Why the Generalized Bass Model leads to odd optimal advertising policies

Gila E. Fruchter a, Christophe Van den Bulte b,⁎

a Graduate School of Business Administration, Bar-Ilan University, Ramat-Gan, 59000, Israel
b The Wharton School, University of Pennsylvania, Philadelphia, PA 19106-6340, USA

A R T I C L E   I N F O
Article history:
First received in January 20, 2010 and was under review for 4½ months.
Available online 2 August 2011

Area Editor: Jacob Goldenberg

Keywords:
Diffusion models
New products
Advertising
Price
Optimal control
Variational approach

A B S T R A C T
We show that the optimal advertising strategy under the Generalized Bass Model (GBM) involves beginning at an extremely low level (the lower the better) and then increasing spending throughout the planning period. This strategy remains optimal in the presence of decreasing prices that affect both margins and diffusion speed. We provide a simple explanation for why this happens. We further show that the intuitively appealing patterns of continuous decrease or increase-then-decrease (both with an uptick towards the end) identified in earlier research are also possible as optimal dynamic advertising paths under the GBM structure, but only if the advertising at launch is constrained to be higher than a particular threshold, which we identify. The constraint necessary to generate intuitively appealing strategies lowers overall profits. Therefore, the GBM generates advertising policy recommendations that most marketers would deem odd. This casts doubt on the value of the GBM for normative purposes. Other existing diffusion models are preferred when seeking normative guidance on optimal dynamic advertising policies for new products subject to word of mouth.

1. Introduction

The diffusion of new products through the marketplace is often affected by both marketing communication efforts and social contagion from adopters to potential adopters. Managing the interplay between these two factors has gained renewed attention as novel findings about their relative importance have emerged (e.g., Iyengar, Van den Bulte, & Valente, 2011; Manchanda, Xie, & Youn, 2008; Van den Bulte & Lilien, 2001) and as marketers increasingly try to leverage word of mouth (WOM) and other contagion dynamics to boost the return on their marketing expenditures (e.g., Lehmann & Esteban-Bravo, 2006; Van den Bulte & Joshi, 2007). Even though referral programs and pure-play viral marketing campaigns allow the initial marketing effort to be quite minimal (e.g., De Bruyn & Lilien, 2008; Schmitt, Skiera, & Van den Bulte, 2011; van der Lans & van Bruggen, 2010), such campaigns form the core of very few marketing strategies. For the great majority of products and services, traditional paid-for marketing communications and free word of mouth are complements rather than substitutes (Armelini & Villanueva, 2010).

Over the years, several diffusion models incorporating both social contagion and marketing mix variables have been proposed. Such models are used not only to describe and predict how these two forces jointly drive new product sales but also to provide normative insights into how to manage advertising to maximize profits while taking into account social contagion. The Generalized Bass Model, or GBM, proposed by Bass, Krishnan, and Jain (1994) has been especially popular in both descriptive and normative applications (e.g., Harirhan, Kwon, & Talukdar, 2010; Krishnan & Jain, 2006; Vakratsas & Kolسارıcı, 2008). In a recent contribution, Krishnan and Jain (2006) have shown that in the GBM, the optimal evolution of advertising expenditures after launch, 𝑎(𝑡) with 𝑡 = 0, is highly dependent on the initial level of advertising, 𝑎(0). This immediately raises the managerial question of what level of initial advertising maximizes the profit stream over the entire diffusion process, a key issue that Krishnan and Jain did not resolve.

Answering the call by Bass, Jain, and Krishnan (2000, p. 119) to find optimization approaches for the GBM that do not require fixing the initial levels of decision variables, we revisit the question of optimal dynamic advertising when a new product diffuses as specified by the GBM. Using a variational approach, we identify both the optimal initial level and the optimal subsequent trajectory of advertising. Because the amount of marketing support at launch is often deemed critical for market success (e.g., Sinha & Zoltner, 2001), finding its optimal level under the GBM is an important contribution. In addition, we identify conditions for the presence and location of turning points in the trajectory, as called for by Krishnan and Jain (2006), and we incorporate the effects that eroding margins, declining prices, and non-zero salvage values have on optimal advertising spending.

Identifying the optimal advertising strategy in the GBM is a technically difficult challenge because it involves a constraint on the control (decision) variable in the form of a differential equation. Krishnan...
and Jain (2006) handled this challenge by transforming the objective function, turning the constraint on the control variable into a constraint on the state variable. This approach, however, does not allow the identification of the optimal level of marketing support at launch.

Our main result is that the optimal advertising policy under the GBM structure is to start at an extremely low level (the lower the better) and then to increase spending throughout the entire planning period. This result, which holds even when accounting for the presence of decreasing prices affecting both margins and diffusion speed, conflicts with the common intuition of using costly marketing communication mainly to “prime the WOM pump” and reducing marketing expenditures once the WOM is sufficiently strong to attract customers (e.g., Szymczak & Shapiro, 1993). Furthermore, why would one want to keep increasing advertising expenses late in the process when the number of potential adopters still to be acquired keeps decreasing? Our main result also contrasts with prior analytical results derived from other extensions of the original Bass diffusion model (e.g., Dockner & Jørgensen, 1988; Horsky & Simon, 1983; Kalish, 1985) and other growth models (e.g., Sethi, Prasad, & He, 2008) that optimal advertising starts decreasing either immediately after launch or after the number of adoptions has peaked. Finally, the GBM-derived optimal policy is inconsistent with evidence that advertising and sales force elasticities decrease over time (e.g., Albers, Mantrala, & Sridhar, 2010; Lodish et al., 1995; Manchanda et al., 2008; Narayanan & Manchanda, 2009; Osinga, Leeflang, & Wieringa, 2010; Parsons, 1975).

We provide a simple explanation for this discrepancy. The GBM was originally developed not to help managers make better decisions but to explain why the original Bass model tends to fit empirical diffusion trajectories quite well despite the absence of price, advertising, and other demand influencing variables (e.g., Bass & Srivivasan, 2002, p. 301; Bass et al., 1994, pp. 203–204). This purpose requires the GBM to have a closed-form solution that is observationally equivalent to that of the regular Bass model when prices decline exponentially. To achieve this objective, the GBM assumes that new product sales are not influenced by the level of marketing mix variables but by proportional changes in these variables. This assumption, for which a behavioral rationale has never been articulated, becomes quite problematic when the model is combined with a standard profit function, as suggested by Krishnan and Jain (2006). Sales increase with proportional changes in advertising, while costs are linear in the level of advertising. As a result, the model structure allows firms to have their cake (fast diffusion) and eat it too (negligible advertising costs throughout the planning period provided the initial level is close to zero). Intuitively more appealing patterns of continuous decrease or increase-then-decrease (with an increase at the end) are also possible as optimal dynamic advertising paths under the GBM structure – even with constant margins over time – but only if the advertising at launch is constrained to be higher than a particular threshold, which we identify. Note that the constraint necessary to generate intuitively appealing strategies lowers overall profits.

As we discuss next, other normative models imply patterns of optimal advertising that not only have more face validity but also are consistent with prior research suggesting that the effectiveness of marketing expenditures declines over time. We expect that most readers will find these models preferable to the GBM for normative applications.

2. Literature review

2.1. Analytical results using models other than GBM

Obtaining informative analytical results on the optimal level of marketing spending over time for a new product subject to social contagion has proven to be a difficult task. Only a small number of studies have taken up the challenge. Kotowitz and Mathewson (1979) and Horsky and Simon (1983) find that when social contagion is at work, the optimal policy is to advertise heavily at launch and then to consistently reduce the level of spending as the product diffuses. Monahan (1984) finds the same result in a stochastic model, provided that the effectiveness of advertising is not much higher than that of contagion. Dockner, Feichtinger, and Sorger (1989) extend the model of Horsky and Simon with price and again find (for the undiscounted case) that optimal advertising levels decrease over time. Using different model structures incorporating both price and advertising, Kalish (1985), Jedidi, Elashberg, and DeSarbo (1989), Swami and Khainrar (2006), and Sethi et al. (2008) all come to the same conclusion. Horsky and Mate (1988) and Krishnamoorthy, Prasad, and Sethi (2010) obtain monotonic decline as the optimal strategy in the case of a dupoly rather than a monopoly. Nguyen and Shi (2006) obtain the same result in a structural-type empirical analysis of the rivalry between Kodak and Polaroid. The common result is intuitively appealing to marketers: spend heavily at first to get the word-of-mouth snowball rolling and then cut back as social contagion takes over and generates “free” adoptions. It is also consistent with the “launch hard” prescription of Sinha and Zoltners (2001) based on their experience in identifying optimal sales force policies for pharmaceutical products.

