Optimal Resource Allocation with Time-varying Marketing Effectiveness, Margins and Costs

Kalyan Raman, a,⁎ Murali K. Mantrala, b Shrihari Sridhar c & Yihui (Elina) Tang d

a Medill School of Journalism, Media, Integrated Marketing Communications, IMC Program, Northwestern University, Evanston, IL 60208, USA
b Marketing, Trulaske College of Business, 438 Cornell Hall, University of Missouri, Columbia, MO 65211-1200, USA
c Marketing, 457, Smeal College of Business, Pennsylvania State University, University Park, PA 16802, USA
d Marketing, College of Business, Illinois State University, Campus Box 5590, Normal, IL 61790-5590, USA

Available online 27 August 2011

Abstract

The importance of optimal marketing communications mix decisions is well-recognized by both marketing scholars and practitioners. A significant volume of work has addressed the problem of dynamic marketing mix optimization assuming constant effectiveness of marketing instruments. However, the effectiveness of marketing communications varies over time for a variety of reasons. Moreover, due to factors such as inflation or deflation in media prices and/or raw material inputs, there can be differential changes in the costs of communications and/or margins on the good (or service) sold over time. The academic literature offers little normative direction on how time-varying marketing effectiveness and costs drive optimal marketing-mix levels and their relative allocation. The authors shed light on these issues by solving a monopoly firm’s finite horizon dynamic marketing communications mix optimization problem involving two marketing instruments with time-varying parameters, i.e., the marketing effectiveness parameters, media costs, and product margin are all allowed to vary over time. First, they find that the structure of the solutions is similar to that of the classic Nerlove–Arrow model, for a completely general nature of time-varying effectiveness. Second, their model can be used by managers to exactly determine whether and when to switch their marketing-mix emphasis (defined by the marketing element receiving the dominant portion of the budget) over a finite planning horizon. In sum, the authors expand knowledge on optimal allocation of marketing resources with time-varying effectiveness. They also extend their solution to incorporate multiple (more than two) marketing instruments.

© 2011 Direct Marketing Educational Foundation, Inc. Published by Elsevier Inc. All rights reserved.

Keywords: Time-varying marketing effectiveness; Time-varying marketing costs; Resource allocation; Finite-horizon optimal control

Introduction

Companies’ marketing communications resource allocation decisions have become considerably more complex as the channels available to reach consumers have expanded to include more interactive marketing vehicles, e.g., online display, paid search, mobile, and social media, in addition to traditional marketing vehicles, e.g., TV, print, radio, and personal selling (Barwise and Farley 2005; Naik and Peters 2009; Marketing Science Institute 2010–2012 Research Priorities). However, a recent global survey by McKinsey & Co. (Doctorow, Hoblit, and Sekhar 2009) reports that companies tend to allocate marketing spending based on historical allocations and rules of thumb far more than quantitative measures. This state of affairs has persisted for decades (e.g., Mantrala 2002) even though optimal allocation can significantly enhance a firm’s profitability, sometimes by as much as 400% (e.g., Raman 2010). Consequently, a large volume of work in marketing has focused on developing normative rules for marketing resource allocation decisions based on models of market response to marketing efforts (e.g. see surveys of the literature by Gupta and Steenburgh 2008; Hanssens, Parsons, and Schultz 2003; Mantrala 2002; Shankar 2009).

A core insight from extant normative analyses is that, subject to cost considerations, a marketing input with higher effectiveness...
(sales response elasticity) deserves more of the overall budget’s allocation than one that is less effective (e.g., Gatignon and Hanssens 1987). However, models for dynamic marketing resource allocation typically assume that marketing effectiveness is constant over time. In reality, marketing effectiveness can vary over time, e.g., consumer segments, values and tastes change as products age, and the competitive landscape or economic conditions change, making the aggregated market less or more responsive over time to marketing efforts, e.g., Shankar (2009). Empirical studies that have documented time-varying effectiveness of marketing instruments include Erickson and Montgomery (1980), Jedidi, Mela, and Gupta (1999), Krishnamurthi and Papatla (2003), Mahajan, Bretschneider, and Bradford (1980), Naik, Mantrala, and Sawyer (1998), Parsons (1975), Winer (1979).

A concise summary of empirical models of marketing dynamics incorporating time-varying parameters is provided by Leeflang et al. (2009) who note that they expect time-varying parameter issues to be increasingly relevant in future research. Contributing to this expectation is the emergence of online media as important channels of communication in addition to traditional offline sources (Naik and Peters 2009). It is acknowledged that offline and online media communications can have varying effectiveness and impacts over different stages of consumer goods buyers’ “path-to-purchase” (e.g., Shankar et al. 2011) and business customers’ “purchase funnels” (e.g., Mantrala and Albers 2010; Wiesel, Pauwels, and Arts forthcoming). As a result, optimal budget allocations between online and offline marketing vehicles are likely to vary over the course of a new consumer or business product marketing campaign. Moreover, even in situations where the effectiveness of marketing inputs is constant, there can be differential changes in the costs of communications over time due to environmental factors such as inflation or deflation in media prices (e.g., Bradshaw 2010). Time-varying costs can also impact optimal media expenditure ratios, since they contribute to the overall effectiveness of a marketing resource from a profitability standpoint.

