On Continuous-Time Optimal Advertising Under S-Shaped Response

Fred M. Feinberg
University of Michigan Business School, 701 Tappan St., Ann Arbor, Michigan 48109
fred.feinberg@umich.edu

Continuous-time monopolistic models of advertising expenditure that rely on strict response concavity have been shown to prescribe eventual spending at a constant rate. However, analyses of discrete analogs have suggested that *S-shaped* response (convexity for low expenditure levels) may allow for the periodic optima encountered in actual practice. Casting the dynamic between advertising and sales in a common format (an autonomous, first-order relationship), the present paper explores extensions along three dimensions: an S-shaped response function, the value of the discount rate, and the possibility of diffusion-like response. Supplementing the treatment by Mahajan and Muller (1986), a flexible class of S-shaped response models is formulated for which it is demonstrated that, in contrast to findings in the literature on discretized advertising models, continuous periodic optima cannot be supported. Further, a set of conditions on the advertising response function are derived, that contains and extends that suggested by Sasieni (1971). Collectively, these results both suggest a set of baseline properties that reasonable models should possess and cast doubt on the ability of first-order models to capture effects of known managerial relevance. (*Advertising Models; Optimal Control; Contagion; Diffusion Models; S-Shaped Response*)

1. Introduction

A great deal of effort has been focused on the modeling of advertising expenditure, with an eye towards more efficiently allocating a given budget. With few exceptions, monopolistic (state), continuous-time models, and their oligopolistic variants, have suggested that advertising spending be evenly allocated over time. In practice, however, a variety of reasons are offered for the suboptimality of this constant level of allocation; more troubling is that it appears to underperform a number of other allocation schedules which allow one's message to be "heard above the noise."

The recent literature has suggested that an avenue for overcoming such difficulties involves the use of *S-shaped* response curves. As is evident in the most recent full review of the literature (Feichtinger et al. 1994), although numerous models of advertising have been analyzed through optimal control methods, analogous optimization for models incorporating S-shaped response is thus far largely unexplored (Vakratsas et al. 2000.) This paper addresses a class of models, differing from that of Sasieni (1971), that generalizes a number of such "classic" models (e.g., Ozga 1960, Vidale and Wolfe 1957, Mahajan and Muller 1986) under S-shaped response, showing through standard control techniques that periodic optima are disallowed.

Although S-shaped response has been analyzed in a general setting by a number of authors, these studies have typically addressed questions of optimization through simulation, rather than formal methods such as optimal control or dynamic programming. The question remains whether it is possible to formulate model variants with S-shaped response and yet retain enough tractability to present a formal analysis of their optimal inputs, if indeed these can be found. Such questions are of clear relevance to managerial

Management Science © 2001 INFORMS
Vol. 47, No. 11, November 2001 pp. 1476–1487

0025-1909/01/4711/1476$5.00
1526-5501 electronic ISSN

practice: Models which rule out, a priori, the possibility of optimal pulsing, blitzing, or flighting strategies may, for all their elegance and tractability, offer demonstrably incorrect advice in budget allocation over time.

It is in this spirit that we seek to understand the intrinsic properties of a class of first-order models that contains the "classic" models as special cases and allows for both S-shaped response and diffusion-like word-of-mouth effects. As a prelude to a discussion of the relevant literature, we briefly discuss standard terminology and formal specifications as they apply to the problem at hand.

1.1. Dynamic Equations and Profit Maximization

The most common specification for relating advertising (u) and sales (x) relies on an input-output framework for aggregate consumer response. In discrete time, such a relationship is typically captured by an analog of the lagged form, $x_{t+\tau} - x_t = \tau \cdot g(x_t, u_t)$, with updating period τ (Bass and Clarke 1972, Dhrymes 1981). In aggregate, where individual responses might take place at any time, this lagged form can be considered in the limit, $\dot{x} = g(x(t), u(t))$, or, disregarding the time argument, $\dot{x} = g(x, u)$. Although this is autonomous (i.e., with no explicit dependence on t), it is possible to include overtly temporal trends as well.

The firm wishes to maximize its discounted profit stream, which after appropriate scaling[1] (e.g., Feinberg 1992) can be taken to be the difference between its sales and advertising expenditure levels, with discount factor r. Thus, the standard approach of the literature involves a dynamic maximization of the following type:

$$\begin{aligned} &\max_{u(t)} \int_0^\infty [x(t) - u(t)]e^{-rt}dt \\ &\dot{x} = g(x, u), \qquad x(0) = x_0. \end{aligned} \tag{1}$$

Modifications of this system are possible in the form of control constraints (e.g., a maximal level of advertising u over a specified period) or state constraints (e.g., sales x should not be allowed to fall below a critical value), the latter addressed by Hartl et al. (1995). Note that, although $u(t)$ is unbounded, $x(t) \in [0, 1]$, so that if it is ever profitable to advertise, continuity alone (over the compact unit interval) suggests the existence of interior optima. It is further possible to amend (1) in a number of ways, including invoking a given finite horizon, appending a salvage value, or by employing a discount function, $r(t)$; a number of authors discuss the merits of doing so (Sethi 1977, Hahn and Hyun 1991, Feichtinger et al. 1994), although we will not pursue these variants here.

1.2. The Long-Run Response Function, $x_\infty(u)$

If the system (1) is considered at steady state (that is, as $\dot{x} \to 0$), a long-run response function, $x_\infty(u)$, is defined implicitly,

$$g(x_\infty(u), u) = 0, \tag{2}$$

for any constant level of advertising, u.[2] Such functions $x_\infty(u)$ exist, by the implicit function theorem, under quite general conditions, subject to the constraint $0 \le x_\infty(u) \le 1$. The function $x_\infty(u)$ relates the correspondence between long-run levels of advertising and of sales; holding discounting aside, this allows the optimal level of long-run advertising to be determined through univariate maximization, $\max_u[x_\infty(u) - u]$. Not being time dependent, optimal long-run u can be found by differentiation as the solution to $g_x \cdot [dx_\infty(u)/du] + g_u = 0$. Because a necessary condition for a non-zero extremum is $dx_\infty(u)/du = 1$, when long-run profits are maximized,

$$g_x + g_u = 0, \tag{3}$$

where both partials are evaluated at $(x_\infty(u), u)$, so that (3) is a univariate equation in u. The condition (3) must hold at any interior solution for u, so that u cannot lie at an endpoint of its set of possible values, here $[0, \infty]$.[3]

[1] Because such scaling restricts sales x to the unit interval, it would be equally correct to use the term "market share;" to avoid ambiguity, "sales" is used consistently here.

