Now or Never Revisited: An Analysis of Market Entry Timing for Successive Product Generation

Zhengrui Jiang  
Iowa State University  
zjiang@iastate.edu

Dipak C. Jain  
INSEAD  
dipak.c.jain@insead.edu

Shaobai Jiang  
Iowa State University  
shjiang@iastate.edu

Abstract
Determining the optimal market entry timing for successive technological innovations is a critical decision for firms. Pioneering studies dealing with this issue have focused on one-time sale (e.g., HDTV), and concluded that a new product should be introduced to the market either now or never, or now or at maturity. However, these prior studies do not examine another commonly seen business practice — revenue is generated from continuous services (e.g., Office 365). In this research, we derive the optimal market entry timing under both one-time sale and continuous service, and check whether the prior findings remain valid under today’s diverse market landscape. We find that under one-time sale, the optimal entry timing is not limited to now, maturity, or never; but it can also lie between now and maturity. More interestingly, our results show that the now or never rule holds only under a scenario not considered in the prior studies.

Keywords
Multi-generation diffusion, market entry timing, business revenue models

Introduction
Most of the IT products we consume today represent improved versions of earlier generations, and such products over time will be substituted by even newer generations. Driven by technological advancements and market needs, the release of successive product generations is frequently observed in the marketplace. Well known examples include major releases of Microsoft Windows and Apple iPhones.

In the presence of successive product generations, when to introduce a new generation to the market is a critical decision for firms. The primary goal of this study is to develop analytical models to decide the optimal market entry timing and compare our findings with those reported in the prior literature.

Market entry timing for successive product versions has been studied in the marketing literature. In one stream of research, Wilson and Norton (1989) develop a multigeneration diffusion model to help decide the optimal market entry timing for a product line extension. They find that in most cases a product line extension should be introduced either at the same time as the main product (now) or not be introduced during the planning horizon (never). Under a different set of assumptions, Mahajan and Muller (1996) analyze market entry timing for an improved new generation. Their general conclusion is that the second generation should be introduced at the same time as the first (now) or when the sale of the first generation has reached its maturity stage (at maturity).

1 For brevity, the term product includes both products and services.
2 The term maturity is not formally defined by the authors. Informally, it refers to the stage after the time of peak adoption.
A second stream of research has taken into account product quality and consumers’ valuations, and has used game-theoretical models to determine the market launch timing for new product versions. For instance, Moorthy and Png (1992) model consumers’ valuations of high-end and low-end versions (e.g., hardcover and paperback books) of the same product. The authors find that sequential is better than simultaneous introduction when cannibalization is a serious problem and the seller is less impatient than the customers. Otherwise simultaneous introduction is preferred.

These two streams of research both assume that firms generate their revenue by charging a one-time price from customers at the time of product sale and no additional revenue is generated afterwards. Examples include the sale of TV sets and computers. We refer to this as the one-time sale (OTS) revenue model.

In today’s market, in addition to OTS, another practice is becoming increasingly common. Some firms generate revenue by providing an ongoing service to their customers, where the fee charged depends on the duration or frequency of the service. For instance, Microsoft, which used to primarily reply on one-time sale of its products, is promoting its subscription-based Office 365 to meet the demand of mobile device users. This is referred to as the continuous service (CNS) revenue model.

In addition to the two revenue models, firms adopt different generation transition strategies. There are cases where after the new generation enters the market, the old generations continue to be sold as long as there is sufficient demand. For instance, digital cellular phone service was introduced in the early 1990s in the U.S, but analog service continued until early 2000. This is referred to as phase-out transition. In some markets, we observe a different practice — a firm discontinues the production and/or sale of the old generation as soon as a new generation is introduced. For example, Microsoft stopped selling older Office versions as soon as a new version is released. We term this generation transition strategy as total transition.

The two revenue models and two generation transition strategies lead to four business scenarios, as presented in Table 1.

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

Table 1. Business Scenarios Corresponding to Revenue Models and Transition Strategies

The four business scenarios illustrated in this table represent a more diverse market landscape than the one (i.e., Scenario I) analyzed by Wilson and Norton (1989) and Mahajan and Muller (1996) decades ago. The primary motivation of the present research is to revisit the entry timing decision and examine whether the findings of the prior studies still hold under these different business scenarios.

We next present the key modeling framework of this study.

**Modeling Framework**

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; switching, on the other hand, represents the behavior of existing adopters of the immediate previous generation making an upgrade to a new generation.