Existing research, however, provides little guidance on the implications of time-varying effectiveness and costs for marketing mix allocations over short and intermediate term horizons (e.g., 2–3 years) even though it is a fundamental managerial issue (Mahajan, Bretschneider, and Bradford 1980). In their review of marketing dynamics research, Leeflang et al. (2009, p 16) note that there is an “acute shortage of normative studies developing navigation systems that allow managers to optimize marketing efforts, or at least investigate what-if scenarios.” Two questions of considerable interest and importance to marketing managers are what the balance between expenditures allocated toward two or more marketing activities, e.g., an online and offline activity, should be, and how that should change over the planning horizon when effectiveness and/or cost parameters are time-varying (see also MSI 2010–2012 Research Priorities, p 9).

In this paper, we derive model-based normative rules for optimal marketing activity planning by a monopoly firm in the presence of time-varying effectiveness. Specifically, we consider a two-variable extension of the well-known Nerlove–Arrow (NA) model with time-varying effectiveness and cost parameters (or ‘TVNA’ model) and solve for the optimal ratio of marketing resources over the planning horizon by applying new developments in finite horizon optimal control (Raman 2006). We obtain a solution that has two notable properties: first, it has a simple, closed-form that generalizes the solution to the classic NA model that assumes constant effectiveness parameters. Second, the solution is general in that it can incorporate any continuous form of time-varying effectiveness of the activities (e.g. linear increases, sinusoidal, state-dependent). That is, our result informs managers about the trajectory of optimal marketing allocations over a finite horizon, whatever be the functional form of the time-variation in the two marketing inputs’ effectiveness they face or wish to examine. For expositional clarity, we focus on two marketing activities but generalize our analysis to multiple activities, as in Naik and Raman (2003), in Appendix B.

The rest of the paper is organized as follows. In the next section, we first develop the TVNA model which proposes a general relationship between a firm’s sales and marketing investments allowing for time-varying effectiveness. Then, using optimal control theory, we analytically determine how a firm whose sales follow the TVNA model should set its marketing investments optimally over a finite planning horizon. Further analysis of the general solution provides insights into our remaining research questions. We conclude with a summary of the managerial takeaways and suggested directions for future research.

Model Development

Market Response Model Formulation

While a variety of aggregate dynamic marketing response models exist in the literature (see e.g., Little 1979), perhaps the most parsimonious is the classic goodwill accumulation model of Nerlove and Arrow (1962). The basic idea of the NA model is that goodwill accumulates as the spending levels of marketing activities increase, and decays exponentially when marketing activities are turned off. In practice, goodwill can be related to relevant observed outcomes such as sales. Specifically, a two-variable form of the NA model is given by

$$\frac{dS}{dt} = -\delta S + \beta_1 u_t + \beta_2 v_t$$  \hspace{1cm} (1)$$

where \(S\) is the sales of the product, \(\delta\) represents the rate of decay in sales, and \(u\) and \(v\) represent the units of the two marketing activities (e.g. number of sales calls, ad exposures, to which we will add cost-multipliers subsequently). Also \(\beta_1\) and \(\beta_2\) represent the effectiveness of each of the marketing activities in generating sales. For tractability, we assume the sales growth rate is a linear function of the two marketing efforts, \(u\) and \(v\), whose monetary costs are time-varying and increase in a convex fashion as their units increase. While the classic Nerlove–Arrow model relates the long-term cumulative effect of advertising to a construct called “goodwill,” it is equivalent to Equation (1) when sales is assumed to be a linear function of goodwill as we do, consistent with previously published research, e.g., Naik and Raman (2003) and Raman (2006, 2010).
Time-varying Nerlove–Arrow (TVNA) Market Response Model

We now extend Equation (1) to incorporate time-varying effectiveness parameters as follows:

\[
\frac{dS}{dt} = -\delta S + \beta_1(t)u_t + \beta_2(t)v_t
\]  

(2)

where all terms are as previously defined, except that \( \beta_1(t) \) and \( \beta_2(t) \) now reflect the time-varying effectiveness of \( u \) and \( v \), respectively. We allow \( \beta_1(t) \) and \( \beta_2(t) \) to be general functions of time with the only restrictions that they be continuous and differentiable at least once. This allows specification of a variety of functions such as polynomial functions (e.g. Winer 1979) in a specific application. An example of a linear polynomial function of time is

\[
\beta_1(t) = a_0 + a_1t.
\]  

(3)

Another example of a TVNA model is the model of advertising effectiveness formulated by Naik, Mantrala, and Sawyer (1998), which allows for state-dependence in the effectiveness term. In practice, researchers can choose the best-fitting form.