[2] It is not necessary that u be constant, but that the state variable x is driven toward a constant value. Thus, technically, u can be a chattering control, although such policies are, by their nature, not continuous.

[3] Given the unit scaling of x, however, $u < 1$ for optimal profit to be positive.

Though it ignores dynamics and discounting, (3) illustrates that essential properties of the optimal solution of systems like (1) are intimately linked to functional properties of g and its gradient. In §4, an analog of (3) is developed that accounts for the discount rate, r (see, for example, Teng and Thompson 1983). Such steady-state equilibria provide a useful benchmark to gauge the relative profitability of alternative advertising paths $u(t)$.

2. Selected Literature on Modeling Advertising Expenditure

From the point of view of having spawned a modeling substream, the first generally recognized model of importance was proposed by Vidale and Wolfe (1957). The V-W model posits that, while advertising directly persuades potential customers not currently buying from the firm, those who are currently buying tend to forget (buy less) over time, cast linearly in the relevant variables,[4] with scaling constants δ and ρ:

$$\dot{x} = \rho u(1-x) - \delta x. \tag{4}$$

An important feature of the V-W model concerns the shape of its response function. Consistent with the traditional economic interpretation of a negative second partial, the linearity of g in x and u is related to constant marginal returns. The V-W model has served as a basis upon which a variety of other models have been built (Mahajan and Muller 1986), typically by altering some aspect of the dynamic equation (4). Analyses of discrete analogs, which exhibit phenomena unique to the assumption of a known minimum period, are provided by Park and Hahn (1991) and Hahn and Hyun (1991). A full control analysis has been worked out for the V-W model (Sethi 1973), showing it is optimal in the long run for the state variable x to be driven to a constant level, the value of which is contingent upon the specific parameterization (i.e., r, δ, ρ); this is further addressed in §4. However, previous control analyses have not addressed variants of (4) for the more general types of response functions considered here.[5]

Sasieni (1971), in a classic paper, studied the optimal policies of a class of models with seemingly sensible economic properties (see also Sasieni 1989). Casting the dynamic problem in standard format (1), it is then shown that if the function g satisfies certain conditions,[6] then the optimal long-run policy entails constant sales x and advertising u.

One such condition, the concavity of g in u ($g_{uu} \leq 0$), has generated a good deal of debate, with several studies finding little evidence for discarding it (Simon and Arndt 1980, Steiner 1987), and others postulating models based on its rejection (e.g., Mahajan and Muller 1986). While there is no real argument that g should be asymptotically nonconvex, it is often conjectured (Steiner 1987), and supported by certain empirical findings (e.g., Eastlack and Rao 1986), that g should be *S-shaped* in u. Loosely stated, this implies that marginal returns to advertising are increasing for some period, after which they decrease; more precisely, g is convex in u for $u < u_{inf}$ (an inflection point), after which it is concave. Sasieni (1971) demonstrates that advertising on the convex portion of the curve is suboptimal, as such a spending policy can be dominated by mixing expenditure levels. That is, one can "chatter" infinitely quickly between $u = 0$ and $u = u_h$, the tangency point, to produce a superior level of response; this amounts to converting g to its convex hull, then re-solving the maximization problem. The appropriate specification of the response function has received considerable attention; Rao and Miller (1975) provide a complete review from a theoretical

[4] The specification is linear in the sense that the partials with respect to x and to u are then independent of their respective variables.

[5] An exception to this is Hartl (1987), who shows that, for a very broad class of continuous, autonomous response functions, the optimal *state* path is monotonic. The generality of Hartl's approach makes the monotonicity or asymptotic behavior of the control path difficult to specify. This is, in fact, the goal of the present study, relative to a more restricted class of response functions.

[6] The conditions for $g(x, u)$ are as follows. For any (x, u):

$g_u > 0$: Sales response would be greater were advertising at a higher level.

$g_x > 0$: Sales response would be greater were sales at a lower level.

$g_{uu} \leq 0$: Sales response exhibits diminishing returns to increases in advertising level.

perspective, while Rao (1986) considers the issue from the point of view of estimation techniques.

It should be noted that Sasieni's (1971) analysis does not, in fact, apply to all direct analogs of the V-W model. For example, the Contagion model (Ozga 1960, Sethi 1979) takes advertising to be essentially a word-of-mouth phenomenon, whereby sales response arises from the interaction of the buying segment x and the nonbuying segment $(1-x)$:

$$\dot{x} = g(x, u) = \rho u x(1-x) - \delta x. \tag{5}$$

Note that, because $g_x = \rho u(1-2x) - \delta$ (which, for $x < \frac{1}{2}$, can be made positive for large enough u), the partial condition $g_x < 0$ does not hold globally, and thus Sasieni's analysis is inapplicable. The optimal policy for the Contagion model has been shown by Gould (1970) and Sethi (1979) to be long-run constant spending (and therefore sales) under nonlinear and linear profit functions, respectively—the main substantive distinction between them being that, at optimum, the former dictates a continuous change of advertising level while the latter an abrupt one (a bang-bang control). In analyzing the response function (5), both these approaches, while adding considerably to the understanding of optima for contagion or diffusion-like processes (embodied by g being "parabolic" in x), are limited by linearity in u, a limitation addressed by the present study.