Teng and Thompson (1985) and Mesak and Clark (1998) obtain a somewhat more nuanced result where monotonicity depends on advertising effectiveness. For the undiscounted case, they find that the optimal advertising level increases whenever the elasticity of the number of adoptions with regard to cumulative adoptions increases with advertising (i.e., whenever the contagion elasticity is boosted by advertising) and that the optimal advertising level decreases whenever the elasticity decreases with advertising. Because new products notably affected by contagion have bell-shaped adoption curves, in which adoptions first increase and then decrease with the cumulative number of adoptions, the latter results imply optimal advertising policies where spending first increases and then decreases over time. Dockner and Jørgensen (1988, pp. 119–120) similarly find that, generally, “it pays to increase (decrease) advertising over time if [adoptions] increase (decrease) as penetration increases.” Teng and Thompson (1983) present a model where, depending on how advertising boosts the effectiveness of contagion, the optimal advertising policy is identified to be of one of three types: (i) zero advertising throughout, (ii) advertising at first followed by zero advertising, or (iii) zero advertising early on, followed by maximum advertising at an intermediate stage, then followed by zero advertising later in the diffusion cycle. Thompson and Teng (1984) obtain similar results when the model is extended with price as a decision variable. This set of increase-then-decrease results is somewhat less intuitive than the continuous decrease results discussed earlier. However, the recommendations are certainly not devoid of intuitive appeal from an optimization point of view because they imply that firms should advertise the most when contagion and advertising have the greatest joint effect on generating new adoptions.

Comparing eighteen different model specifications, Mesak and Clark (1998) conclude that the optimal dynamic advertising paths depend on how exactly one incorporates the advertising and contagion effects into the diffusion model. As a result, researchers and managers may prefer to conduct their analysis using a model that is empirically well validated (e.g., Dockner & Jørgensen, 1988; Little, 1998). Monahan (1984) shows that optimal advertising decreases once half of the market has adopted the new product.

1. Our literature review focuses on monopoly results and does not consider models without contagion or models where advertising is taken to essentially be a contagion phenomenon (e.g., Feinberg, 2001; Sethi, 1979). We do not consider studies focusing on allocation across multiple segments or markets (e.g., Lehmann & Esteban-Bravo, 2006).

2. Without the proviso, Monahan (1984) shows that optimal advertising decreases once half of the market has adopted the new product.
Because of this last consideration, Krishnan and Jain (2006) have advocated using the GBM for normative analyses.

2.2. Analytical results using the GBM

Assuming that new product diffusion follows the GBM structure and that the initial level of advertising is predetermined and not under the control of the decision maker, Krishnan and Jain (2006) reached the following conclusions. (1) If the initial expenditures are low, then the optimal policy is to increase advertising throughout the entire planning horizon. (2) If the initial expenditures are high, then the optimal policy is to first decrease but later increase advertising. In short, the recommendation for how advertising should evolve over time within the GBM hinges critically on the initial level of advertising. This immediately implies that to truly characterize the optimal dynamic advertising policy $a(t)$, one should also identify the optimal level of advertising support at launch $a(0)$. In addition, marketers operating under constraints making it impossible for them to choose $a(0)$ freely would still benefit from more precise insights into the critical level of initial spending $a(0)$ determining which advertising trajectory is more profitable (i.e., what is “high” or “low”). The issues of optimal and critical initial advertising levels, however, have remained open questions.

2.3. Empirical research

Empirical research strongly suggests that the optimal strategy in real markets is likely to involve decreasing advertising over time, especially late in the diffusion process. Research on packaged goods shows that advertising is more effective shortly after launch (Lodish et al., 1998), which suggests that it is profitable to advertise heavily early rather than late. Direct evidence to this effect is provided by Mesak and Clark (1998), who found by fitting a large number of different Bass-type diffusion model specifications that advertising affected only the coefficient of innovation $p$ and that the advertising elasticity decreased over time. Additional support for heavier marketing effort early rather than late is provided by evidence that marketing communication is especially effective early on (Albers et al., 2010; Lehmann & Mayzlin, 2007; Narayanan, Manchanda, & Chintagunta, 2005; Narayanan & Manchanda, 2009; Osinga et al., 2010; Parsons, 1975). The most direct evidence of the superiority of providing especially strong marketing support at launch arguably comes from work by Sinha and Zoltners (2001) in the area of sales force allocation. Fig. 1 documents such a pattern, where a pharmaceutical company marketing a new drug reduced the sales force efforts targeted towards potential adopters over time, even on a per-capita basis accounting for the decreasing number of physicians who had yet to adopt as the product gained marketplace acceptance (Iyengar et al., 2011).

2.4. Conclusion from prior research

In short, the great majority of research suggests that the optimal dynamic advertising strategy consists of either a monotonic decrease from the time of launch onwards or an initial increase followed by a monotonic decrease. Only the GBM leads to a recommendation of increasing advertising throughout the entire planning horizon, but does so only if the initial expenditures are low. This raises the question of whether high or low initial advertising is optimal under the GBM.

3. Model and assumptions

3.1. The Generalized Bass Model

We consider the diffusion of a new product in a monopolistic market. Marketing instruments other than price and advertising are held constant, though what we refer to as “advertising” can include any demand-inducing marketing expenditure, including sales force support, direct mail, and so on. Focusing on the diffusion process, we do not consider repeat purchases. The diffusion ceiling, the ultimate number of adopters, is constant and normalized to 1. We assume that
the diffusion process evolves as described by the GBM (Bass et al., 1994):
\[ \dot{F}(t) = [p + qF(t)](1 - F(t))x(t), \quad F(0) = 0. \] (1a)

In Eq. (1a), \( F(t) \) is the proportion of ultimate adopters that have already adopted the product at time \( t \), \( p \) and \( q \) are the innovation and imitation coefficients (non-negative and constant) as in the traditional mixed-influence or Bass model, and \( x(t) \) represents the current impact of dynamic marketing variables. The latter is specified as
\[ x(t) = 1 + \alpha \frac{\dot{p}(t)}{p(t)} + \beta \frac{\dot{q}(t)}{q(t)}, \quad \alpha \leq 0, \beta > 0, \] (1b)
where \( p(t) \) is the current price, \( a(t) \) is the current level of advertising expenditure, \( \dot{p}(t) \) and \( \dot{q}(t) \) are the proportional change in price and advertising, respectively, and the coefficients \( \alpha \) and \( \beta \) reflect how sensitive the diffusion process is to price and advertising, respectively. Note that the speed of adoption at a particular point in time is affected not by the level of price or advertising but by the proportional change in those marketing mix variables at that time. Previous changes in price and advertising are also significant because they carry through into \( \dot{F}(t) \) via \( F(t) \). The model reduces to the Bass model when price and advertising have no effect, when they are constant, or when they increase or decrease at a constant proportional rate.

3.2. Assumptions

3.2.1. Price

We focus on optimal advertising and assume that price is not a decision variable. Specifically, we assume that price is constant or decreases exponentially, the latter being consistent with empirical evidence for many new technologies and durables (Bass et al., 1994). Setting the initial price \( p(0) \) equal to \( v \), we specify the price as
\[ p(t) = ve^{-\gamma t}, \quad \gamma \geq 0. \] (2a)

As a result, the current impact of all dynamic marketing variables in the GBM equals
\[ x(t) = 1 - \alpha \gamma + \beta \frac{\dot{a}(t)}{a(t)} = k + \beta \frac{\dot{a}(t)}{a(t)}, \quad k = 1 - \alpha \gamma \geq 1. \] (2b)

3.2.2. Advertising

We denote the rate of change in advertising at time \( t \) by \( R(t) \) such that
\[ \dot{a}(t) = R(t)a(t), \] (3a)
and thus,
\[ x(t) = k + \beta R(t). \] (3b)

Like Krishnan and Jain (2006), we assume that \( R(t) \) is bounded from below and above:
\[ R \leq R(t) \leq \tilde{R}, \] (4)
where \( R > 0 \) and \( \tilde{R} < 0 \), and both are fixed. We denote the admissible set for \( R(t) \) by \( \Omega_{R(t)} = [R, \tilde{R}], t \in [0, T] \). In many applications, \( \Omega_{R(t)} \) is determined by physical or economic constraints (Sethi & Thompson, 2000). For advertising, the constraints can result from limitations on how easily the firm can raise money or from inertia in reallocating marketing resources across divisions and product lines.

We also assume that the advertising \( a(t) \) is positive. Because Eq. (3a) implies \( a(t) = a(0) \exp \left( \int_{0}^{t} R(\tau) d\tau \right) \), it is sufficient to impose a constraint on the initial level of advertising \( a(0) \):
\[ a(0) \geq \tilde{a}_0 > 0, \] (5)
where \( \tilde{a}_0 \) is the lowest feasible positive initial advertising.\(^4\) We denote the admissible set for \( a(0) \) by \( \Omega_{a(0)} = [\tilde{a}_0, \infty) \). The lower bound can be minute, constrained only by the expenditure for a minor communication effort through a cheap medium.

3.2.3. Profits

With the sales process given by the GBM and assuming a unit cost of zero, the profit \( \pi(t) \) at each instant \( t \) is given by
\[ \pi(t) = w(t) \dot{F}(t) - a(t) = w(t)[p + qF(t)](1 - F(t))[k + \beta R(t)] - a(t). \] (6a)
where
\[ w(t) = p(t) = ve^{-\gamma t}, \quad \gamma \geq 0 \] (6b)
is the contribution margin per unit sold (i.e., per adoption), scaled such that the market size is 1. Following Krishnan, Bass, and Jain (1999) and Krishnan and Jain (2006), we assume zero marginal costs to avoid needless complexity.