In the diffusion literature, several multigeneration models (e.g., Mahajan and Muller 1996, Jun and Park 1999, Danaher, Hardie, and Putsis 2001, Jiang and Jain 2012) explicitly capture leapfrogging and switching based on a diffusion or choice modeling framework. We adopt and extend the Generalized Norton-Bass (GNB) model (Jiang and Jain 2012) because this model provides closed-form expressions for both the number of units-in-use and the instantaneous adoption rate, and can help project the profit for all considered business scenarios.
**Profit Projection under Phase-out Transition**

As mentioned earlier, the GNB model provides closed-form expressions for the number of units-in-use and the adoption rate. As an example, the adoption rate curve represents the rate of initial adoptions of a cellular service (e.g., analog or digital), while the units-in-use curve captures the number of active subscribers of the service. Figures 1a and 1b illustrate the key differences between the number of units-in-use and the adoption rate for a two-generation case.

Without loss of generality, we assume that generation 1 (G1) is introduced at time 0 and generation 2 (G2) at time $\tau_2 \geq 0$. 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 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.

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

$$
S_1(t) = m_1 F_1(t) - m_1 F_1(t) F_2(t - \tau_2) = m_1 F_1(t) [1 - F_2(t - \tau_2)],
$$

$$
S_2(t) = m_2 F_2(t - \tau_2) + m_1 F_1(t) F_2(t - \tau_2) = [m_2 + m_1 F_1(t)] F_2(t - \tau_2),
$$

The instantaneous adoption rates for the two generations are

$$
y_1(t) = m_1 F_1(t)[1 - F_2(t - \tau_2)],
$$

$$
y_2(t) = [m_2 + m_1 F_1(t)] F_2(t - \tau_2) + m_1 F_1(t) F_2(t - \tau_2),
$$

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

$$
Y_1(t) = m_1 F_1(t) - m_1 \int_{\tau_2}^{t} F_1(\theta) F_2(\theta - \tau_2) d\theta,
$$

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

In Equations (1) – (6), $m_1$ represents the market potential for generation 1, and $m_2$ is the incremental market potential specific to generation 2, i.e., potential adopters who are only interested in generation 2. $F_G(t)$ and $f_G(t)$ denotes the cumulative and noncumulative diffusion rates, both in terms of the fraction of potential adopters, for generation $G$ ($G = 1, 2$). Specifically,

$$
F_G(t) = \int_{0}^{t} f_G(\theta) d\theta = \begin{cases} 
0, & t < 0, \\
\frac{1 - e^{-(p_G+q_G)t}}{(p_G+q_G)}e^{-(p_G+q_G)(t+1)}, & t \geq 0.
\end{cases}
$$

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 derive the optimal market entry timing, we consider a planning horizon, denoted by $D$. During this planning horizon, we assume that the cost and the price of a product (service) both increase at the same
rate as the discount rate. Therefore, the time-discounted profit per unit sale or per unit time’s service remains constant during the planning horizon.\(^3\)

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 generation G by \(\pi_G\). Similarly, we define *unit contribution margin (for service)* as the present value of the profit generated from providing one unit time’s service for one customer, and denote the unit contribution margin for generation G by \(\varphi_G\). We assume that all profit margins are positive, i.e., \(\pi_G > 0\), \(\varphi_G > 0\).

Under one-time sale (OTS), the profit at any given time is proportional to the adoption rate at that time, hence the total time-discounted profit for the two product generations during the entire planning horizon (from time 0 to \(D\)) equals

\[
\pi(t) = \pi_1 \int_0^D y_1(\theta)d\theta + \pi_2 \int_{\tau_2}^D y_2(\theta)d\theta = \pi_1 Y_1(D) + \pi_2 Y_2(D)
\]

(8)

When 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, therefore the total profit during the planning horizon is

\[
\pi(t) = \varphi_1 \int_0^D s_1(\theta)d\theta + \varphi_2 \int_{\tau_2}^D s_2(\theta)d\theta
\]

(9)

Equations (8) and (9) both assume that the fixed cost of introducing a new generation is insignificant when compared to the variable costs and the revenues generated from product sale or service, hence the fixed cost is not considered in our analysis. This assumption also ensures a fair comparison between our findings and those of Wilson and Norton (1989) and Mahajan and Muller (1996), because the same assumption is also implicitly adopted by the two prior studies.