Decision Problem Formulation and Solution

We consider a firm that is interested in maximizing its profits over a finite time horizon \( T \), which could be several years long in many companies’ marketing planning process (e.g. McDonald and Keegan 2001) or just a few weeks for short life-cycle campaigns or products such as motion pictures. We allow the firm’s margin \( m(t) \) on sales (price less cost of goods sold) to vary over time in an arbitrary continuous manner, e.g., due to fluctuations in raw material input costs even though the price stays fixed. We assume that at the end of the planning horizon of length \( T \), the firm seeks to salvage a fraction \( \theta \) of its contribution revenues \( m(T)S(T) \), where \( S(T) \) is the value of the final level of the state variable at the end of the planning horizon. Therefore, after accounting for a discount factor \( \rho \) that the firm places on its revenue, the discounted salvage value at the end of the planning horizon is given by \( m(T)S(T)e^{-\rho T} \) where the parameter \( \theta > 0 \), captures a number of substantively interesting scenarios (see, e.g., Raman’s 2006 use of this salvage value formulation in the optimal planning of advertising over a finite planning horizon with constant advertising effectiveness).

Incorporating these assumptions, the firm seeks to optimally allocate \( u(t) \) and \( v(t) \) over its planning horizon \( T \) to maximize discounted long-term profits. This problem is mathematically expressed as:

Maximize \( J(u, v) = \int_0^T e^{-\rho t} \pi(S(t), u(t), v(t))dt + mSTe^{-\rho T} \),

(4)

where \( J \) is the objective functional of the firm,

\[
\pi(S, u, v) = m(t)S[c_1(t)u^2 - c_2(t)v^2],
\]  

(5)

subject to the dynamics in Equation (2) and the salvage value \( mSTe^{-\rho T} \).

We introduce a quadratic cost of effort structure as given by the squared-terms pertaining to \( u \) and \( v \) in Equation (5). Additionally, we allow for the monetary costs per unit of the two marketing efforts to be different. For example, Gopalakrishna and Chatterjee (1992) state that the unit cost of a sales call is much higher than the unit cost of a print advertisement. Similarly, the cost per 1000 impressions (CPM) or “cost per action” (CPA) varies across online and offline media ads (e.g., Evans 2009). We also allow these costs of marketing activities to be time-varying, e.g., due to factors like media price inflation. Specifically, we capture this through the parameters \( c_1(t) \) and \( c_2(t) \) in Equation (5). Thus, \( c_1(t) > c_2(t) \), implies that the unit cost of \( u \) is higher than that of \( v \) at \( t \).

Solution Procedure

As highlighted earlier, a majority of the work on dynamically optimal marketing resource allocation assumes constant effectiveness. A notable exception is the work by Aravindakshan, Peters, and Naik (2011) that addresses spatio-temporal optimal allocation of advertising with time-varying parameters. However, this work, akin to the majority of extant research on integrated marketing communications (IMC), studies the optimal allocation of resources over an infinite planning horizon. Mathematically, the assumption of infinite horizon makes the dynamic optimization problem more tractable to solve. However, most real-world-planning scenarios occur over finite horizons. Time-varying effectiveness parameters in conjunction with optimization over a finite horizon makes discovery of the analytical solution mathematically challenging but the payoff is its greater suitability for managerial implementation. Below we present the key normative results of managerial interest while relegating more details about the solution procedure to Appendix A for the interested reader.

Step 1. Defining the value function

The key term in the problem is defined by the value function \( V(s, t, T) \),

\[
V(s, t, T) = \max_{\pi(w, t)} \left[ e^{-\rho (T-t)} \pi(S(w), U, w)dw \right]
\]

\( \pi(.) \) denotes the maximum profit that can be achieved by using the optimal marketing mix over the remaining horizon \( [t, T] \), starting from an arbitrary state \( s \), and \( U \) denotes the control vector — \( U = (u(t), v(t)) \). The instantaneous profit is \( \pi(t) = mTS(t) - c_1(t)u(t)^2 - c_2(t)v(t)^2 \) where \( S(t) \) is the instantaneous sales rate at time “\( t \).”