3.2.4. Objective function

The firm’s problem is choosing the advertising expenditure over time, \( a(t) \), to maximize the discounted profits over a time horizon \([0, T] \), where \( T \) may (but need not) go to infinity. Because the rate of change in advertising, \( R(t) \), governs the decision variable \( a(t) \) for \( t \leq 0 \), the firm chooses the initial advertising expenditure \( a(0) \) and subsequent change rates \( R(t) \) to maximize the discounted profits over a time horizon \([0, T] \). Because we are concerned with the flow of adoptions and profits over time, it is important to account for the time value of money. We denote the discount rate by \( r > 0 \) and assume that the firm wants to maximize the discounted profits over a planning horizon \([0, T] \). To increase the generality and realism of the analysis, we allow the firm to set its advertising policy for a finite period rather than forcing the planning horizon to be infinitely long. Additionally, we do not require the firm to myopically ignore the value of sales to be realized beyond its planning horizon. The total discounted profits, \( \Pi \), can be expressed as
\[ \Pi = \int_{0}^{T} [w(t)[p + qF(t)](1 - F(t))[k + \beta R(t)] - a(t)]e^{-\gamma t} dt + S(F(T), T). \] (7)

The integral term in \( \Pi \) represents the discounted profits from the adoptions during the period \([0, T] \). The second term represents the salvage value and can be zero or positive. We use three specifications for the salvage value:

Case 1.
\[ S(F(T), T) = 0, \] (7a)

\(^4\) It is clear from Eq. (1b) that for the GBM to hold and not lead to instantaneous full diffusion as soon as the firm advertises, one must have \( a(0) > 0 \). No other constraints on \( a(0) \) are necessary for our analysis, nor are they necessary to obtain the standard closed-form solution of \( F(t) \) and \( \dot{F}(t) \) in the GBM, which assume that \( F(0) = 0 \) and \( X(0) = 0 \), where \( X(t) = kt + \int_{0}^{t} \dot{R}(\tau) d\tau \). Note that \( X(0) = 0 \) holds for any value of \( a(0) \).
Case 2.

\[
S(F(T), T) = \frac{1 - F(T)}{F_B(T)} \int_{0}^{\infty} w(t) \hat{F}_b(t)e^{-rt} dt = v_{rg}(1 - F(T))e^{-rt},
\]

\[
v_{rg} = \frac{1}{1 - F_B(T)} \int_{0}^{\infty} w(t) \hat{F}_b(t)e^{-rt} dt < w(T), \quad (7b)
\]

\[
F_B(t) = \frac{1 - e^{-p + qt}}{1 + (q/p)e^{-p + qt}}.
\]

Case 3.

\[
S(F(T), T) = \left( \int_{0}^{\infty} w(t) \hat{F}_b(t)e^{-rt} dt \right) F(T)e^{-rt} = v_{rg}F(T)e^{-rt},
\]

\[
v_{rg} = 0.8 \int_{0}^{\infty} w(t) \hat{F}_b(t)e^{-rt} dt.
\]

The first salvage value is simply zero and assumes that the firm myopically ignores what happens after the planning horizon. The second salvage value is the discounted profit from the remaining adoptions to occur at the end of the advertising period and hence decreases with \( F(T) \). Specifically, we use the value of the adoptions of the focal product taking place beyond \( T \) when advertising is stopped, and the diffusion proceeds according to the traditional Bass model. The integral \( \int_{0}^{\infty} w(t) \hat{F}_b(t)e^{-rt} dt \) is the discounted value of the adoptions occurring after \( T \) without advertising if the process had been without advertising between \( 0 \) and \( T \) as well. This integral is renormalized by the factor \( \frac{1 - F(T)}{1 - F_B(T)} \) to account for the fact that the fraction of adoptions occurring after time \( T \), given the advertising support between \( 0 \) and \( T \), is \( 1 - F(T) \) rather than \( 1 - F_B(T) \). For example, if advertising increases (decreases) throughout the planning period, then \( F(T) \) in the GBM is larger (smaller) than \( F_B(T) \) in the traditional Bass model. The third salvage value is the discounted profit if adopters of the focal product form the market potential for a next-generation product launched at the end of advertising period and hence increases with \( F(T) \). Specifically, we assume that 80% of those who have adopted the focal product by time \( T \) constitute the pool of adopters of a next generation product launched at time \( T \) and diffusing according to a classic Bass model. Setting the fraction at 80% is arbitrary.

3.3. The optimal control problem

The firm’s goal is to maximize \( \Pi \) while recognizing that it can influence the adoption rate \( \hat{F}(t) \) by judiciously setting its advertising level. Given the constraint Eqs. (3a), (4) and (5), the optimal policy must be a solution to the following problem:

\[
\begin{align*}
\max_{a(0:0:T)} & \int_{0}^{T} \left[ w(t)[p + qF(t)][1 - F(t)] + k + \beta(R(T) - a(t))e^{-rt} dt + S(F(T), T) \right] \\
\text{s.t.} & \text{Eqs. (3a),(4),(5), and where } F(t) = [p + qF(t)][1 - F(t)][k + \beta(R(t)), \quad F(0) = 0
\end{align*}
\]

(8)

The technical challenge in this problem is that it involves a constraint on one of the control variables in the form of a differential equation (Eq. 3a). The optimization problem in Eq. (8) generalizes the one defined and solved in Krishnan and Jain (2006) in two respects:

(i) It involves the two decision variables of key managerial interest, \( a(0) \) and \( R(t) \), rather than only post-launch decisions \( R(t) \);

(ii) It allows for a non-zero salvage value.

The first generalization leads to very different theoretical and substantive insights than those obtained by Krishnan and Jain.5

4. Results

We now present our analytical results for the optimal advertising policy under the GBM structure. The analytic procedure uses a variational approach (e.g., Bryson & Ho, 1975), and the proofs are reported in the Appendix.

4.1. Main result: start extremely low and keep increasing

We prove that (i) the optimal policy always involves advertising as little as possible at launch, (ii) lowering the minimum allowed initial advertising always increases profits, and (iii) if the advertising at launch is allowed to be below a specific threshold, which we identify, the optimal policy is to always increase advertising. These analytical results hold regardless of whether prices and margins are constant or decrease, and regardless of the choice of salvage value. In other words, the optimal GBM policy is to spend extremely little at first (the lower the better) and then increase spending throughout the planning period.

This policy runs counter to common sense supported by the analytical studies reviewed previously. The latter suggest either monotonic decreases based on a “get the snowball rolling and then free-ride word of mouth” logic or increase-then-decrease trajectories based on a “spending should follow the advertising curve” logic. The optimal GBM policy also runs counter to empirical findings favoring high initial spending (e.g., Lodish et al., 1995; Sinha & Zoltner, 2001). Our key result, however, becomes intuitively obvious once one reflects on the profit Eq. (6a) in which sales are boosted by proportional changes in advertising, \( \beta(R(t) = \frac{\alpha(b - \alpha)}{\text{cost}} \), but costs are linear in the level of advertising \( a(t) \). Maximum profits will be achieved by having a very high growth rate in advertising \( R(t) \) but a very low actual level of advertising \( a(t) \). Both can be attained simultaneously by choosing an extremely low initial level of advertising \( a(0) \). This suggests that starting advertising at the lowest level possible and then increasing it steadily is a very attractive strategy under the GBM structure (assuming a positive discount rate) because it quickly moves adoptions from the future to the present at little or no cost. This allows marketers to “have their cake” in the form of accelerated adoption and “eat it too” by avoiding sizable advertising costs.

The numerical analyses reported in Table 1 illustrate how the results we have just reported are robust to the presence of decreasing prices and margins and to the choice of salvage value. Note that decreasing prices (\( \gamma > 0 \)) can have two effects: they always depress the gross margins, and they affect the speed of diffusion if \( \alpha > 0 \). To distinguish between these two effects, we investigate three types of cases: constant prices and margins (\( \gamma = 0 \)), decreasing prices and margins without boost in diffusion speed (\( \gamma > 0, \alpha = 0 \)), and decreasing prices and margins with boost in diffusion speed (\( \gamma > 0, \alpha > 0 \)).

As Table 1 shows, margin erosion obviously affects the NPV, and price declines paired with a non-zero price coefficient in the diffusion equation obviously boost the NPV (net of margin effects). What kind of salvage value one assumes, if any, also affects the NPV. However,

---

5 Krishnan et al. (1999) considered optimal pricing policies under the GBM, considering both initial price \( P(0) \) and subsequent price evolution \( R(t) \). Their work, however, differs from ours in four important respects. First, they did not work with the traditionally, empirically validated GBM specification but rather with a variation of the model that has not been empirically supported. Second, they did not solve for the optimal initial price and subsequent evolution analytically but had “to necessarily resort to numerical analyses to arrive at the optimal pair \( (P(0), R(t)) \)” (p. 1658). Third, they did not provide analytical expressions to identify the switching times at which the optimal pricing path changes direction (e.g., from increase to decrease). Fourth, price and advertising enter in the profit function quite differently: price levels affect gross margins but not fixed costs, whereas the opposite holds for advertising.
the optimal strategy of starting advertising as low as possible and increasing it throughout the planning period holds across the different types of salvage values and regardless of whether price erosion affects margins or diffusion speed. Spending as little as possible at launch and then increasing spending continually over time remains the optimal GBM strategy even when the planning horizon extends to infinity, provided that the discount rate, \( r \), is higher than the upper bound of the growth rate in spending, \( \overline{R} \). Because we assume the latter to be positive, the optimal GBM strategy implies infinite advertising spending over the entire diffusion process when the planning horizon extends to infinity and \( r=\overline{R} \). When the planning horizon extends to infinity but \( r<\overline{R} \), in contrast, the optimal policy involves at least one decrease in advertising.