**Profit Projection under Total Transition**

As stated earlier, the GNB model considers only phase-out transition; we now extend it for profit projection under total transition. In Equations (1)-(4), the term \(F_2(t - \tau_2)\) represents the *leapfrogging multiplier*, i.e., the proportion of potential adopters who leapfrog to G2. Under total transition, since G1 is discontinued once G2 is introduced, 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

\[
\begin{align*}
\dot{y}_1(t) &= \begin{cases} 
m_1 f_1(t), & t < \tau_2, \\
0, & t \geq \tau_2,
\end{cases} \\
\dot{y}_2(t) &= \begin{cases} 
m_2 + m_1 F_1(\tau_2), & t = \tau_2, \\
m_2 + m_1 F_1(t), & t > \tau_2,
\end{cases}
\end{align*}
\]

(10)

(11)

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

\[
\begin{align*}
\hat{y}_1(t) &= \begin{cases} 
m_1 F_1(t), & t < \tau_2, \\
0, & t \geq \tau_2,
\end{cases} \\
\hat{y}_2(t) &= \begin{cases} 
m_2 + m_1 F_1(\tau_2) - m_1 F_1(t), & t = \tau_2, \\
m_2 + m_1 F_1(\tau_2), & t > \tau_2,
\end{cases}
\end{align*}
\]

(12)

(13)

We next derive the number of units-in-use for the two generations. 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.

\(\text{---}\)

\(^3\) This assumption is adopted primarily for mathematical tractability. Numerical analyses show that even if the assumption is relaxed, the key results of this study still remain valid qualitatively.
Similar to Equations (8) and (9), under total transition, the total profits corresponding to one-time sale (OTS) and continuous service (CNS) are

\( \pi(t_2) = \pi_1 Y_1(t_2) + \pi_2 Y_2(t_2), \) and

\( \pi(t_2) = \varphi_1 \int_0^D \hat{S}_1(\theta)d\theta + \varphi_2 \int_{t_2}^D \hat{S}_2(\theta)d\theta. \) (17)

We next derive the optimal market entry timing for the four business scenarios shown in Table 1.

**Market Entry Timing under One-Time Sale (OTS)**

We now analyze the two-generation case under the One-Time Sale (OTS) revenue model. Our goal is to find the market entry timing for the second generation (G2) that maximizes the total profit. The two generation transition strategies, i.e., phase-out transition and total transition, are separately examined.

**Scenario I: Phase-out Transition**

As explained earlier, many consumer products (e.g., computers, TVs) fall under Scenario I. The total profit for this business scenario is given in Equation (8); hence the decision problem for deciding the optimal market entry time for G2 is formulated as

\[ \max_{a_2 \geq t_2 \leq D} \pi(t_2) = \pi_1 Y_1(D) + \pi_2 Y_2(D), \] (18)

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

Regarding the values of the coefficients of innovation and imitation across generations, the prior literature has adopted different assumptions and reported different empirical findings. In a recent study, based on data for 39 product generations in twelve product markets, Stremersch et al. (2010) find that the changes in the coefficients of innovation and imitation across generations are insignificant for all but one product category (steel making). 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 (18) then becomes

\[ \max_{a_2 \geq t_2 \leq D} \pi(t_2) = \pi_1 \left[ m_1 F(D) - m_1 \int_{a_2}^D f(D) F(D - t_2)d\theta \right] + \pi_2 \left[ m_2 + m_2 F(D) \right] F(D - t_2). \] (19)

Using 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.

For the optimal market entry timing under Scenario I, we find the following result:

**Proposition 1. Under one-time sale and phase-out transition, introducing the new generation as early as possible is an optimal solution, while not introducing it during the planning horizon is not.**

We now compare Proposition 1 with findings from the prior literature. The first part of Proposition 1 is consistent with the “now” solution observed by Wilson and Norton (1989) and Mahajan and Muller (1996). In particular, our results show that if the unit contribution margin for the second generation is

---

4 Proofs and some derivations are omitted from the paper because of space limitation and are available from the authors upon request.
equal to or greater than that for the first generation \((\pi_2 \geq \pi_1)\), and the planning horizon is shorter than the time of peak sales for the first generation, then it is optimal to introduce the second generation as early as possible.

We find that “never” introducing the new generation during the planning horizon is not an optimal solution. An explanation for this finding is as follows. Assuming that the expected profit per unit sale is fixed for both generations, the total profit depends on the numbers of adoptions of G1 and G2 and their relative unit contribution margins. When the unit contribution margin for G2 is at least as high as that for G1 \((\pi_1 \leq \pi_2)\), introducing G2 at any 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.