Step 2. Hamilton–Jacobi–Bellman (HJB) equation

The value function \( V(s, t, T) \) satisfies the Hamilton–Jacobi–Bellman (HJB) equation (Fleming and Rishel 1975),

\[
V_t + \max_u \left[ e^{-\rho t} \pi(s, U, t) + V_s f(s, U, t) \right] = 0
\]

where \( V_s = \partial V / \partial s \), and \( V_t = \partial V / \partial t \). The function \( f(s, U, t) \) is defined by the right-hand side of Equation (2).

Step 3. Define boundary values

The firm seeks to ensure that its salvage value at the end of the planning horizon is captured. Therefore the value function at time \( t = T \) is given by \( V(S, T, T) = mTSTe^{-\rho T} \).

Step 4. Finding a value function

The solution procedure involves finding an appropriate value function \( V(s, t, T) \) that satisfies the Hamilton–Jacobi–Bellman (HJB) equation (Fleming and Rishel 1975). We solve this two-point boundary value problem, using the methodology developed in Raman (2006), which uses the method of undetermined parameters.
Normative Analysis and Insights

Optimal Marketing Activities with Time-varying Parameters

Applying the above solution procedure, we find that the optimal marketing efforts \( u^*(t) \) and \( v^*(t) \) are given by

\[
u(t) = \frac{F(t)\beta_1(t)}{2\epsilon_1(t)}
\]

(6)

where,

\[
v(t) = \frac{F(t)\beta_2(t)}{2\epsilon_2(t)}
\]

Hereafter, we will call \( F(t) \) the finite horizon effect. It is noteworthy that a model with completely general time-varying marketing effectiveness parameters yields closed-form optimal solutions very similar in structure to the classic NA model. The optimal spending levels in Equations (6) and (7) also exhibit fundamental similarities to prior results (Naik and Raman 2003; Naik, Raman, and Winer 2005) in that they are proportional to the effectiveness parameters. Specifically, when we set \( \beta(i) = \beta_i \) and \( \epsilon(i) = \epsilon_i \) for \( i = 1, 2 \), i.e., we restrict the marketing effectiveness and cost parameters to be constant, assume \( m(t) = m \) (constant margin), and set the salvage value to zero, we retrieve the Nerlove and Arrow (1962) model solution. Hence, our solution retains the parsimonious structure of constant-parameter settings while being completely general with respect to time-varying marketing effectiveness and costs.

Examining Equation (6), we see that the optimal allocations to the two marketing activities are directly proportional to the effectiveness and inversely proportional to their costs. This would explain why companies spend on interactive activities like Google Ad Words, Facebook and Twitter even when their effectiveness is formally undocumented and unproven — namely, because they are relatively cheaper than TV.

Insights into Dynamics of Optimal Trajectories

We see from the solutions in Equations (6) and (7), that optimal marketing trajectories \( u^*(t) \) and \( v^*(t) \) can follow increasing (e.g., \( du^*/dt > 0 \)), decreasing (e.g., \( du^*/dt < 0 \)) or constant patterns (e.g., \( du^*/dt = 0 \)) over the planning horizon depending on the time-varying nature of the effectiveness parameters. Furthermore, past literature on optimal advertising has shown that the functional form of the market-response heavily influences the outcome of the optimal advertising schedule. Specifically, the choice of a concave market-response function always yields an “even” (i.e., constant level of spending) schedule at optimality (see, e.g., Mahajan and Muller 1986). However, the solutions in Equations (6) and (7) imply that if the parameters are indeed time-varying, the optimal policy’s trajectory can be different from even spending. For example, if \( \beta_k(t) \) follows a second-order polynomial function in time (e.g. Winer 1979), the optimal policy will not be even.

Moreover, even if the effectiveness of the marketing activities remains constant over time, the optimal spending trajectories are influenced by how media costs change over time, e.g. inflation trends. Specifically, spending on a marketing activity should increase (decrease) as its cost decreases (increases) over time.

The dynamics of the optimal policies we have derived may not seem surprising since they are proportional to the time-varying parameters. However, the interesting point here is that, even if the parameters were constant, the optimal policies will be time varying due to the finite horizon effect, \( F(t) \) (Equation (7)).

Insights into Finite Horizon Effect

To better understand the finite horizon effect, consider the special case where the margin and media costs are constant. In that case, the finite horizon effect \( F(t) \) reduces to the expression \( \text{FHE}(t) \), where,

\[
\text{FHE}(t) = \left( 1 - e^{-(\theta + \rho)(T-t)} \right) \left( 1 - \theta (\delta + \rho) \right).
\]

(8)

Equation (8) rewards close analysis. First, note that

\[
\lim_{T \to \infty} \text{FHE}(t) = \lim_{T \to \infty} \left( 1 - e^{-(\theta + \rho)(T-t)} \left( 1 - \theta (\delta + \rho) \right) \right) = 1.
\]

(9)

Thus, the finite-horizon effect is unity as the size of the horizon becomes infinite. Consequently, provided that the effectiveness parameters \( \beta_k(t) \), media costs \( \epsilon_k(t) \) and margin \( m(t) \) all converge to steady state levels over an infinitely long time horizon, the finite-horizon policies intuitively reduce to the infinite horizon policies.