4.2. Alternative, constrained optimal advertising policies

We also prove the following: (i) if the lowest level of advertising allowed at launch is too high for a monotonic increase to be optimal, then the optimal advertising path includes at least one cut in advertising (making decrease–increase or increase–decrease–increase patterns optimal), and (ii) the switching times can be identified analytically as functions of the diffusion parameters, advertising and price effectiveness, advertising growth/reduction rates and price erosion rates, and discount rate. These analytical results hold regardless of whether prices and margins are constant or decrease, and regardless of the choice of salvage value.

In other words, optimal GBM policies that are consistent with common sense are possible, but only if the lowest level of advertising allowed at launch is constrained to be higher than a particular value (identified in Propositions 2–4 in the Appendix). The constraint necessary to generate intuitively appealing strategies lowers overall profits. To illustrate, consider a situation where \( p=0.3, q=0.3, \beta=0.5, \alpha=0, r=0, \overline{R}=1, \overline{R}=-1, w(t)=10^{3}, \) and \( T=10 \) such that \( F_{p_{0}}(T)=70\% \). If the minimum admissible advertising at launch is very small, \( a_{0}=0.1 \), then the optimal policy is to set \( a^{*}(t)=1e^{1t} \), throughout the entire planning period from \( t=0 \) to \( T \), which generates a net present value of 317,367 using a salvaging value based on remaining adoptions and 284,295 using one based on next-generation sales (Table 1). When the minimum admissible advertising level at launch is increased from \( 0.1 \) to \( 20,000 \) but all other parameters are kept the same, then the optimal policy is to start at \( 20,000 \), to decrease from \( t=0 \) to \( 9.78 \) when \( a(t)=20,000e^{-0.978t}=7521 \), and to then start increasing again for the very short duration until \( t=10 \) to 7689. Being forced to start at a level that is too high cuts the net present value by 37\%, from 317,367 to 201,139, when the salvage value

is based on remaining adoptions. For the case where salvage value is based on next-generation sales, the optimal strategy is to cut advertising until \( t=9.42 \). The NPV decreases by 29\%, from 284,295 to 201,399. The detrimental impact on the NPV is driven both by the advertising costs that are too high and by the fact that cutting advertising to avoid such unjustifiably high costs pushes adoptions from the present into the future.

The numerical analyses reported in Table 2 illustrate how the results we have just reported are robust to the presence of decreasing prices and margins and to the choice of salvage value. Margin erosion obviously affects the NPV, and price declines paired with a non-zero price coefficient in the diffusion equation obviously boost the NPV (net of margin effects), but the optimal strategy of starting advertising as low as possible and consistently decreasing it until very late in the planning period holds across the different types of salvage value and regardless of whether price erosion affects margins or diffusion speed.

In our numerical analyses, we also identified situations with two switching times and an increase–then-decrease advertising pattern (with an uptick again at the end). We did not find any situations with more than two switching times. One example of two switching times is the situation in which we take the same scenario as above but raise \( \beta \) from 0.05 to 0.1 and set the minimum admissible launch advertising at 5000. The optimal policy is then to start advertising at 5000, increase spending for a while, then decrease spending through most of the planning period with a minor uptick very late in the process. Table 3 shows that this pattern is robust to the choice of salvage value and to price declines affecting both margins and diffusion speed.

Margin erosion obviously leads one to switch from increasing to decreasing advertising sooner when margins are constant \( (t_{2} \) is lower when \( \gamma=0 \) than when \( \gamma=0.1 \)) and to continue decreasing it longer \( (t_{1} \) is higher). Furthermore, when the salvage value stems from the remaining adoptions of the focal product, obtaining a penetration rate at the end of the planning period, \( F(T) \), less than 1 is not purely lost profit. Marketers who take this into account \( (Case \ 2) \) will therefore advertise less aggressively than myopic marketers \( (Case \ 1) \), leading the former to switch from increasing to decreasing advertising sooner \( (t_{2} \) is lower in Case 2 than Case 1) and to continue decreasing it longer \( (t_{1} \) is higher). That pattern reverses when the salvage value stems from sales of a next-generation product to adopters of the focal product. The firm then has an incentive to advertise more than the myopic firm to boost the installed base. As Table 3 shows, the installed base effect leads the firm to switch from increasing to decreasing advertising later than in the absence of an installed base generating subsequent profits \( (t_{1} \) is higher in Case 3 than Case 1) and to decrease it for a shorter time \( (t_{1} \) is lower).

Finally, comparing the corresponding cells across Tables 1–3 illustrates how imposing higher minimum initial advertising levels depresses the profitability (NPV) of the optimal advertising strategy.
under the GBM. In short, while both the “continuous decrease” policy (e.g., Horsky & Simon, 1983) and the “increase then decrease” policy (e.g., Teng & Thompson, 1985) can be recommended under the GBM, even with constant margins, such intuitively appealing strategies result only if the initial advertising is constrained to be sub-optimally high.

4.3. Upper bounds on initial advertising that avoid advertising cuts

As we prove in the Appendix and illustrate with numerical examples in the preceding sections, the truly optimal strategy under the GBM is to start advertising at an extremely modest level and increase the expenditures at a constant rate throughout the planning period. However, if the initial level is not allowed to be sufficiently low, the optimal policy will involve advertising cuts. This naturally raises the question of how high the initial level can be before the firm is forced to cut its expenditures at some point, either immediately after launch or later. As Krishnan and Jain (2006) note, this is an important question, and our analysis allows us to answer it by identifying how high $g_0$ can be before the inequality in Proposition 1, condition (A2S) in the Appendix, is violated. Transforming this condition shows that for sustained increases in advertising throughout the entire planning period to be optimal, (i) the ratio between the effectiveness of the rate of change in advertising and the current multiplicative impact of all dynamic marketing variables must, at any point under the optimal advertising policy, be higher or equal to that (ii) the optimal advertising-to-sales ratio weighted by a factor scaling the cost of increasing advertising against the profit impact from increasing current adoptions.

Numerical analysis provides additional insight. Table 3 reports the upper bounds on the initial advertising that allow for subsequent sustained advertising increases for several values of the advertising effectiveness $\beta$, the advertising growth rate $R$ (the rate of advertising decrease $R$ is irrelevant for this question), and the discount rate $r$ when $p = .03, q = .30, \alpha = \gamma = 0$, and $\beta = 0.1, \gamma = 0.1, T = 10$, and the firm takes into account the salvage value based on remaining adoptions (Case 2).

Table 3

<table>
<thead>
<tr>
<th>Salvage value</th>
<th>Case 1</th>
<th>Case 2</th>
<th>Case 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) Optimal advertising policy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma = 0, \alpha = 0$</td>
<td>$t_2 = 4.39, t_1 = 8.36$</td>
<td>$t_2 = 3.64, t_1 = 8.99$</td>
<td>$t_2 = 4.72, t_1 = 8.22$</td>
</tr>
<tr>
<td>$\gamma = 0.1, \alpha = 0$</td>
<td>$t_2 = 2.97, t_1 = 9.14$</td>
<td>$t_2 = 2.79, t_1 = 9.49$</td>
<td>$t_2 = 3.13, t_1 = 8.90$</td>
</tr>
<tr>
<td>$\gamma = 0.1, \alpha = -0.2$</td>
<td>$t_2 = 2.99, t_1 = 9.15$</td>
<td>$t_2 = 2.81, t_1 = 9.49$</td>
<td>$t_2 = 3.15, t_1 = 8.90$</td>
</tr>
<tr>
<td>(b) Resulting NPV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\gamma = 0, \alpha = 0$</td>
<td>240,496</td>
<td>263,744</td>
<td>256,073</td>
</tr>
<tr>
<td>$\gamma = 0.1, \alpha = 0$</td>
<td>157,517</td>
<td>161,156</td>
<td>163,689</td>
</tr>
<tr>
<td>$\gamma = 0.1, \alpha = -0.2$</td>
<td>161,030</td>
<td>164,547</td>
<td>167,163</td>
</tr>
</tbody>
</table>

In all cases, the optimal policy is to set $\alpha'(0) = g_0 = 5000$, increase advertising from $t = 0$ until $t_2$, then decrease advertising until $t_1$, and finally increase it again.

However, if the initial level is not allowed to be sufficiently low, the optimal policy will involve advertising cuts. This naturally raises the question of how high the initial level can be before the firm is forced to cut its expenditures at some point, either immediately after launch or later. As Krishnan and Jain (2006) note, this is an important question, and our analysis allows us to answer it by identifying how high $g_0$ can be before the inequality in Proposition 1, condition (A2S) in the Appendix, is violated. Transforming this condition shows that for sustained increases in advertising throughout the entire planning period to be optimal, (i) the ratio between the effectiveness of the rate of change in advertising and the current multiplicative impact of all dynamic marketing variables must, at any point under the optimal advertising policy, be higher or equal to that (ii) the optimal advertising-to-sales ratio weighted by a factor scaling the cost of increasing advertising against the profit impact from increasing current adoptions.