In case the unit contribution margin for G2 is less than that for G1 \((\pi_1 > \pi_2)\), 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, and the total benefit can exceed the total cost of introducing G2 before the end of the planning horizon.

We believe that this finding is consistent with what we observe in the marketplace. Many firms would release a new product generation as soon as it is ready for the market. However, there are many reported examples that firms intentionally delayed the introduction of a new innovation to achieve a higher benefit. Examples include Intel Camino chipset, DVD video recording, and 3G Cellular networks (Wang and Hui 2005, 2010). These examples show that the entry timing can indeed lie between “now” and the end of the planning horizon.

To gain a better understanding of the analytical results, we perform numerical analyses to further examine how the market entry timing of G2 affects the generational adoptions and the total profit. In order to have a broad representation of today’s market, we estimate the Bass model parameter values based on the 1999-2011 sales data for three popular consumer electronics products (standard cell phone, digital TV, and MP3 player) and adopt their averages, i.e., \(p = 0.00855\) and \(q = 0.429\). The market potentials for G1 and G2 are set to \(m_1 = m_2 = 10\) million. The unit contribution margins are assumed to be \(\pi_1 = \pi_2 = $100\). In addition, we assume that G2 is available for market introduction at time zero.

![Figure 2. How the Profit and Cross-Generation Adoptions Change with Entry Timing](image)

(a) \(D = 10\) Years

(b) \(D = 20\) Years

We first try two different planning horizon at \(D = 10\) and 20 years, and record how the total profit changes with the entry timing of G2. The results are shown in Figures 2(a) and 2(b). From the figures, we observe that the total profit decreases monotonically when \(D = 10\) years, whereas it first increases and then decreases when \(D = 20\) years. Therefore, when the planning horizon is \(D = 10\) years, the optimal market entry timing for G2 is \(\tau_2^* = 0\), implying that it is optimal to introduce G2 at the same time as G1, and the resulting total profit is \(\pi^* = $1.39\) billion. When the planning horizon is extended to \(D = 20\) years, the optimal market entry timing changes to \(\tau_2^* = 4.73\) years, and the corresponding profit becomes \(\pi^* = $2.68\) billion.

In another set of analyses, we adopt the average parameter values \((p=0.03, q=0.38)\) reported in a meta-analysis by Sultan et al. (1990), and the results are found to be qualitatively similar.
Proposition 1 concludes that *never* cannot be optimal entry timing an optimal solution under Scenario I. The numerical solutions further confirm that the optimal entry timing for G2 can be *now*, before *maturity* (i.e., $T^* = 8.95$ years for G1), or after *maturity*. This differs from the findings reported by both Wilson and Norton (1989) and Mahajan and Muller (1996).

**Scenario II: Total Transition**

Under total transition, since G1 is discontinued after the introduction of G2, all potential adopters who would have adopted G1 will leapfrog to G2 instead. Neither Wilson and Norton (1989) nor Mahajan and Muller (1996) have considered total transition in their studies.

The total profit under Scenario II can be obtained based on Equation (16). Therefore, the problem for deciding the profit-maximizing market entry timing for G2 is formulated as:

$$\max_{\tau_2 \leq D} \pi_2(t_2) = \pi_1 \hat{Y}_1(t_2) + \pi_2 \hat{Y}_2(D),$$  

where $\hat{Y}_1(t)$ and $\hat{Y}_2(t)$ are defined in Equations (12) and (13), 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 $f(t) = f_1(t) = f_2(t)$. Then, (20) becomes

$$\max_{\tau_2 \leq D} \pi_2(t_2) = \pi_1 m_1 F(t_2) + \pi_2 m_2 + m_4 F(t_2) [F(D - \tau_2) + \pi_2 m_4 [F(D) - F(t_2)].$$

Unlike under Scenario I (one-time sale and phase-out transition), we are able to obtain a closed-form solution for (21). In addition, we find analytically that *never* could be an optimal solution under Scenario II, as stated in the following proposition.

**Proposition 2. Under one-time sale and total 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 sometime during the planning horizon. If the unit contribution margin for the second generation is less than that for the first generation, not introducing the second generation during the planning horizon could be an optimal solution.