Mahajan and Muller (1986) examine an analog of the V-W model, generalizing the response function g in term of its advertising component:

$$\dot{x} = f(u)(1-x) - \delta x. \tag{6}$$

Although the qualitative nature of the optimal policy is derivable from Sasieni's analysis, Mahajan and Muller algebraically examine a variety of possible policy-types to compare their profit outputs. Since g is S-shaped in u (and $f(u)$ is chosen so that the optimum u falls below u_h), the best policy was found explicitly to be one approaching chattering. However, their analysis is based on numerical comparisons of simulated responses to a specific case of (6) for which no closed-form solution can be attained.

The goal of this paper is to examine the control properties of a class of models—distinct from those of Sasieni type (Sasieni 1971)—which both generalize prior models and offer a rationale as to when nonconstant advertising paths may be optimal. This model class allows for the inclusion of two important constructs, which have not been analyzed previously in a unified fashion: a response function that may increase in x, and which may be S-shaped in u. To this end, we examine models that are *separable* in their encoding of sales effects and the ostensive S-shapedness of advertising response. In the next section, this class of models is developed and analyzed.

3. A Generalized Linear Model

We examine a class of models which allows for S-shaped response and contains the classic models as special cases. We proceed by model specification and by deriving necessary conditions for "interior" optima; such conditions will then be shown to be economically sensible in two cases, depending on the number and location of zeros in a particular part of the response function, g; finally, control analyses will be presented for both cases, showing the impossibility of periodic optima. We note that this model class differs from that of Sasieni (1971)—which allowed for neither regions of positivity in g_x nor response functions that are not "chattered out"[7]—and allows one to understand their control properties with considerably less machinery.

We therefore propose to show that no optimal cyclical continuous policies exist for the following general model:

$$\begin{aligned} &\max_{u(t)} \int_0^\infty (x-u)e^{-rt}\,dt \\ &\dot{x} = f(u)\cdot a(x) - b(x), \qquad x(0) = x_0, \end{aligned} \tag{7}$$

where $f(u)$ is S-shaped. Models of type (7) offer the possibility of decoupling the S-shapedness of the response function from properties of the "acceleration" and "decay" functions, $a(x)$ and $b(x)$, of sales, as well as allowing optimization through a Hamiltonian

[7] Sasieni's analysis applied to strictly nonconvex response functions: Regions of convexity were assumed to be "chattered out" by mixing of expenditure levels. The characterization obtained below in Equations (11) and (14) for the phase-plane analysis, by contrast, are applicable for response functions of any shape.

mechanism. It is not necessary to place restrictions on the formal properties of the functions a and b at this point, though Sasieni-like restrictions could take the form $a_x < 0$, $b_x > 0$; we require, however, that a, b, and f possess continuous second derivative in line with prior literature. Application of the maximum principle yields the following necessary conditions for interior extrema (Sethi and Thompson 1981); subsequent functional arguments are suppressed for clarity:

$$H = (x-u) + \lambda(t)\cdot[f(u)\cdot a(x) - b(x)] \tag{8}$$

$$\partial H/\partial x = -\dot{\lambda} + r\lambda = 1 + \lambda\cdot[f\cdot a_x - b_x] \tag{9}$$

$$\partial H/\partial u = 0 = -1 + \lambda\cdot f_u\cdot a \tag{10}$$

$$\partial H/\partial \lambda = \dot{x} = f\cdot a - b. \tag{11}$$

The two singular curves, $\dot{x} = 0$ and $\dot{u} = 0$, must be characterized, and the first of these follows directly from (11). For the u singular curve, we proceed by taking the time derivative of (10):

$$-\dot{\lambda}\cdot\lambda^{-2} = f_{uu}\cdot a\cdot\dot{u} + f_u\cdot a_x\cdot\dot{x}. \tag{12}$$

Because $\lambda^{-1} = af_u$ (from (10)), $-\dot{\lambda} = 1 + \lambda[a_x f - b_x - r]$ (from (9)) and $\dot{x} = af - b$ (from (11)), (12) becomes:

$$(af_u)^2[1 + (a_x f - b_x - r)/af_u] = (af_{uu})\dot{u} + (a_x f_u)(af - b).$$

Then, $\dot{u}$ can be isolated as:

$$\begin{aligned}\dot{u} &= \left[(af_u)^2 + (af_u)(a_x f - b_x - r) - a_x f_u(af - b)\right](af_{uu})^{-1}\\ &= \left[(af_u)^2 - af_u b_x - af_u r + a_x f_u b\right](af_{uu})^{-1}\\ &= \left[af_u - r - b_x + b\frac{a_x}{a}\right]\frac{f_u}{f_{uu}}.\end{aligned} \tag{13}$$

This implies that, for $\dot{u}$ to be identically 0, the following condition must hold:

$$f_u = \frac{r}{a} + \frac{ab_x - ba_x}{a^2} = \frac{r}{a} + \frac{d}{dx}\left[\frac{b}{a}\right] \equiv h(x). \tag{14}$$

Thus, a complete characterization of the phase plane is offered by the two singular curve equations, (11) and (14), the second of which merely requires constructing the function $h(x)$. In general, h is unbounded in a neighborhood of any zero of a, so that such zeros serve to characterize functional types with "sensible" response properties.

The zeros of a respond to the question "when can the firm's share not increase *in principle*?" with two possibilities: when the firm already has the entire market ($x = 1$) or when there is, for example, nothing such as word of mouth to get it started ($x = 0$), the first of which will hold by definition. Because it is unreasonable from an economic vantage point for a to possess more than two zeroes, or for those zeroes to fall at points other than $x = 0$ or $x = 1$ (indicating advertising might be completely ineffective at some intermediate level of penetration), a logical and fairly general class of models can be addressed as to its control properties through application of (14).

Note that, for the Contagion model, there are two such zeroes (at $x = 0$ and $x = 1$), while for the Vidale-Wolfe, there is one (at $x = 1$). Generally speaking, the absence of response for $x = 0$ is a classic indicator of a diffusion-like response mechanism; as such, the S-shaped versions of these two models serve as exemplars for models with two, and with but a single, zero in the acceleration function, a. Because phase-plane analysis focuses on local properties of singular points, we examine both these specific cases, proofs for which are stated (except for algebraic manipulation) in terms of the general cases of two and one zero for a. Simply put, qualitative features of the phase plane for models of the form (7) are equivalent to one of the two cases analyzed below, neither of which, we will demonstrate, admit periodic optima.