Numerical analysis provides additional insight. Table 4 reports the upper bounds on the initial advertising that allow for subsequent sustained advertising increases for several values of the advertising effectiveness $\beta$, the advertising growth rate $R$ (the rate of advertising decrease $R$ is irrelevant for this question), and the discount rate $r$ when $p = .03, q = .30, \alpha = \gamma = 0$, and $\beta = 0.1, \gamma = 0.1, T = 10$, and the firm takes into account the salvage value based on remaining adoptions (Case 2).

Table 4

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$\alpha$</th>
<th>$R$</th>
<th>$\beta$</th>
<th>$r$</th>
<th>$\text{NPV}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.05</td>
<td>0.05</td>
<td>0.1</td>
<td>0.2</td>
<td>5180</td>
</tr>
<tr>
<td>0.00</td>
<td>0.05</td>
<td>0.05</td>
<td>0.1</td>
<td>0.3</td>
<td>5598</td>
</tr>
<tr>
<td>0.00</td>
<td>0.05</td>
<td>0.05</td>
<td>0.1</td>
<td>0.5</td>
<td>5840</td>
</tr>
</tbody>
</table>

5. Conclusion

We identify the optimal dynamic advertising strategy for a new product that diffuses according to the Generalized Bass Model (GBM). Prior research by Krishnan and Jain (2006) dealt with this challenge only partially, leaving the problem of initial spending unsolved even though prior experience shows that marketing support at launch is of critical importance (e.g., Sinha & Zoltners, 2001). Using a variational approach, we solve the problem of finding the optimal level of advertising spending at launch and the optimal subsequent evolution, as called for by Bass et al. (2000). We also derive conditions for the presence and location of switching times (or turning points) in the optimal trajectory, as called for by Krishnan and Jain (2006), and properly account for the salvage value at the end of the firm’s planning period.

We have seven new insights. First, the optimal policy always involves advertising as little as possible at launch, regardless of how advertising levels evolve later. Second, if the advertising at launch is allowed to be below a specific threshold, which we identify, the optimal policy is to always increase advertising. Third, numerical analyses indicate that this threshold, the maximum value of advertising at launch for a subsequent monotonic increase to be the optimal policy, increases with the effectiveness of advertising, $\beta$, and the discount rate, $r$, and decreases with the maximum growth rate of advertising, $R$. Fourth, if the advertising at launch is too high for a monotonic increase to be optimal, then the optimal advertising path is either decrease–increase or increase–decrease–increase, even when margins are constant over time. Fifth, conditions for the switching times can be identified analytically as functions of the diffusion parameters, advertising and price effectiveness, advertising growth/reduction and price erosion rates, and discount rate. Sixth, decreasing prices, affecting both margins and diffusion speed, do not affect the main result that the best strategy is to set initial advertising extremely low and then increase advertising throughout. Finally, ignoring the salvage value stemming from remaining adoptions leads to overinvestment rather than underinvestment in advertising, which is the opposite pattern of applications where salvage value stems from next-generation or repeat purchases by acquired customers rather than from adoptions from yet to be acquired customers. Each of these seven results goes beyond the earlier
contribution by Krishnan and Jain (2006) that the optimal trajectory \( a^* (t) \) for \( t > 0 \) is highly dependent on \( a(0) \).

The most important result is that the optimal GBM policy is to spend extremely little at first and then increase spending throughout the planning period. The growing number of low-cost marketing technologies and media available for triggering social contagion (tweets, emails, videos, etc.) makes very low launch budgets increasingly feasible. Our results prove that, under such conditions, the GBM implies that optimal policies will increasingly exhibit a very low launch budget followed by monotonic increases in spending, even though the number of prior adopters generating free word of mouth keeps growing. In contrast, we do not believe that the availability of new marketing technologies allowing for small initial investments fundamentally changes the normative “start high then decrease” or “increase then decrease” policies derived from the many other models discussed in our literature review.

The optimal GBM policy, spending extremely little at first and then increasing spending throughout the planning period, runs counter to common sense and marketing intuition as well as many prior analytical studies, all of which imply either monotonic decreases based on a “get the snowball rolling and then free-ride word-of-mouth” logic or increase-then-decrease trajectories based on a “spending should follow the adoption curve” logic. The optimal GBM policy also runs counter to empirical findings favoring initial spending (e.g., Albers et al., 2010; Lodish et al., 1995; Narayanan & Manchanda, 2009; Sinha & Zoltners, 2001). Our key result, however, becomes quite intuitive once one takes into account the odd structure of the GBM where the firm benefits from proportional increases in spending but pays only for levels of spending, allowing the firm to “have its cake” in the form of accelerated adoption and “eat it too” by avoiding sizable advertising costs. That a critically important yet dubious policy recommendation stems from such an odd model feature casts doubts on the value of the GBM for normative purposes.

Science advances by a process of variation, selection and retention, and progress comes not only from developing new models or theories but also from weeding out unfit models or theories, as many observers have noted (e.g., Hull, 1988; Popper, 1959). One response to our work might be that diffusion researchers and marketing analysts consider the GBM to have been “selected out” for normative purposes and turn to other model structures in normative applications. An alternative response might be to adapt the model to neutralize its “have your cake and eat it too” characteristic. One such modification might be to make sales an increasing function of the initial level of advertising (ceteris paribus), as Krishnan et al. (1999) did for price. Such modifications to the GBM, however, have not been empirically validated. Until that happens, researchers seeking to develop normative analytical insights might be advised to use other diffusion models meeting two important criteria: (i) having advertising levels rather than only changes in levels affecting sales, and (ii) having been documented to have at least some descriptive validity (e.g., Horsky & Simon, 1983). The first requirement, as we have shown, is important for the model to be robust in its normative applications—both in the sense of Little’s (1970) robustness criterion of avoiding absurd implications and of Wimsatt’s (1981, 2007) robustness criterion of avoiding extreme sensitivity of policy and other conclusions to unique model assumptions. The second criterion, documented descriptive validity, is less critical for theoretically orientated normative modeling, but it may be a useful “tie breaker” when choosing among many possible model specifications that are equally attractive otherwise.

When trying to identify optimal policies, managers and marketing scientists have to live with the consequences of the assumptions they make. As we have shown, the GBM’s specification of how advertising affects new product sales leads to very odd policy recommendations. Other existing diffusion models are preferred when seeking normative guidance on optimal dynamic advertising policies for new products subject to word of mouth.

**Acknowledgments**

Much of this research was completed while Gila E. Frucher was a visitor at The Wharton School of the University of Pennsylvania, on sabbatical leave from the Graduate School of Business Administration, Bar-Ilan University. We benefited from comments by Fred Feinberg, Ernan Haruvy, Trichy Krishnan, Gary Lilien, Renana Peres, Ashutosh Prasad, and the IJRM review team.

### Appendix I. Necessary conditions for optimal advertising

We first formulate the necessary conditions for the solution to the optimization problem stated in Eq. (8). These conditions allow us to derive important propositions regarding the optimal advertising policy over time.

Let \( \lambda_a(t) e^{-\lambda_a(t)} \), \( \lambda_a(t) e^{-\mu_a(t)} \), \( \mu_a(t) e^{-\mu_a(t)} \), \( \mu_a(t) e^{-\mu_a(t)} \) be the Lagrange multipliers or shadow prices associated with the differential equation constraints, and let \( \mu_a(t) e^{-\mu_a(t)} \), \( \mu_a(t) e^{-\mu_a(t)} \), \( \mu_a(t) e^{-\mu_a(t)} \), \( \mu_a(t) e^{-\mu_a(t)} \) be the Lagrange multipliers associated with the inequality constraints on \( a(0) \) and \( R(t) \), respectively. In our context, \( \lambda_a = \lambda_a(t) \) and \( \lambda_a = \lambda_a(t) \) capture the profit impact from increasing the proportion of ultimate adopters that have already adopted the product and increasing the rate of advertising change, respectively, by one unit at time \( t \). The shadow prices \( \mu_a(t) = \mu_a(t) \), \( \mu_a(t) = \mu_a(t) \), and \( \mu_a(t) = \mu_a(t) \) will be zero if \( (a(0) - \frac{R(t)}{R}) \), \( (R - R) \), and \( (R + R) \), respectively, are positive and will be non-negative otherwise.

Mathematically,

\[
\mu_a(t) \left\{ \begin{array}{ll}
\geq 0 & \text{if } a(t) = a(0) = 0,
\mu_a(t) \leq 0 & \text{if } R(t) > \frac{R}{R},
\mu_a(t) \geq 0 & \text{if } R(t) = \frac{R}{R},
\mu_a(t) \leq 0 & \text{if } R(t) = \frac{R}{R}.
\end{array} \right.
(A1)
\]

Using a variational approach (e.g., Bryson & Ho, 1975), we obtain the following result.

