_1)m_1 F(\tau_2)f(\theta - \tau_2) + (\varphi_2 - \varphi_1)m_1 f(\tau_2)F(\theta - \tau_2)\}d\theta
\]
\[
= (\varphi_1 - \varphi_2)m_1 f(\tau_2)(D - \tau_2) - \varphi_2 m_2 F(D - \tau_2) - (\varphi_2 - \varphi_1)m_1 F(\tau_2)F(D - \tau_2)
\]
\[
+ (\varphi_2 - \varphi_1)m_1 f(\tau_2) \int_{\tau_2}^{D} F(\theta - \tau_2)d\theta.
\]

If \(\varphi_1 \leq \varphi_2\), we have
\[
\frac{d\pi(\tau_2)}{d\tau_2} < (\varphi_1 - \varphi_2)m_1 f(\tau_2)(D - \tau_2) - \varphi_2 m_2 F(D - \tau_2) - (\varphi_2 - \varphi_1)m_1 F(\tau_2)F(D - \tau_2)
\]
\[
+ (\varphi_2 - \varphi_1)m_1 f(\tau_2)) \int_{\tau_2}^{D} d\theta
\]
\[
= (\varphi_1 - \varphi_2)m_1 f(\tau_2)(D - \tau_2) - \varphi_2 m_2 F(D - \tau_2) - (\varphi_2 - \varphi_1)m_1 F(\tau_2)F(D - \tau_2)
\]
\[
+ (\varphi_2 - \varphi_1)m_1 f(\tau_2)(D - \tau_2)
\]
\[
= -\varphi_2 m_2 F(D - \tau_2) - (\varphi_2 - \varphi_1)m_1 F(\tau_2)F(D - \tau_2).
\]

Hence,
\[
\begin{cases}
\frac{d\pi(\tau_2)}{d\tau_2} < 0, & \text{if } \tau_2 < D, \\
\frac{d\pi(\tau_2)}{d\tau_2} = 0, & \text{if } \tau_2 = D.
\end{cases}
\]

Therefore, it is optimal to release G2 as early as possible, i.e., \(\tau^*_2 = \alpha_2\).

**Proof of Proposition 7**

If \(\varphi_1 > \varphi_2\), we have
\[
\frac{d\pi(\tau_2)}{d\tau_2} = -\varphi_2 m_2 F(D - \tau_2) - (\varphi_2 - \varphi_1)m_1 F(\tau_2)F(D - \tau_2)
\]
\[
+ (\varphi_1 - \varphi_2)m_1 f(\tau_2) \int_{\tau_2}^{D} [1 - F(\theta - \tau_2)]d\theta.
\]
\[
\frac{d\pi(\tau_2)}{d\tau_2} \geq 0
\]

\[\Leftrightarrow -\varphi_2 m_2 F(D - \tau_2) + (\varphi_1 - \varphi_2) m_1 F(\tau_2) F(D - \tau_2) + (\varphi_1 - \varphi_2) m_1 f(\tau_2) \int_{\tau_2}^{D} [1 - F(\theta - \tau_2)] d\theta \geq 0\]

\[\Leftrightarrow \varphi_2 m_2 F(D - \tau_2) \leq (\varphi_1 - \varphi_2) m_1 F(\tau_2) F(D - \tau_2) + (\varphi_1 - \varphi_2) m_1 f(\tau_2) \int_{\tau_2}^{D} [1 - F(\theta - \tau_2)] d\theta\]

\[\Leftrightarrow m_2 \leq \frac{(\varphi_1 - \varphi_2) m_1}{\varphi_2} \left[ F(\tau_2) + \frac{f(\tau_2)}{F(D - \tau_2)} \int_{\tau_2}^{D} [1 - F(\theta - \tau_2)] d\theta \right].\]

The second term on the RHS of the equation is always positive, i.e.,

\[F(\tau_2) + \frac{f(\tau_2)}{F(D - \tau_2)} \int_{\tau_2}^{D} [1 - F(\theta - \tau_2)] d\theta > 0, \forall \tau_2 \in [\alpha_2, D].\]

We let

\[\mu = \min_{0 \leq \tau_2 \leq D} f \left[ F(\tau_2) + \frac{f(\tau_2)}{F(D - \tau_2)} \int_{\tau_2}^{D} [1 - F(\theta - \tau_2)] d\theta \right].\]

Then,

\[
\frac{d\pi(\tau_2)}{d\tau_2} \geq 0, \quad \text{if } m_2 \leq \frac{(\varphi_1 - \varphi_2) m_1}{\varphi_2} \mu.
\]

Similarly, if we let

\[\nu = \max_{0 \leq \tau_2 \leq D} f \left[ F(\tau_2) + \frac{f(\tau_2)}{F(D - \tau_2)} \int_{\tau_2}^{D} [1 - F(\theta - \tau_2)] d\theta \right].\]

It can be shown that

\[
\frac{d\pi(\tau_2)}{d\tau_2} \leq 0, \quad \text{if } m_2 \geq \frac{(\varphi_1 - \varphi_2) m_1}{\varphi_2} \nu.
\]