Next, the derivative of \( \text{FHE}(t) \) with respect to time ‘t’ is

\[
\frac{\partial \text{FHE}(t)}{\partial t} = (\delta + \rho) e^{-(\theta + \rho)(T-t)} \left( 1 - \theta (\delta + \rho) \right),
\]

(10)

which is negative for \( \theta > \frac{1}{(\delta + \rho)} \), positive for \( \theta < \frac{1}{(\delta + \rho)} \), and zero when \( \theta = \frac{1}{(\delta + \rho)} \).

These expressions show the critical role played by the salvage value parameter \( \theta \) in a finite horizon budgeting and allocation problem. In the first case, the finite horizon effect decreases over time, in the second case, the finite horizon effect increases over time and in the third case, the finite horizon effect is zero, which means that all temporal variation in the optimal policies is driven purely by the temporal variation in \( \beta_k(t) \), \( \epsilon_k(t) \) and \( m(t) \).
To understand the influence of the horizon size on the finite horizon effect, consider the derivative of \( FHE(t) \) with respect to ‘\( T \)’ in Equation (11), below

\[
\frac{\partial (FHE(t))}{\partial T} = (\delta + \rho)e^{-(T-t)(\delta + \rho)}(-1 + \theta(\delta + \rho)).
\]  \hspace{1cm} \text{(11)}

Equation (11), exhibits, as expected, exactly the opposite behavior as it does with respect to time ‘\( t \)’ (Equation (10)).

Finally, the influence of the salvage value parameter \( \theta \) on the finite horizon effect is given by the derivative of \( FHE(t) \) with respect to ‘\( \theta \)’:

\[
\frac{\partial (FHE(t))}{\partial \theta} = (\delta + \rho)e^{-(T-t)(\delta + \rho)}
\]  \hspace{1cm} \text{(12)}

which is always positive. Consequently, the finite horizon effect increases with size of the salvage value parameter \( \theta \) a result that agrees with intuition because the salvage value of terminal sales \( S(T) \) is proportional to \( \theta \). Thus, when \( \theta = 0 \), the finite horizon effect is one (i.e., there is no finite horizon effect), and then, as \( \theta \) increases, the finite horizon effect steadily increases with \( \theta \).

Thus, the optimal policies in Equation (6) need not simply track the time-variation of the effectiveness parameters even though they are proportional to the latter because the finite horizon effect may be dominant. Consider, for example, the case where \( \beta_1(t) \) is increasing over time: if the decay and discount rate are large enough, the finite horizon effect may decrease at a rate greater than the rate at which \( \beta_1(t) \) increases, making it optimal to decrease \( u(t) \) over time even though its effectiveness \( \beta_1(t) \) is increasing. That is, the optimal allocations over a finite horizon are fundamentally different from those optimized over an infinite planning horizon because they are affected by the length of the horizon and need not always track the temporal patterns of the effectiveness parameters—a finding of substantive interest for managers.

*Insights into Switching Emphasis between Marketing Activities over Time*

Shifts in emphasis from one marketing variable to the other (by way of the dominant portion of the budget going toward one activity or the other) over a campaign planning horizon are observed in practice and have been documented in the marketing literature (see, e.g., Shankar 2009). For example, in the pharmaceutical industry, a combination of a “pull” marketing strategy through the use of journal advertising and a “push” marketing strategy through the use of a sales rep detailing (and growing e-Detailing) is employed (Kotler and Keller 2008). In the beginning of a new drug’s life-cycle, while there is larger uncertainty surrounding its efficacy, a pull strategy is shown to be more effective. However, as the physician learns through experience, the uncertainty about a drug’s efficacy is substantially reduced, and the effects of detailing (i.e. personal selling) are likely to be more direct and to dominate the effect of advertising (Narayanan, Manchanda, and Chintagunta 2005).