**Theorem 1.** Consider the profit maximization (Eq. (8)). The necessary conditions for the initial advertising \( a(0) \) and the rate of change in advertising at time \( t, R = R(t) \), to be optimal are as follows:

(i) There exists \( \lambda_a \) which satisfies the following conditions:

\[
\lambda_a(T) = \frac{d}{dt} \Delta(t)(1 - \sum_{k} p_k \Delta(t)) / \Delta(T), \quad \lambda_a(T) = \Delta(T) / \Delta(T). \tag{A2}
\]

(ii) The optimal policy of initial advertising and rate of change in advertising \((a^*(0), R^*(t))\) satisfies the following conditions:

\[
a(0) = a_0, a_0 \in \Omega(0)
\]

and

\[
R(t) = \begin{cases}
\mathbb{R} & \text{if } (R^*(0) - R) \leq 0, \\
\mathbb{R} & \text{if } (R^*(0) - R) > 0, \quad \mathbb{R} \in \Omega(0).
\end{cases}
\tag{A4}
\]

where \( a \) is as in Eq. (3a).

Conditions (A3) and (A4) are the necessary conditions for the control parameter \( a(0) \) and control variable \( R(t) \). Note that \( a^*(0) \) always takes the lowest permissible value, whereas \( R^*(t) \) takes only the lowest and highest permissible values (bang-bang behavior). Condition (A2) consists of a differential equation for \( \lambda_a \) and a terminal condition at \( T \). The differential equation indicates the profit impact of increasing the cumulative number of adoptions at each point in time.
The terminal condition indicates the profit impact when the diffusion process stops at $F(T)$ rather than continuing to full penetration.

**Proof of Theorem 1. Derivation of the first-order necessary conditions**

Adjoining the constraint equations in Eq. (8) with the multiplier functions $\lambda_\epsilon(t)e^{-\eta t}$, $\lambda_\mu(t)e^{-\eta t}$ and $\mu_\beta(t)e^{-\eta t}$, $\mu_\delta(t)e^{-\eta t}$, $\mu_F(t)e^{-\eta t}$, we obtain

$$\Pi = S(F(T), T) + \int_0^T \left[ \left( w + qF \right)(1-F)(k + \beta R) - a + \lambda_\epsilon(Ra - a) 
+ \lambda_\mu[(p + qF)(1-F)(k + \beta R) - \dot{F}] + \mu_\beta(a(0) - g_0)
+ \mu_\delta(R-R) + \mu_F(R-R) \right] e^{-\eta t} dt.$$  \hspace{1cm} (A5)

For convenience, we define a scalar $H$ (the Hamiltonian), as follows:

$$H(F, a(0), R, \lambda_\epsilon, \lambda_\mu, \mu_\beta, \mu_\delta, \mu_F)$$

$$= \left[ (w + \lambda_\epsilon)(p + qF)(1-F)(k + \beta R) - a + \lambda_\mu Ra + \mu_\beta(a(0) - g_0) 
+ \mu_\delta(R-R) + \mu_F(R-R) \right] e^{-\eta t}.$$  \hspace{1cm} (A5a)

Integrating the remaining terms in Eq. (A5) by parts yields

$$\Pi = S(F(T), T) - \lambda_\epsilon(T)e^{-\eta T}a(T) + \lambda_\mu(t)e^{-\eta T}a(0) - \lambda_\mu(T)e^{-\eta T}F(T)$$

$$+ \int_0^T \left[ H + \left( \lambda_\mu - \lambda_\epsilon \right)ae^{-\eta t} + \left( \lambda_F - \lambda_\mu \right)Fe^{-\eta t} \right] dt.$$  \hspace{1cm} (A5b)

Now consider the first-order variation in $\Pi$ due to variations in the control variable $R(t)$ and the control parameter $a(0)$ for fixed times $t = 0$ and $T$:

$$\delta \Pi = -\lambda_\epsilon(T)e^{-\eta T}a(T) + \lambda_\mu(0)\delta a(0)$$

$$+ \int_0^T \left[ \frac{\partial H}{\partial a} + \left( \lambda_\mu - \lambda_\epsilon \right)ae^{-\eta t} + \left( \lambda_F - \lambda_\mu \right)Fe^{-\eta t} \right] dt.$$  \hspace{1cm} (A6)

Define $\lambda_\mu, \lambda_F$ to cause the coefficient of $\delta a, \delta F$, respectively, to vanish, such that

$$\dot{\lambda}_\epsilon = r\lambda_\epsilon - \frac{\partial H}{\partial a} e^{\eta t}, \quad \dot{\lambda}_F = r\lambda_F - \frac{\partial H}{\partial F} e^{\eta t},$$  \hspace{1cm} (A7a) with the boundary conditions

$$\lambda_\epsilon(T) = \lambda_\mu(T) = \lambda_\delta(T) = \lambda_F(T) = 0.$$  \hspace{1cm} (A7b)

Considering Eq. (A5a), conditions (A7a–b) become

$$\dot{\lambda}_\epsilon = r\lambda_\epsilon - (w + \lambda_\epsilon)(q-p-2qF)(k + \beta R), \quad \lambda_F(T) = \delta S(F(T), T) / \partial F e^{\eta T}$$  \hspace{1cm} (A8a)

and

$$\dot{\lambda}_\epsilon = r\lambda_\epsilon + 1 - \lambda_\mu R, \quad \lambda_\mu(T) = 0.$$  \hspace{1cm} (A8b)

Eq. (A6), after adding the zero sum

$$0 = \left[ -\lambda_\epsilon(0) - \int_0^T \frac{d}{dt}(\lambda_\epsilon e^{-\eta t}) dt \right] \delta a(0),$$

becomes

$$\delta \Pi = \int_0^T \left[ \frac{\partial H}{\partial a} \delta a + \left( \frac{\partial H}{a(0)} - \frac{d}{dt}(\lambda_\epsilon e^{-\eta t}) \right) \delta a(0) \right] dt.$$  \hspace{1cm} (A9)

Assume first that $a(0)$ is given and fixed, so that the variation on this parameter is zero. Then, for an extremum, $\delta \Pi$ must be zero for arbitrary $\delta R$. This can only happen if

$$\frac{\partial H}{\partial a} = \delta R, \quad t \in [0, T] \text{ or } (w + \lambda_\epsilon)(p + qF)(1-F) + \lambda_\mu a + \mu_\delta a = 0.$$  \hspace{1cm} (A10)

Thus, considering Eq. (A11), $\mu_\beta$ will be nonnegative, and thus, $(w + \lambda_\epsilon)(p + qF)(1-F) + \lambda_\mu a \leq 0$ if $R(T) = R$. Similarly, $\mu_\beta$ will be nonnegative, and thus, $(w + \lambda_\epsilon)(p + qF)(1-F) + \lambda_\mu a \geq 0$ if $R(T) = R$.

We can repeat this optimization process for a series of given $a(t)$. We want to consider the initial value $a(0)$ as an additional control parameter and solve the series of the above identical optimization problems with different $a(0)$. The particular value of $a(0)$ that yields the maximal value of $\Pi$ for the series of optimization problems must be the solution to the problem with unspecified initial value $a(0)$. Thus, we can expect that all necessary conditions derived above hold. Considering Eq. (A9), to have an extremum, we must also have the following hold for arbitrary $\delta a(0)$:

$$\lambda_\epsilon = r\lambda_\epsilon + \frac{\partial H}{\partial a(0)} e^{\eta t}.$$  \hspace{1cm} (A12)

Considering Eq. (A5a), $\frac{\partial H}{\partial a} e^{\eta t} = \mu_\beta$, and because Eq. (A8b) must hold, we obtain that the additional necessary condition is

$$\mu_\beta = 1 - \lambda_\mu R.$$  \hspace{1cm} (A13)

Considering Eq. (A11), the optimal value of $R(t)$, $R^*$, is constant. Thus, Eq. (A8b) can be integrated, yielding

$$\lambda_\mu(t) = -1 + e^{-(T-t)(r-R^*)} \frac{r-R^*}{R^*}.$$  \hspace{1cm} (A14)

Using the Taylor expansion of $e^{-(T-t)(r-R^*)} = 1 - (T-t)(r-R^*) + (T-t)^2(r-R^*)^2 - \ldots$, we obtain

$$\lambda_\mu(t) = -1 + e^{-(T-t)(r-R^*)} = -1 + (T-t) \left[ 1 - (T-t)(r-R^*) + \ldots \right] = 0.$$  \hspace{1cm} (A14a)

Substituting Eq. (A14a) in Eq. (A13) leads to

$$\mu_\beta = 1 + R'(T-t)e^{-(T-t)(r-R^*)} e^{\eta t} e^{(T-t)(r-R^*)} \geq 1 + R'(T-t)e^{(T-t)(r-R^*)} e^{\eta t} e^{(T-t)(r-R^*)}$$

$$= 1 + R'(T-t) \left[ 1 + R'(T-t) + |R'(T-t)|^2 + \ldots \right] e^{(R^*-r)(T-t)} > 0.$$  \hspace{1cm} (A15)
Considering Eqs. (A15) and (A1), we obtain
\[ a(0) = a_0. \]  
(A16)

Substituting Eqs. (A14a) and (A16) in Eq. (A11) and considering Eqs. (3a) and (3b), we obtain
\[ R(t) = \begin{cases} \frac{R}{(w + \lambda_F)(p + qF)(1 - F)} \beta \frac{e^{-(T-t)(r-R)}}{a \leq 0} & \text{if } t = T, \\ \frac{R}{(w + \lambda_F)(p + qF)(1 - F)} \beta \frac{e^{-(T-t)(r-R)}}{a \geq 0}. & \end{cases} \]  
(A17)

However, Eq. (A23) contradicts our assumption that \( \lambda_F(t) + w(t) \) continuously decreases from positive value at \( t=T \) to zero at \( t_1 \). This completes the proof of our lemma.