3.1. Two-Zero Case: Contagion Model Analog

In keeping with the profit formulation in (1), we examine the following Contagion model analog:

$$\dot{x} = f(u)\cdot x\cdot(1-x) - \delta x, \qquad x(0) = x_0, \tag{15}$$

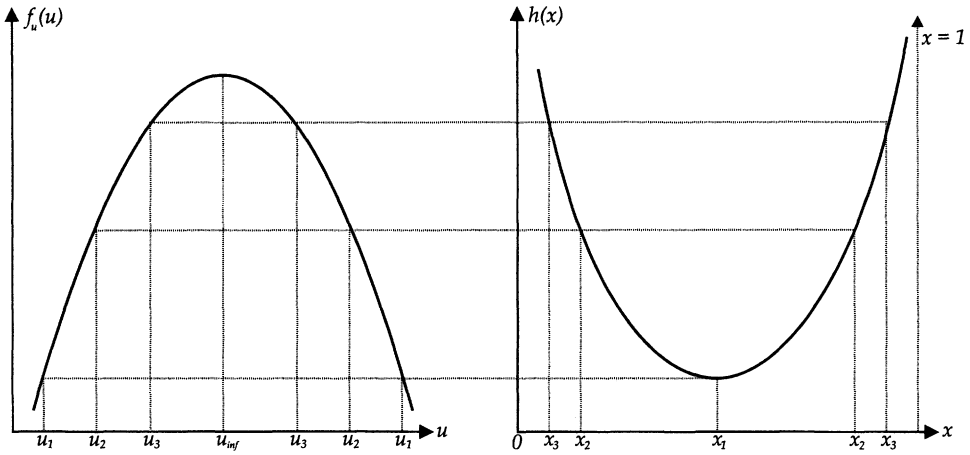
where $f(u)$ is S-shaped. To perform a phase-plane analysis, we first obtain the singular curves, $\dot{u} = 0$ and $\dot{x} = 0$:

$$\begin{aligned}\dot{x} = 0 &\Rightarrow 0 = f(u)\cdot x\cdot(1-x) - \delta x\\ &\Rightarrow x_\infty(u) = \max[0, 1 - \delta/f(u)].\end{aligned} \tag{16}$$

Note that f being S-shaped implies nothing about whether $1/f$ is as well;[8] however, it is not difficult to

[8] For example, if $f(u) = u^2/(2+u^2)$, both f and $1+f$ are S-shaped; however, while $-1/(1+f)$ is S-shaped, $-1/f$ is *not*.

Figure 1 Correspondence Between x and u



show that, if f is asymptotically concave, so is $-1/f$, hence the same is true for $x_\infty(u) = \max[0, 1-\delta/f(u)]$ by the Envelope Theorem.

We have from (14) that $\dot{u} = 0$ when:

$$f_u(u) = h(x) = \frac{r}{x(1-x)} + \frac{d}{dx}\left[\frac{\delta x}{x(1-x)}\right] = \frac{r}{x(1-x)} + \frac{\delta}{(1-x)^2}. \tag{17}$$

So, the cure $f_u(u) = h(x)$ must be characterized. Since it is assumed that f is S-shaped, f_u first increases and then decreases, reaching its maximum at the point of inflection of f. For any non-zero value of r, $h(x)$ has singularities at $x = 0$ and $x = 1$, and reaches a unique minimum between these values, as depicted in Figure 1.

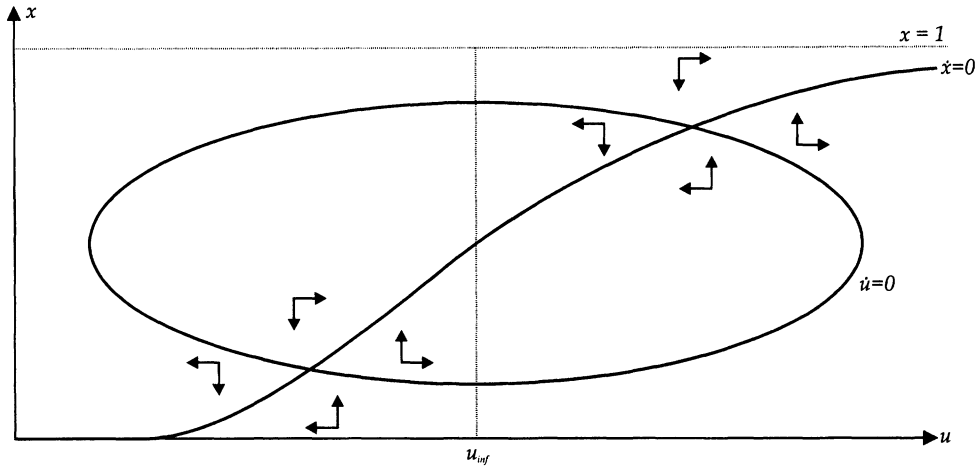
As this illustrates a particular, though representative, example, care must be exercised to distinguish essential qualitative features of the relation (17) from those peculiar to the example of Figure 1. For example, f_u needn't be zero for u in a neighborhood of zero, but can take on any positive value, as can $\lim_{u\to\infty} f_u(u)$, so long as they're each less than $f_u(u_{\text{inf}})$. Additionally, the f_u graph is depicted for a "significantly" S-shaped f; that is, f only has a relatively brief domain of convexity (near $u = 0$), so that a correspondingly brief segment of the f_u graph will be increasing.