Notwithstanding such empirical evidence, normative marketing-mix guidance on the optimal planning of marketing resource allocation efforts through the length of the planning horizon is missing in the literature. Our model results indicate, first, that as a function of the time-varying effectiveness of each marketing activity, the optimal allocation ratio of marketing activities is also time-varying. Second, the resource costs per unit influence the optimal allocation ratio. Specifically, let us define \( x^*(t) = u^*(t)/v^*(t) \) to be the ratio which reflects the allocation emphasis placed on \( u \) relative to \( v \) during the planning period. The mathematical expression for \( x^*(t) \) is given as:

\[
x^*(t) = \frac{\beta_1(t)c_2(t)}{\beta_2(t)c_1(t)} \hspace{1cm} \text{(13)}
\]

Equation (13) shows that the allocation emphasis ratio does not depend on the discount rate and carry-over parameter, and is larger when \( \beta_1(t) \) is larger, when \( c_1(t) \) is smaller, \( c_2(t) \) is larger, or when \( \beta_2(t) \) is smaller. Another perspective on this follows by noting that \( \frac{\beta_1(t)c_2(t)}{\beta_2(t)c_1(t)} \geq 1 \Leftrightarrow \frac{\beta_1(t)}{\beta_2(t)} \geq \frac{c_1(t)}{c_2(t)} \). Thus, instrument 1 should be emphasized more than instrument 2 if its relative effectiveness exceeds its relative cost compared to instrument 2. Equation (13) shows that, depending upon the nature of time variation in the four parameters \( \beta_1(t) \), \( \beta_2(t) \), \( c_1(t) \), \( c_2(t) \), the emphasis on one instrument versus the other can completely reverse itself over the planning horizon. Thus, it can be optimal for an instrument heavily used in the early stage of the PLC to be de-emphasized in a later stage of the PLC, while the other instrument, used sparsely early on could later start receiving the lion’s share the total marketing budget.

In sum, our results shed light on the optimal allocation of marketing resources when marketing efforts have time-varying effectiveness. In addition, they highlight the role of changing effectiveness in determining the dominance of a variable (as defined by its allocated budget) in a planning horizon.

*Managerial Implementation of Solution*

The managerial import of Equations (6) and (7) is that in practice, managers must perform two tasks to obtain dynamically optimal marketing plans under any general time-varying structure.

First, they need to estimate a time-varying response model — which can be achieved with the Kalman filter as it overcomes a degrees of freedom problem that would arise with ordinary least squares (OLS) estimation, even with only one marketing activity with time-varying effectiveness. Specifically, considering the model intercept and effectiveness coefficient continuously change over the time for which the data is observed, the number of model parameters to be estimated in the case of OLS will exceed the number of observations. In contrast, the Kalman Filter tackles the degrees of freedom problem in the following way. First, it separates the dynamics of the setting into the evolution of unobserved states that are linked to the observed data. Second, the unobserved state
formulation can be specified to capture both deterministic aspects (e.g., an AR(1) formulation) or stochastic aspects (e.g., an MA(1) formulation) in the coefficient $\beta_i(t)$. Next, it derives the conditional likelihood of jointly observing the sequence of unobserved states and the observed data, by decomposing the joint density into the product of conditional density and marginal density. The advantage of using the conditional density in the Kalman filter is that the inter-temporal dependence in the observed data, induced by the time-varying parameters, is fully captured without severely losing degrees of freedom, which would be the case if the marginal density is used (such as in OLS). The parameters that maximize the likelihood are then obtained, followed by inference (see Naik 1999; Naik, Mantrala, and Sawyer 1998 for examples of Kalman filter based time-varying response estimation in marketing and Harvey 1994, pp 104–107 for a description of the Kalman recursions).

Once the time-varying coefficient/s have been estimated, decision-makers can compute and utilize the marketing effectiveness value at any point “t,” (e.g. if $\beta_1 = .3 + .01t$, it would be 3.1 at $t=10$), as well as the costs and the margin at each time “t,” in Equations (6) and (7) to obtain the optimal solutions with relative ease. This also enables managers to compare the optimal spending trajectory from the time-varying effectiveness case to that of the constant effectiveness case.

**Conclusion**

The daunting mathematical problem of simultaneously optimizing the firm’s profitability with respect to multiple marketing mix variables in a dynamic sales response environment makes the use of simplifying yet sub-optimal heuristics and decision rules attractive to managers (Mantrala 2002). However, even though repeated surveys suggest that managers are unlikely to directly use sophisticated optimization solutions in their decision-making, they can benefit from the directional insights and insights for effective spending offered by normative marketing models that capture the essence and complexity of their environment and decision-making process. Over time, such insights can become part of industries’ “best practices” that improve profitability. In this vein, this paper aims to contribute to both effective marketing budgeting theory and practice by deriving rules for optimal marketing mix expenditures over time that account for two real but neglected aspects of practical problems in models to date: marketing instruments’ time-varying effectiveness and costs, and the length of planning horizons.