Because we know that \( g(T)>0 \) always holds, we can now begin evaluating the switching function \( g(t) \) by going backward in time from \( T \). Given Lemma 1, the first term in Eq. (A18) is always positive, and if \( t<T \), the second term is negative. Thus, the function \( g(t) \) may reach zero as the time span \( T-t \) increases. Let \( t_1 \) denote the point in time at which this happens. For notational efficiency, let us define two quantities \( \Delta \) and \( \bar{A} \) as follows:
\[ \Delta = (w + \lambda_F)(p + qF)(1 - F)|_{R = \bar{R}} \]  
(A24a)
\[ \bar{A} = (w + \lambda_F)(p + qF)(1 - F)|_{R = \bar{R}}. \]

Let us also define the conditional switching function \( g(t) \) as
\[ g(t) = \beta \Delta - (T-t) \frac{e^{-(T-t)(r-R)}}{a} \]  
(A24b)
\[ g(t) = \beta \bar{A} - (T-t) \frac{e^{-(T-t)(r-R)}}{a}. \]

Let \( g(t) > 0 \) for \( t_1 < t < T \) and \( g(t_1) = 0 \). If, in addition, \( g(t) < 0 \) for \( t < t_1 \) then \( t_1 \) can be termed a switching time (Sethi & Thompson, 2000). At time point \( t_1 \) the firm switches from the low to the high rate of advertising change.

We now proceed to characterize the optimal advertising paths after launch.

**Proposition 1.** Assume that for a given set of parameters \( \Theta = \{a_0, p, q, \alpha, \beta, \gamma, r, v, R, R, T \} \),
\[ a_0 \leq \frac{\beta}{(T-t)e^{-(T-t)(r-R)}} \bar{A} \forall t \in [0, T], \]  
(A25)

where \( \bar{A} \) is as in Eq. (A24a). Then, \( a^*(t) \) increases monotonically from \( a_0 \) to \( a_0 e^{RT} \), there is no switching time.

**Proof of Proposition 1.** The proof follows directly from the fact that Eq. (A25) is equivalent to \( g(t) \geq 0, R = \bar{R}, a(t) = a_0 e^{RT} \forall t \in [0, T] \).

Condition (A25) can be rewritten as
\[ \frac{\beta}{k + \beta \bar{R}} \leq \frac{-\lambda_F(t) a(t)}{(w + \lambda_F)(p + qF)(1 - F)} \]  
(A25a)

where \( \lambda_F(t) = -(T-t)e^{-(T-t)(r-R)} \) captures the future loss from increasing the rate of advertising change by one more unit at time \( t \) (see Eq. (A14a)). The inequality in Eq. (A25a) can be interpreted as follows. It requires, at any point in time under the optimal advertising policy, (i) the ratio between the effectiveness of the rate of change in advertising and the current multiplicative impact of all dynamic marketing variables to be higher or equal than (ii) the optimal advertising-to-sales ratio weighted by a factor scaling the cost of increasing advertising against the profit impact from increasing current adoptions.

Because the conditions under which there is no switching time in the interval \([0, T]\) also depend on the level of the initial advertising, we first find the minimum initial advertising level \( q_{\text{min}} \) that maximizes the total discounted profits \( \Pi \). If the firm were able to choose the lower bound for its initial advertising, what would that minimum be?

The value of \( a_0 \) that maximizes \( \Pi \) satisfies the necessary condition that it maximizes the maximized Hamiltonian, \( H^* \), where
\[ H^* = \left( (w + \lambda_F)(p + qF)(1 - F)(k + \beta \bar{R}) + a_0 e^{RT} (1 + \lambda_R) \right) e^{-rt}. \]
In the following lemma, we find this necessary condition.

Lemma 2. Let \( g_0 \) be the lower bound for \( a(0) \), i.e., \( \Omega_0(0) = |g_0| = 0 \). Then, the optimal \( g_0^* \) is the lowest possible positive value.

Proof of Lemma 2. Considering Eq. (A5a), Theorem 1 and Proposition 1, the maximized Hamiltonian will be

\[ H^* = (v + \lambda_R)(p + qF)(1-F) + q_0e^{Rt}(-1 + \lambda_R)|e|^{e^{-R}}. \]

Considering Eq. (A13), we obtain

\[ H^* = \left[ (v + \lambda_R)(p + qF)(1-F) + q_0e^{Rt}(-1 + \lambda_R) \right] |e|^{e^{-R}}. \quad (A26) \]

Considering Eqs. (A26) and (A15), we obtain

\[ \partial H^*/\partial q_0 = -e^{Rt}H_0e^{e^{-R}} < 0. \quad (A27) \]

Lemma 2 establishes that, under the GBM, it is optimal for the firm to advertise as little as possible at the time of launch.

Because the best initial value for advertising is as low as possible according to Lemma 2, the firm can choose it so that Eq. (A25) always holds when \( g_0 \) is arbitrarily small. Hence, given Lemma 2 and Proposition 1, it is optimal to start advertising at the lowest possible level and subsequently increase spending over the time at a constant rate \( R \).

In other words, the optimal policy \( a^*(t) \) increases monotonically from the lowest possible level. Hence, we have proven the intuition obtained from simply reflecting on the nature of the profit function under the GBM where the firm benefits from proportional increases in advertising regardless of levels but incurs costs for levels rather than changes in advertising.

Starting as low as possible and increasing monotonically is the optimal policy when the minimal admissible value \( g_0 \) satisfies Eq. (A25) in Proposition 1. If the firm cannot choose a sufficiently low value for the initial advertising, then a different strategy may be optimal. We identify these conditions and strategies in the following propositions.

Proposition 2. For a given set of parameters \( \Theta = (g_0, p, q, \alpha, \beta, \gamma, r, v, R, R, T) \), assume that condition (A25) is not satisfied and that there is a time point \( t_0 \), \( 0 < t_0 < T \), such that

\[ g(t) = \left\{ \begin{array}{ll}
0, & R_a(t) = g_0 e^{Rt}, \quad \forall t \in [0, t_0] \\
0, & R_a(t) = g_0 e^{R(t-t_0)}, \quad \forall t \in [t_0, T].
\end{array} \right. \quad (A28) \]

where \( \Delta \) and \( \Delta \) are as in Eq. (A24a). Then, \( a^*(t) \) first decreases from \( a_0 \) to \( g_0 e^{Rt} \) and then increases to \( g_0 e^{R(t-t_0)} \); there is one switching time at \( t_1 \).

Proof of Proposition 2. The proof follows directly from the fact that Eq. (A28) is equivalent to stating that there is \( 0 < t_1 < T \) such that

\[ g(t) = \left\{ \begin{array}{ll}
\leq 0, & R_a(t) = g_0 e^{Rt}, \quad \forall t \in [0, t_1] \\
\leq 0, & R_a(t) = g_0 e^{R(t-t_0)}, \quad \forall t \in [t_1, T].
\end{array} \right. \]

Condition (A28) can be rewritten as

\[ (-\lambda_R(t))a^*(t) = \frac{\lambda}{(w + \lambda_R(t))F(t)} \leq \frac{\beta}{k + \beta R} \quad t \in [0, t_1], \quad t \in [t_1, T]. \quad (A28a) \]

We can interpret the inequalities in Eq. (A28a) in a similar fashion as that in Eq. (A25a). Under the optimal policy, (i) the ratio between the effectiveness of the rate of change in advertising and the impact of advertising is lower than (ii) the advertising-to-sales ratio weighted by a factor scaling the cost against the gross margin generated by deviating from the optimal policy for all \( t \) in the interval \([t_1, T]\), but it is higher for all \( t \) in the interval \([0, t_1]\). Thus, it is optimal to first decrease and then increase advertising.

Proposition 3. For a given set of parameters \( \Theta = (g_0, p, q, \alpha, \beta, \gamma, r, v, R, R, T) \), assume that Eq. (A25) is not satisfied and that there are two time points \( t_1 \) and \( t_2 \), \( 0 < t_2 < t_1 < T \), such that

\[ g(t) = \left\{ \begin{array}{ll}
\leq 0, & R_a(t) = g_0 e^{Rt}, \quad \forall t \in [0, t_2] \\
\leq 0, & R_a(t) = g_0 e^{R(t-t_2)}, \quad \forall t \in [t_2, t_1] \\
\leq 0, & R_a(t) = g_0 e^{R(t-t_1)}, \quad \forall t \in [t_1, T].
\end{array} \right. \quad (A29) \]

Then, \( a^*(t) \) first increases from \( g_0 \) to \( g_0 e^{Rt_2} \), then decreases to \( g_0 e^{R(t_1-t_2)} \), and finally increases again to \( g_0 e^{R(t_1-t_1)} \); there are two switching times \( t_1 \) and \( t_2 \), \( 0 < t_2 < t_1 < T \).