When the unit contribution margin for G2 is at least a high as that for G1 ($\pi_1 \leq \pi_2$), the conclusions of Proposition 1 and Proposition 2 are the same, and the interpretations for Scenario I (see discussion after Proposition 1) remain valid for Scenario II. In case $\pi_1 > \pi_2$, the difference findings for Scenario I and II can be explained as follows.

Under Scenario II, introducing G2 is less profitable for two reasons. First, all else being equal, there are more leapfrogging under Scenario II than under Scenario I. Since $\pi_2 < \pi_1$, more leapfrogging leads to higher revenue loss. Second, although the introduction of G2 can lead to switching and hence cross-generation repeat purchases, all else being equal, the number of repeat purchases is lower under Scenario II than under Scenario I. This implies that the benefit derived from repeat purchases is lower. With both factors considered, it is clear that the cost of introducing G2 is higher and the benefit is lower under Scenario II than under Scenario I, hence not introducing G2 could be an optimal solution for Scenario II.

To further illustrate, we adopt the same parameter values used in the previous subsection. With a short planning horizon of 10 years, the optimal solution is obvious, i.e., $\tau_2 = 0$. When the planning horizon increases to 20 years, the optimal market entry timing for G2 equals to $\tau_2 = 7.40$ years. Again, as the duration of planning horizon increases, it is beneficial to delay the market entry timing.

We also examine the less likely scenario with $\pi_1 > \pi_2$. Specifically, we let $\pi_1 = $100, $\pi_2 = $40, and $D = 10$ years. The optimal solution is found to be $\tau_2 > 10$ years, implying that G2 should not be introduced during the planning horizon, a result consistent with Proposition 2.

**Market Entry Timing under Continuous Service (CNS)**

We now derive the market entry timing for the two-generation case under continuous service (CNS). Unlike one-time sale (OTS), with CNS a customer does not pay a one-time fee to gain permanent access to a service; instead, the fee is calculated based on how long the customer consumes the service. From a modeling perspective, a key difference between OTS and CNS is that under the former, the profit at any
given point in time depends on the instantaneous adoption rate; while for the latter, the profit depends on the number of units-in-use at any given time.

In terms of revenue implications, there are two important differences between CNS and OTS. First, under CNS, whether a customer adopts G2 through leapfrogging or switching does not affect the firm’s revenue from G2 because the revenue is not generated through one-time sale. Under OTS, however, switching is more beneficial than leapfrogging. 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 revenues.

We examine both Scenario III and Scenario IV in Table 1. Under Scenario III, the total profit can be estimated from Equation (9). Therefore, the optimal market entry time for G2 can be obtained by

$$\max_{\alpha_2 \leq \tau_2 \leq D} \pi(\tau_2) = \varphi_1 \int_0^D S_1(\vartheta)d\vartheta + \varphi_2 \int_{\tau_2}^D S_2(\vartheta)d\vartheta,$$

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

For Scenario IV, the total profit can be estimated from Equation (17). Hence the decision problem is formulated as

$$\max_{\alpha_2 \leq \tau_2 \leq D} \pi(\tau_2) = \varphi_1 \int_0^D \hat{S}_1(\vartheta)d\vartheta + \varphi_2 \int_{\tau_2}^D \hat{S}_2(\vartheta)d\vartheta,$$

where \(\hat{S}_1(t)\) and \(\hat{S}_2(t)\) are given in Equations (14) and (15), respectively.

We again let \(p = p_1 = p_2\) and \(q = q_1 = q_2\). Problem (21) then becomes

$$\max_{\alpha_2 \leq \tau_2 \leq D} \pi(\tau_2) = \varphi_1 \int_0^D m_1 F(\vartheta)[1 - F(\vartheta - \tau_2)]d\vartheta + \varphi_2 \int_{\tau_2}^D [m_2 + m_1 F(\vartheta)] F(\vartheta - \tau_2)d\vartheta,$$

and problem (28) changes to

$$\max_{\alpha_2 \leq \tau_2 \leq D} \pi(\tau_2) = \varphi_1 m_1 \int_0^{\tau_2} F(\vartheta)d\vartheta + \varphi_1 \int_{\tau_2}^D \hat{S}_1(\vartheta)d\vartheta + \varphi_2 \int_{\tau_2}^D \hat{S}_2(\vartheta)d\vartheta.$$

It is worth noting that under CNS, whether the generation strategy is phase-out or total transition is less critical than under OTS. Furthermore, if the unit contribution margins for the two generations are close, when a customer leapfrog/switch from G1 and G2 has little impact on the total profit. Therefore, despite the difference in model formulations, our analytical and numerical findings are similar for Scenarios III and IV. For this reason, unless necessary, we do not differentiate Scenarios III and IV in the remaining discussion.