In this paper, we developed a new analytical framework to optimize the marketing mix allocation – focusing on the two-variable case – over a finite horizon when the response function parameters are changing over time. We obtained closed-form analytical results showing that the optimal allocations are proportional to the effectiveness parameters, consistent with earlier results of Naik and Raman (2003) and Naik, Raman, and Winer (2005), but also providing the new insight that this need not always be the case because of the finite horizon effect. Further, due to the time-varying parameters, the optimal allocation ratio will also change over time, thereby directing managers to emphasize different marketing mix elements at different times over the planning horizon. We also found that the allocation ratio will switch over the planning horizon under certain conditions, causing complete reversals in the emphasis placed on one instrument versus the other. While conventional wisdom on the product lifecycle (PLC) concept recommends doing so – for example the recommendation that advertising should be emphasized over personal selling in the introductory phase while personal selling should receive greater weight later in the PLC – analytical proof that such actions are optimal is missing in the literature up to this point. We establish the precise nature – quantitatively and qualitatively – of the optimal variation in spending on different marketing instruments over time such as offline and online media. Thus, our research is a valuable complement to recent empirical research in marketing dynamics emphasizing time-varying effectiveness parameters (e.g., Leeflang et al. 2009) in that it provides normative rules for resource allocation. Managers can combine these rules with empirically derived parameter estimates to improve their marketing resource allocation.

**Directions for Future Research**

A limitation of the analysis in this paper is that it does not incorporate competition (e.g., Shankar 2009). This is primarily because solving a finite horizon competitive marketing resource allocation problem with time-varying parameters is analytically intractable, e.g., Bass et al. (2005). Earlier analyses of dynamic marketing resource allocation optimization that incorporate competition assume constant marketing effectiveness parameters, e.g., Raman and Naik (2004) and Naik, Raman, and Winer (2005). Considering that the extant literature offers very limited insights into optimal marketing resource allocation over time with time-varying parameters even for the case of a monopoly firm we focus on this problem in this paper. Our analysis also does not consider the effects of uncertainty. Incorporating competition and uncertainty in marketing models with time-varying parameters is a worthy direction for future research. However, the solutions are unlikely to be analytically tractable and will probably require numerical simulation analyses to gain managerially relevant insights.

Another interesting direction for research is to investigate dynamic marketing mix problems that involve both time-varying effectiveness parameters and synergy between the marketing instruments, thereby extending the analyses of Naik and Raman (2003) and Schultz, Block, and Raman (2009a,b, in press), all of which assumed constant effectiveness parameters. A third emerging direction for research is dynamic marketing mix optimization by platform firms that have multiple revenue sources (e.g., Sridhar et al. 2009).

**Appendix A. Analytical Derivation of Optimality Conditions**

The key player in the methodology is defined by

$$V(s,t,T) = \max_U \left[ \int_t^T e^{-\rho w} \pi(S(w), U, w) dw \right]$$

(A1)
where \( V(s, t, T) \), called the value function, denotes the maximum profit that can be achieved by using the optimal marketing mix over the remaining horizon \([t, T]\), starting from an arbitrary state ‘s’ at time \( t \). The value function \( V(s, t, T) \) satisfies the Hamilton–Jacobi–Bellman (HJB) equation (Fleming and Rishel 1975), where \( V_s = \partial V / \partial s \), and \( V_t = \partial V / \partial t \):

\[
V_t + \max_U \left[ e^{\rho t} \pi((s, U, t) + V_s f(s, U, t)) \right] = 0. \tag{A2}
\]

The function \( f(s, U, t) \) is defined as: \(-\delta S + \beta_1(t)u(t) + \beta_2(t)v(t)\).

In our problem, the vector \( U \) consists of the two decision variables \( -U = \{u(t), v(t)\} \) and the instantaneous profit is

\[
\pi(t) = m(t)s - c_1(t)u^2 - c_2(t)v^2 \tag{A3}
\]

where \( S(t) \) is the instantaneous sales rate at time “t.” The function \( f(s, U, t) \) is defined by the right-hand side of Equation (3) in our paper. We obtain the necessary conditions for dynamic optimality by differentiating the maximand in Bellman’s equation with respect to \( U \) and equating the gradient vector to zero, i.e.

\[
e^{-\rho t} \pi V_U + V_s f_U = 0. \tag{A4}
\]

This yields the system of equations below.

First Order Condition For \( u \):

\[-2c_1(t)e^{\rho t}u + \beta_1(t)V_s = 0 \tag{A5}\]

First Order Condition For \( v \):

\[-2c_2(t)e^{\rho t}v + \beta_2(t)V_s = 0 \tag{A5}\]

We conclude from the above that

\[
u = e^{\rho t} \beta_2(t) V_s / 2c_2(t) \tag{A7}\]

\[
v = e^{\rho t} \beta_2(t) V_s / 2c_2(t) \tag{A7}\]

Substituting these expressions for \( u \) and \( v \) into the HJB equation yields the following second-order nonlinear partial differential equation for our problem.

\[
V_t + \frac{1}{4} \left[ 4e^{\rho t}m(t)s - 4\delta s V_s + \frac{(c_1(t)\beta_2^2 + c_2(t)\beta_1^2)e^{\rho t}V_s^2}{c_1(t)c_2(t)} \right] = 0. \tag{A8}
\]

The above nonlinear partial differential equation must be solved subject to the initial condition \( x = x_0 \) and the boundary conditions at the terminal time \( t=T \), given by the salvage value at the end of the planning horizon: \( V(S, T, T) = mS0e^{-\rho T} \). We next solve this two-point boundary value problem, using the methodology developed in Raman (2006).