Proof of Proposition 3. The proof follows directly from the fact that Eq. (A29) is equivalent to stating that there are \( 0 < t_2 < t_1 < T \) such that

\[ g(t) = \left\{ \begin{array}{ll}
\leq 0, & R_a(t) = g_0 e^{Rt}, \quad \forall t \in [0, t_2] \\
\leq 0, & R_a(t) = g_0 e^{R(t-t_2)}, \quad \forall t \in [t_2, t_1] \\
\leq 0, & R_a(t) = g_0 e^{R(t-t_1)}, \quad \forall t \in [t_1, T].
\end{array} \right. \]

Condition (A29) can be rewritten as

\[ g(t) = \left\{ \begin{array}{ll}
\leq 0, & R_a(t) = g_0 e^{Rt}, \quad \forall t \in [0, t_1] \\
\leq 0, & R_a(t) = g_0 e^{R(t-t_2)}, \quad \forall t \in [t_2, t_1] \\
\leq 0, & R_a(t) = g_0 e^{R(t-t_1)}, \quad \forall t \in [t_1, T].
\end{array} \right. \quad (A29a) \]

As before, the inequalities in Eq. (A29a) can be interpreted in terms of marketing mix effectiveness compared to the advertising-to-sales ratio weighted by profit impact from deviating from the optimal policy. In this case, the balancing of those considerations implies an optimal time policy with an increase–decrease–increase pattern.

More generally, similar results can be obtained for a sequence of \( n \) switching times \( t_1, t_2, ..., t_n \). We state the result in Proposition 4.

Proposition 4. For a given set of parameters \( \Theta = (g_0, p, q, \alpha, \beta, \gamma, r, v, R, R, T) \), assume that Eq. (A15) is not satisfied and that there are \( n+1 \) time points \( t_1, t_2, ..., t_n \) such that

\[ g(t) = \left\{ \begin{array}{ll}
\leq 0, & R_a(t) = g_0 e^{Rt}, \quad \forall t \in [0, t_1] \\
\leq 0, & R_a(t) = g_0 e^{R(t-t_2)}, \quad \forall t \in [t_2, t_1] \\
\leq 0, & R_a(t) = g_0 e^{R(t-t_1)}, \quad \forall t \in [t_1, T].
\end{array} \right. \quad (A30) \]

As before, the inequalities in Eq. (A30) can be interpreted in terms of marketing mix effectiveness compared to the advertising-to-sales ratio weighted by profit impact from deviating from the optimal policy.
Then, in case n is even (odd), $a'(t)$ first increases (decreases) from $g_0$ to $g_0e^{R_0}a_0g_0e^{R_0}$ (and continues alternating until it increases in the last interval $[t_1, T]$ from $g_0e^{R_0}(-R_0)(n-\tau_0 + (-1)^{n-1}\tau_0)$ to $g_0e^{R_0}R_0(n-\tau_0 + (-1)^{n-1}\tau_0)$). There are n switching times $t_1, t_2, ..., t_n$, $0 < t_n < ... < t_2 < t_1 < T$.

**Proof of Proposition 4.** The proof follows directly from the fact that Eq. (A30) is equivalent to stating that there are $0 < t_0 < ... < t_2 < t_1 < T$ such that

$$g(t) \begin{cases} 
\geq 0, R = R(t) = g_0e^{R_0}, & \forall t \in [0, t_1] \text{ if n is even} \\
\leq 0, R = R(t) = g_0e^{R_0}, & \forall t \in [0, t_0] \text{ if n is odd} \\
\leq 0, R = R(t) = g_0e^{R_0}(-R_0)(n-\tau_0 + (-1)^{n-1}\tau_0), & \forall t \in [t_2, t_1] \\
\geq 0, R = R(t) = g_0e^{R_0}R_0(n-\tau_0 + (-1)^{n-1}\tau_0), & \forall t \in [t_1, T] 
\end{cases}$$

Propositions 1–4 characterize the optimal advertising paths and determine sufficient conditions under which there is one, or more than one switching time. The propositions require one to check whether the conditions in Eqs. (A25), (A28), (A29) and (A30) are met. However, the expressions on the right-hand side of these inequality conditions depend on the number of switching times. As a result, one must proceed as follows. First, assume no switching and check whether Eq. (A25) is satisfied. If it is, then stop. If it is not, assume one switching time and check whether Eq. (A28) is satisfied. If it is, then stop. If it is not, assume two switching times and check whether Eq. (A29) is satisfied, and so on. While there is no closed-form expression for the number of switching times, they can be identified through this iterative procedure. We present an algorithm for this purpose below.

**Appendix III. An algorithm to find switching times**

**Step 0.** Set the optimal $a(0) = g_0$, and enter it into

$$h_0(t) = \frac{\beta}{(T-t)e^{-(T-t)r}} - a(0), \quad t \in [0, T]. \quad (A31)$$

If $h_0(t) \geq 0$ for all $t$ in the interval $[0, T]$, we have no switching time. Note that $h_0(T) = \lim_{t \to T^-} h_0(t) = + \infty$. If the $h_0(t) \geq 0$ condition is not satisfied, go to Step 1.

**Step 1.** Assume we have one switching time. Set the optimal $a(0) = g_0$, and enter it into

$$\tilde{h}_1(t) = \frac{\beta}{(T-t)e^{-(T-t)r}} - a(0), \quad t \in [t_1, T].$$

Again, $\tilde{h}_1(T) = \lim_{t \to T^-} \tilde{h}_1(t) = + \infty$. We must find $t_1 < T$ in the neighborhood of $T$ such that $\tilde{h}_1(t_1) = 0$. This implies $\tilde{h}_1(t) \geq 0$ in $[t_1, T]$. Then, check if

$$\tilde{h}_1(t) = \frac{\beta}{(T-t)e^{-(T-t)r}} - a(0) \leq 0, \quad t \in [0, t_1].$$

We call the double condition that $\tilde{h}_1(t) \geq 0$ in $[t_1, T]$ and $\tilde{h}_1(t) \leq 0$ in $[0, t_1]$ the **alternating sign condition**. Note that equality can happen strictly only at $t = t_1$. If the condition is met, then $t_1$ is the unique switching time. If the condition is not met, go to Step 2.

**Step 2.** Assume we have two switching times. Set the optimal $a(0) = g_0$ and enter it into

$$\tilde{h}_2(t) = \frac{\beta}{(T-t)e^{-(T-t)r}} - a(0), \quad t \in [t_1, T], t_2 - t_1$$

and

$$\tilde{h}_2(t) = \frac{\beta}{(T-t)e^{-(T-t)r}} - a(0), \quad t \in [t_2, t_1].$$

Again, $\tilde{h}_2(t) = \lim_{t \to T^-} \tilde{h}_2(t) = + \infty$. We must find $t_2 < t_1 < T$ such that $\tilde{h}_2(t_2) = 0$ and $\tilde{h}_2(t_1) = 0$. This implies $\tilde{h}_2(t) \geq 0$ in $[t_2, T]$ and $\tilde{h}_2(t) \leq 0$ in $[t_1, t_2]$. Then, check if

$$\tilde{h}_2(t) = \frac{\beta}{(T-t)e^{-(T-t)r}} - a(0) \geq 0, \quad t \in [0, t_2].$$

The conditions $\tilde{h}_2(t) \geq 0$ in $[t_1, T]$, $\tilde{h}_2(t) \leq 0$ in $[t_2, t_1]$ and $\tilde{h}_2(t) \geq 0$ in $[0, t_2]$ are the **alternating sign conditions for two switching times**, where equality can happen strictly only at the switching times. If these conditions are met, then $t_1$ and $t_2$ are the two switching times. If the alternating sign conditions in Step 2 are not met, we repeat Step 2 but search for three switching times, and so on, until we find the right number of switching times according to Proposition 4.

**Appendix IV. Optimal strategy with infinite planning horizon**

Spending as little as possible initially remains the optimal GBM policy when $T \to \infty$ and the planning period extends to $[0, \infty)$. This can be established by checking whether Eq. (A16) holds when $T \to \infty$ and then using Lemma 2. Condition (A16) indeed holds because Eq. (A15) holds:

$$h_0RHS = 1 + R^e(T-t)e^{-(T-t)(r-R)}$$

Increasing spending throughout the planning period remains optimal as well, provided that the discount rate, $r$, is higher than the upper bound of the growth rate in advertising spending, $R$. Under that condition, the inequality in Eq. (A25) holds for any finite value of the lower bound on initial advertising because

$$RHS \text{ of Eq. (A25) } = \frac{\beta}{(T-t)e^{-(T-t)r}} - a(0) \leq 0$$

The total optimal advertising over the entire diffusion cycle of the product will then go to infinity. That total equals $\int_0^T a'(t)dt = g_0 \int_0^T e^{pr}dt$ when Eq. (A25) holds at any time, and this integral does not converge to a finite value because we impose $R < 0$.

Note that the condition in Eq. (A25) is violated when $T \to \infty$ and $r < R$. In such cases, the optimal policy will involve at least one decrease in advertising spending and it can be identified using Propositions 2–4.

**References**