Figure 1 depicts the correspondence between values of u and values of x, when the conditions imposed by the maximum principle (9–11) are fulfilled. A particular u value, as pictured, can be seen to correspond to (at most) two values of x, although if u becomes sufficiently large or small, it will match no x value; the precisely analogous statement holds for x. Distilling this information pictorially, it can be seen that, for a significantly S-shaped f, the $\dot{u} = 0$ singular curve resembles an ellipse-like closed contour in the $x - u$ plane. Drawing the $\dot{x} = 0$ and $\dot{u} = 0$ singular curves on the same set of coordinate axes yields the phase-plane diagram of Figure 2.

It is straightforward to determine the signs of $\dot{x}$ and $\dot{u}$ at each point of the phase plane. According to the defining equation (17) for $\dot{u}$, there are two factors determining sign, the left bracketed expression analyzed to obtain the singular carve, and an additional factor of $[f_u/f_{uu}]$. Because f is increasing and S-shaped, $f_u > 0$, with a point of inflection at $u = u_{\text{inf}}$, so that $f_{uu}(u_{\text{inf}}) = 0$. Thus, for u in a neighborhood of u_{inf}, $\dot{u}$ approaches a singularity, and the sign of $\dot{u}$ changes abruptly as the line $u = u_{\text{inf}}$ is crossed. Consistent with the functional forms (17) and (16), the phase-plane directions, in each of eight relevant areas, are as indicated in Figure 2 (an explicit direction field for a fully specified example is depicted in §3.3, showing the absence of optimal periodic controls).

Because the defining equations for $\dot{x}$ and $\dot{u}$ don't discriminate between local maxima and minima, there are two extremal trajectories. Note that, for the dynamic equation in (15), spending ever more on

Figure 2 **Phase-plane Diagram for the $\dot{x} - \dot{u}$ Singular curves**



advertising is eventually unprofitable, and spending nothing will eventually produce no profit (i.e., $f(0) = 0$). The upper and lower stationary points in the phase diagram correspond to, respectively, local maxima and minima; the locations of these stationary points are functions not only of the components of the response function, g, but also of the discount rate, r (as addressed by Sethi 1973 and in §4).

The lower stationary point, representing indefinite maintenance of poor profits, is of little practical interest. However, the presence of this minimum point implies that, for $f(u)$ sufficiently S-shaped, it is possible that in the long-run advertising could be unprofitable below a certain "threshold" level, after which it becomes increasingly profitable, until the optimal point is reached. By contrast, when $f(u)$ is not S-shaped (and $f(0) = 0$), advertising below the optimal long-run level is always profitable, though less so than at the optimal level.

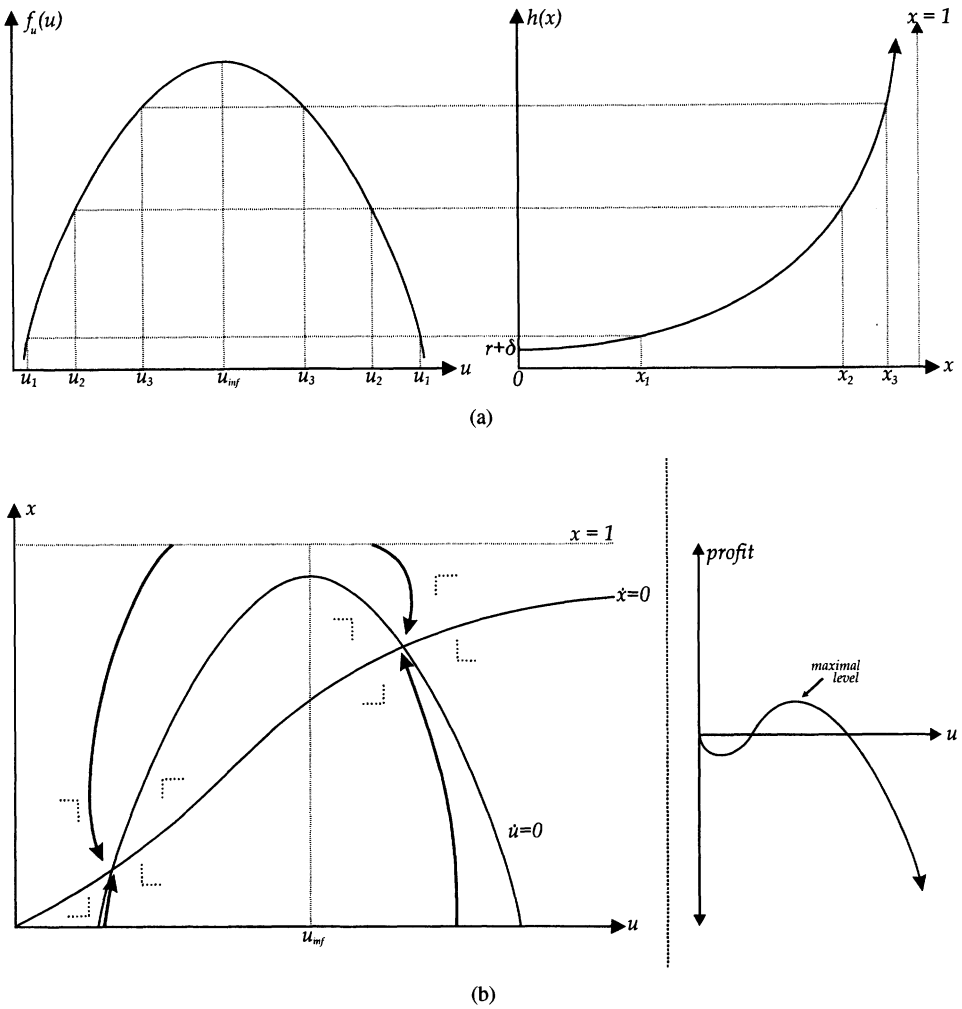
3.2. Single-Zero Case: Vidale-Wolfe Analog

For the Vidale-Wolfe model, $a(x) = (1-x)$ and $b(x) = \delta x$. Direct application of (14) implies that the $\dot{u} = 0$ singular curve is the solution of $f_u = h$, where $h(x) = r \cdot (1-x)^{-1} + \delta \cdot (1-x)^{-2}$. As in the case for the contagion model analog, h here is strictly convex. By analogy with the previous analysis, we construct Figure 3a. Having represented the relationship between u and x for the $\dot{u} = 0$ singular curve, we graph it along with the $\dot{x} = 0$ singular curve, $x_\infty(u) = 1 - (\delta/(\delta + f(u)))$ in Figure 3a and 3b; note that there is no x value corresponding to any u for which $f_u(u) < (r + \delta)$.