We would like to emphasize that under the CNS revenue model, the total profit depends on not only the number of adopters of each service, but also the duration of each service being consumed. Therefore, all else being equal, delaying the introduction of G2 is more costly under CNS because it reduces the average duration of G2 service.

For most service types, G2 is expected to be at least as profitable as G1 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, 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.

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) or Equations (14) and (15) that the sum of the numbers of leapfrogs and switchings can only increase. Because G2 is at least as profitable as G1, more leapfrogs or switchings from G1 to G2 can never decrease the revenue. 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. Therefore, G2 should be introduced as soon as possible.
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.

For numerical illustration, we adopt a dataset that includes the numbers of analog and digital cellular phone subscribers in the US. The estimated parameter values are \( p = 0.0158, q = 0.279, m_1 = 38.68 \) million, and \( m_2 = 318.06 \) million. The unit contribution margin for one year’s cellular service is set to \( \varphi_1 = \varphi_2 = $200 \). We again assume that G2 is available for release at time zero. We find that regardless of the duration of planning horizon (e.g., \( D=10, 20 \) years), the optimal introduction time for G2 is always \( \tau_2^* = 0 \), a result consistent with Proposition 3.

To understand the solution under the less likely scenario where G2 is less profitable than G1, we let \( \varphi_1 = $200 \) and \( \varphi_2 = $100 \), and vary the value of \( m_2 \), which represents the incremental market size for G2 after it is introduced. The planning horizon is fixed at \( D = 10 \) years.

As shown in Figure 3(a), 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, i.e., now. When \( m_2 \) decreases to 12 million, the total profit first decreases and then increases as G2’s introduction is postponed (see Figure 3(b)). Since the highest profit is achieved at \( \tau_2^* = 0 \) years, G2 should again be introduced now. When \( m_2 \) is further reduced to \( m_2 = 11 \) million, as shown in Figure 3(c), the impact of the entry time on the total profit is also non-monotonic; the highest profit, however, is achieved at \( \tau_2^* > 10 \) years, hence G2 should not be introduced during the planning horizon, i.e., never.

![Figure 3. Now or Never Depending on the Incremental Market Size for G2](image)

To verify whether the optimal market entry timing can lie between now and never, we vary the values of \( m_2 \) by increasingly smaller incrementals, and find that the optimal solution still exhibits the interesting now or never pattern, similar to the finding reported by Wilson and Norton (1989). Specifically, the optimal entry timing for G2 is always now (\( \tau_2^* = 0 \)) when \( m_2 \geq 11.64 \) million, and the solution jumps to never (\( \tau_2^* > 10 \) years) when \( m_2 \leq 11.63 \) million. In addition, we find that a similar threshold exists for the unit contribution margin for G2 (\( \varphi_2 \)), around which a very small change in the parameter value can change the optimal entry timing from now to never.

We would like to point out that there are similarities and differences between our 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, whereas our finding is only valid for the CNS model. It is very interesting to observe that conclusion of the prior study remains valid in this study, although for a completely different business scenario.

As a summary, we show in Table 2 how the now or never rule differs in different studies and under different business scenarios. These results show that it is indeed necessary for firms to adjust their market entry timing strategies based on their underlying business models.
Table 2. Comparison of Findings Concerning Now or Never

<table>
<thead>
<tr>
<th>Scenario I</th>
<th>Scenario II</th>
<th>Scenario III</th>
<th>Scenario IV</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilson and Norton (1989)</td>
<td>now or never</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Mahajan and Muller (1996)</td>
<td>now or at maturity</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Present Study</td>
<td>now or during planning horizon, but not never</td>
<td>now, during planning horizon, or never</td>
<td>now or never</td>
</tr>
</tbody>
</table>

**Conclusion**

In today’s business environment, continuous product improvement in the form of successive releases of product generations is critical for market success. In this research, we conduct a comprehensive analysis of market entry timing for successive product generations under multiple business scenarios. The proposed models can help firms make informed decisions when managing the introduction of successive product generations.

**Acknowledgement**

This research was supported in part by a National Natural Science Foundation of China grant (71428003).

**REFERENCES**