Although this phase-plane diagram is rather different from that for the Contagion model, they both share the essential feature that, for f sufficiently S-shaped, it is possible that there is a unique path toward a local minimum profit, a minimum that may in fact (if $f(0) = 0$, for instance) entail a negative profit. For both models, therefore, there can be a non-zero level of advertising under which it is unprofitable to advertise and over which it becomes increasingly lucrative up to some finite maximum point. It is important to realize that this is driven by the extent of the S-shape in f: A slightly S-shaped f will give substantially the same results as a non-S-shaped f; that is, advertising becomes increasingly profitable from $u = 0$ through some optimal u, after which profitability falls off.

The main difference between the two models lies with their optimal trajectories. Although it wasn't belabored in the phase-plane analysis of the Contagion analog, for every positive value of r, there is a positive value of x_0, the initial sales level, under which it is not possible to access the optimal path. This highlights that, in a model of contagion, the "seed," or proportion of the population needed to inform the remainder, must be large enough to warrant any spending of advertising funds at all, since advertising merely amplifies the voice of the seed. Because any results driven merely by the size of r

Figure 3 **(a) $x - u$ Correspondence (b) Extremal Trajectories**



alone are suspect, and the primary concern lies in the infinite horizon problem, the problem might be best approached taking the initial sales level to exceed this critical value. This assumption is also made by Gould (1970), presumably, whether it is reasonable to advertise at all would be quickly settled in an empirical setting. By way of contrast, there is no such minimal level for a "direct communication" model such as the V-W analog.

The optimal trajectory for both models (with a sufficiently S-shaped f), then, is to monotonically increase or decrease sales to its optimal value, represented by the stationary point in the phase diagram. In the V-W analog, this monotonicity is accomplished through a monotonic increase or decrease in u: For x_0 below the long-run optimal level, advertising should optimally start out above its long-run optimal value and gradually decrease as x (and u) approach the stationary point. The opposite is true when x_0 is above the long-run optimal level.

The Contagion model is not so easily characterized. As in the Vidale-Wolfe model, both x and u go monotonically to their long-run optimal values (x decreasing, u increasing) when x_0 is above the long-run optimum. However, when x_0 is below its long-run optimum, u should begin at some value (not necessarily lower or higher than its long-run value), gradually increase to greater than its long-run value,

and then decrease again. This demonstrates that, with contagion, a large initial advertising outlay to rapidly increase the effectiveness of word-of-mouth advertising is unwise; rather, u should start at a moderate level to slowly increase x and rise to higher levels as x becomes large enough to justify a large u in an attempt to amplify word of mouth.

We point out that neither of these models is in any sense the "correct" one, as advertising works its effects through a combination of mechanisms, direct conversion and word of mouth prominent among them. Since there are other models formally and conceptually close to these, and because the important qualitative aspects of the phase-plane diagram are driven mainly by formal properties of the response function (S-shaped $f(u)$, zeros of $a(x)$), a wide range of related models will admit optimal trajectories of the type presented here. In the following section, we present a fully specified example, for which an explicit phase plane, exhibiting no periodicity, can be calculated.

3.3. A Numerical Example

It is possible to illustrate the preceding derivation and methodology by choosing specific forms for the various functions and parameters making up the system (1), that is, a particular S-shaped f and values for the discount (r) and decay (δ) rates. The only consideration in such a choice concerns $\dot{x}$, namely that it can take on both positive and negative values.[9] For the purposes of illustration, the following system suffices:

$$\begin{aligned} &\max_{u(t)} \int_0^\infty (x-u)e^{-rt}\,dt, \\ &\dot{x} = f(u)x(1-x) - \delta x, \qquad x(0) = x_0 \\ &f(u) = 80u^3/(1+10u^3) \\ &r = 1/5, \qquad \delta = 1. \end{aligned} \tag{18}$$

It is easily verified that the function f is S-shaped with $f(0)=0$, $f'(0)=0$, $f(u)_{u\to\infty}=8$, $f'(u)_{u\to\infty}=0$, and an inflection point at $u=(20)^{-1/3}$. From (14), the singular curve for u can be determined by setting $h(x)=f'(u)$:

$$\frac{1}{5}[x(1-x)]^{-1} + (1-x)^{-2} = \frac{240u^2}{(1+10u^3)^2}. \tag{19}$$

Similarly, the singular curve for x can be determined by using (2):

$$x_\infty(u) = \max[0, 1-\delta/f(u)] = \max\left[0, \frac{70u^3-1}{80u^3}\right]. \tag{20}$$

Approximately 56% of the unit square lies below $x_\infty(u)$, where $\dot{x}$ takes positive values. These singular curves, along with the corresponding direction fields, are depicted in Figure 4.

Note that the main difference between the phase diagram and that drawn previously in Figure 2 concerns the shape of the x singular curve, which is strictly concave, while that for u is flattened at the bottom, both artifacts of the specific functional forms chosen; the singularity appears, as predicted, at $u=(20)^{-1/3}\approx 0.368$, with the direction field pointing inward within the u singular curve and outward elsewhere, and an absence of periodic controls at optimum.

4. Effect of the Discount Rate on Long-Run Optima

Given a model with constant-state long-run optimum, it remains to specify the actual levels of advertising u and sales x involved; these will allow an analysis of the state and controls themselves, as well as for profit, in terms of sensitivity to problem parameters and the specification of the dynamic equation governing their interrelations. This can be accomplished by reconsidering the current-value Hamiltonian for the system (1):

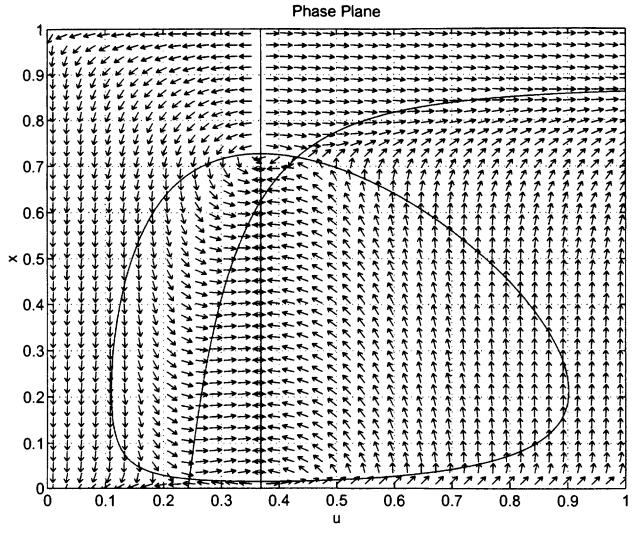
$$\begin{aligned} H &= (x-u) + \lambda g; \\ \partial H/\partial x &= -\dot{\lambda} + r\lambda = 1 + \lambda g_x; \\ \partial H/\partial u &= 0 = -1 + \lambda g_u. \end{aligned} \tag{21}$$

The last of these implies:

$$\lambda = -1/g_u, \qquad \dot{\lambda} = -\frac{g_{xu}\cdot\dot{x} + g_{uu}\cdot\dot{u}}{g_u^2}. \tag{22}$$

[9] Specifically, for any given value of sales x, it should be possible to choose a low enough value of advertising u to cause sales to fall, and a high enough value to cause it to rise (except perhaps for limiting cases, where u or x are near the edges of the unit square). This can be accomplished by having the curve $x_\infty(u)$ divide the unit square more or less evenly.

Figure 4 Phase Plane and Direction Field for Specified Model



The positivity of g_u implies that the denominator g_u^2 is non-zero, so $\dot{\lambda}$ vanishes as the stationary point $(\dot{x}, \dot{u}) = (0, 0)$ is approached. The adjoint variable λ can be eliminated from (21), yielding the following conditions at the optimal pair (see Teng and Thompson 1983):

$$g = 0, \qquad g_x + g_u = r. \tag{23}$$

The conditions (23) define a system in the unknowns (x^*, u^*), valid for any Sasieni-type model (1), and agree with (3) when $r = 0$ (Sasieni 1971). By way of example, application of (23) to the V-W formulation replicates the optimal long-run pair (x^*, u^*) found by Sethi (1973) through use of Green's theorem, $u^* = [-2\delta + (r^2 + 4\rho\delta)^{1/2} - r]/2\rho$, $x^* = \rho u^*/(\delta + \rho u^*)$. It is similarly straightforward to apply (23) to the Contagion model or to any functionally specified model of the Mahajan-Muller (7) form. The conditions (23) further allow the calculation of the direction of optimal changes in long-run advertising, sales and profit when the discount rate is changed. Construing x^* and u^* as functions of r, (23) can be differentiated:

$$\begin{aligned} g_x \frac{\partial x^*}{\partial r} + g_u \frac{\partial u^*}{\partial r} &= 0, \\ \left[g_{xx} \frac{\partial x^*}{\partial r} + g_{xu} \frac{\partial u^*}{\partial r} \right] + \left[g_{ux} \frac{\partial x^*}{\partial r} + g_{uu} \frac{\partial u^*}{\partial r} \right] &= 1. \end{aligned} \tag{24}$$

This is a 2×2 system in $\partial x^*/\partial r$ and $\partial u^*/\partial r$, with solution:

$$\begin{aligned} \partial x^*/\partial r &= -g_u/D, \qquad \partial u^*/\partial r = g_x/D, \\ \partial \pi^*/\partial r &= -r/D, \\ D &= g_x[g_{ux} + g_{uu}] - g_u[g_{xx} + g_{xu}], \end{aligned} \tag{25}$$

where $\pi^* = x^* - u^*$. For a Sasieni-type model (1), we have the conditions $\{g_x < 0, g_u > 0, g_{uu} < 0\}$, which imply that $\partial u^*/\partial r$, $\partial x^*/\partial r$ and $\partial \pi^*/\partial r$ are all of the same sign, opposite of the denominator D. However, these conditions neglect to specify the signs of several second derivatives of g.

Using the V-W (4), Contagion (5) and Mahajan and Muller (6) specifications as guides, the following partial conditions, containing Sasieni's, are suggested:

$$\begin{aligned} &g_x < 0, \qquad g_u > 0 \\ &g_{uu} < 0, \qquad g_{ux} = g_{xu} \le 0, \qquad g_{xx} \le 0. \end{aligned} \tag{26}$$

Jointly, the conditions (26) imply that D is nonnegative, and positive if g is strictly concave in u, in which case $\partial u^*/\partial r$, $\partial x^*/\partial r$ and $\partial \pi^*/\partial r$ are negative. In short, if the discount rate is increased, optimal long-run response prescribes a lower level of advertising and, consequently, of sales and profit; these results hold for any Sasieni-type model (1). In a sense, such results are to be expected: If future earnings are discounted, because advertising's benefits are reaped on

a time scale determined by a fixed g, firms should be less willing to invest in the future, dictating lower levels of advertising and, consequently, of sales and profit. However, the fact that such results—seemingly dictated by standard economic reasoning—hinge on a variety of partial conditions (26) which have not appeared in prior theoretical analyses, suggests that the subset of conditions put forth by Sasieni (1971) may not capture all "reasonable" assumptions regarding the specification of the response function $g(x, u)$.

5. Discussion and Managerial Implications

Many models proposed to explain advertising spending have tended to fall into the form explored by Sasieni (1971), and therefore prescribe eventually spending at a constant rate. What has been relatively underexplored, vis-à-vis optimal monopolistic advertising in "standard" format, has been the qualitative effects of varying analytic specifications. In the present paper, the explicit focus was on generalization along three dimensions: the partial convexity of $g(\cdot, u)$, the value of the discount rate r, and the number of zeroes of the "acceleration" portion of $g(x, \cdot)$. An S-shaped relationship between the rate of increase in sales $\dot{x}$ and advertising u has the effect, from an analytic vantage point, of generating a curve of joint zeroes for the $(\dot{x}, \dot{u})$ pair; by comparison, the number of zeroes for the "acceleration" portion of the response function (for V-W-like models), dictates whether there is a roughly parabolic or elliptical shape for the $\dot{u} = 0$ contour, and thus the maximal number of critical points. These findings attest to the applicability of classical phase-plane analysis in cases where there are unspecified nonlinearities in the response function, g. The discount rate is seen to affect the qualitative nature of resulting optima, even at steady-state; further, under certain sufficient conditions derived here, increasing the discount rate has a "damping" effect on the values of optimal advertising, sales and profit in the long-run. Finally, a simple criterion was derived to calculate explicit solutions for models which admit a steady-state optimum, so that the effects of parameterization can be directly analyzed. As Sasieni points out, because models with some convexity in response can be converted to the form considered here, a variety of models with long-run constant optima conform to the present characterization.

A main implication of the analysis presented, in terms of the ability to model and optimize advertising expenditure, concerns the very possibility of adequately capturing advertising effects through a first-order mechanism, that is, one that does not at least implicitly account for stock effects, which can be accomplished through a second-order approach (Feinberg 1992, Bronnenberg 1998). In short, first-order models displaying separability of sales and advertising effects, and imposing a priori reasonable response assumptions, lead to the conclusion that "flighting" or pulsing policies can never be optimal. A second implication of the present study is that S-shaped advertising response, taken alone, does not lead to pulsing optima in a broad class of models. Thus, capturing the empirically justified phenomenon of pulsing in the fullest temporal generality—through a continuous-time mechanism—will require a departure from the types of response functions on which classical models have been devised.

Acknowledgments

The author wishes to thank John D. C. Little for his help and guidance, as well as the associate editor and anonymous reviewers.

References

Bass, F. M., D. G. Clarke. 1972. Testing distributed lag models of advertising effect. *J. Marketing Res.* **9** 298–308.

Bronnenberg, Bart J. 1998. Advertising frequency decisions in a discrete Markov process under a budget constraint. *J. Marketing Res.* **35** 399–406.

Dhrymes, P. J. 1981. *Distributed Lags*. North-Holland, Amsterdam, The Netherlands.

Eastlack, J. O., A. Rao. 1986. Modeling response to advertising and pricing changes for V8 cocktail vegetable juice. *Marketing Sci.* **5**(3) 245–259.

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

Feinberg, Fred M. 1992. Pulsing policies for aggregate advertising models. *Marketing Sci.* **11**(3) 221–234.

Gould, J. P. 1970. Diffusion processes and optimal advertising policy. E. S. Phelps et al., eds., *Microeconomic Foundations of Employment and Inflation Theory*. Norton, New York.

Hahn, Minhi, Jin-Sok Hyun. 1991. Advertising cost interactions and the optimality of pulsing. *Management Sci.* **37**(2) 157.

Hartl, R. F. 1987. A simple proof of the monotonicity of the state trajectories in autonomous control problems. *J. Econom. Theory* **41** 211–215.

——, S. P. Sethi, R. G. Vickson. 1995. A survey of the maximum principles for optimal control problems with state constraints. *SIAM Rev.* **37**(2) 181.

Little, J. D. C. 1979. Aggregate advertising models, the state of the art. *Oper. Res.* **27** 629–667.

Mahajan, V., E. Muller. 1986. Advertising pulsing policies for generating awareness for new products. *Marketing Sci.* **5**(2) 89–106.

Ozga, S. 1960. Imperfect market through lack of knowledge. *Quart. J. Econom.* **74** 29–52.

Park, Sehoon, Minhi Hahn. 1991. Pulsing in a discrete model of advertising competition. *J. Marketing Res.* **28**(4) 397.

Rao, A. G., P. B. Miller. 1975. Advertising/Sales Response Functions. *J. Advertising Res.* **15**(2) 1–15.

Rao, R. 1986. Estimating continuous time advertising-sales models. *Marketing Sci.* **5**(2) 125–142.

Sasieni, Maurice W. 1971. Optimal advertising expenditures. *Management Sci.* **18**, 64–72.

——. 1989. Optimal advertising strategies, *Marketing Sci.* **8**(4).

Sethi, S. P. 1973. Optimal control of the Vidale-Wolfe advertising model. *Oper. Res.* **21** 998–1013.

——. 1977. Dynamic optimal control models in advertising. *SIAM Rev.* **19**(4) 685–725.

——. 1979. Optimal advertising with the contagion model. *J. Optim. Theory Appl.* **29**(4) 615–627.

——. Gerald L. Thompson. 1981. *Optimal Control Theory: Applications to Management Science.* M. Nijhoff Publishers, Hingham, MA, and Kluwer, Boston, MA.

Simon, J. L., J. Arndt. 1980. The shape of the advertising response function. *J. Advertising Res.* **20** 11–28.

Steiner, R. L. 1987. Point of view: The paradox of increasing returns to advertising. *J. Advertising Res.* **27** 45–53.

Teng, Jin-Tsair, Gerald L. Thompson. 1983. Oligopoly models for optimal advertising. *Management Sci.* **29**(9) 1087–1101.

Vakratsas, D., G. K. Gurumurthy, F. Bass, F. M. Feinberg. 2000. The shape of the advertising response function revisited: A model of advertising effects with dynamic probabilistic thresholds. Working paper, University of Texas, Dallas, TX.

Vidale, M. L., H. B. Wolfe. 1957. An operations research study of sales response to advertising. *Oper. Res.* **5** 370–381.

Accepted by Dipak Jain; received May 11, 1999. This paper was with the author for 2 revisions.