Using the method of undetermined parameters, we conjecture a polynomial (in \( s \)) solution with time-varying coefficients to the above PDE. After experimenting with different orders of the polynomial, we find that a quadratic does the job. Consequently, we conjecture the following solution for \( V(s, t, T) \).

\[
V(s, t, T) = e^{-\rho t} \left[ s^2 k_1(t) + sk_2(t) + k_3(t) \right] \tag{A9}
\]

Next we compute the partial derivatives \( V_t \) and \( V_s \) under the above conjecture, put the resulting expressions back into Equation (A8). Substituting the quadratic conjecture and the partial derivatives \( V_t \) and \( V_s \) from the previous step into the PDE yields a set of coupled nonlinear ordinary differential equations for the unknown coefficient functions \( k_1(t), k_2(t), \) and \( k_3(t) \) of the conjectured quadratic. While the equations determining function \( k_1(t), k_2(t), \) and \( k_3(t) \) are nonlinear and coupled, they are ordinary – not partial – differential equations.
We simultaneously solve these differential equations subject to the boundary conditions on $V(s, t, T)$ at $t=T$. The system of equations we solve is shown below:

\[k_1(t) = 0\]
\[4c_1(t)c_2(t) \frac{dk_2(t)}{dt} - 4(\rho + \lambda)c_1(t)c_2(t)k_2(t) + 4c_1(t)c_2(t)m(t) = 0\]

\[4c_1(t)c_2(t) \frac{dk_3(t)}{dt} - 4\rho c_1(t)c_2(t)k_3(t) + (c_1(t)\beta_2(t) + c_2(t)\beta_1(t))k_2^2(t) = 0\]

\[
\left( e^{(b+\rho)t} \int_0^1 e^{(b+\rho)s}m(s)ds - e^{(b+\rho)t} e^{(b+\rho)T} \int_0^1 e^{(b+\rho)s}sm(s)ds T \right) \beta_1^2(t)c_2(t) \\
\left( e^{(b+\rho)t} \int_0^1 e^{(b+\rho)s}m(s)ds - e^{(b+\rho)t} e^{(b+\rho)T} \int_0^1 e^{(b+\rho)s}sm(s)ds T \right) \beta_2^2(t)c_2(t) \\
-4\rho k_3(t)c_1(t)c_2(t) + 4c_1(t)c_2(t)k_3'(t) = 0 \\
V(s, t, T) = e^{-\nu T} k_3(t) + e^{(b+\rho)T} \left( e^{(b+\rho)t} \int_0^1 e^{(b+\rho)s}m(s)ds - e^{(b+\rho)t} e^{(b+\rho)T} \right) \left( e^{(b+\rho)t} \int_0^1 e^{(b+\rho)s}sm(s)ds T \right) \right)
\]

Substituting the functional forms for $k_1(t), k_2(t)$, and $k_3(t)$ back into Equation (A9), we finally obtain the solution to the PDE. For brevity, we omit this in the Appendix. We checked that the solution satisfies the boundary conditions. The mathematical derivations are tedious and were implemented in Mathematica. Interested readers may obtain the rest of the details of the optimization algorithm from the authors. To specify $V(s, t, T)$ completely, we need to find $k_3(t)$. This is done as follows — $k_3(t)$ satisfies the ODE:

\[
\left( e^{(b+\rho)t} \int_0^1 e^{(b+\rho)s}m(s)ds - e^{(b+\rho)t} e^{(b+\rho)T} \right) \beta_1^2(t)c_2(t) \\
\left( e^{(b+\rho)t} \int_0^1 e^{(b+\rho)s}m(s)ds - e^{(b+\rho)t} e^{(b+\rho)T} \right) \beta_2^2(t)c_2(t) \\
-4\rho k_3(t)c_1(t)c_2(t) + 4c_1(t)c_2(t)k_3'(t) = 0
\]

In the above, $k_3(t)$ denotes $\frac{dk_3(t)}{dt}$. For any given, arbitrarily specified continuous time-varying functions $m(t), c_1(t), c_2(t), \beta_1(t), \beta_2(t)$, we have the complete solution to our model. We simply plug into the functions into the above linear ODE for $k_3(t)$, which always has an integrating factor, and therefore, although the solution may be algebraically complicated, it will always exist and can be found as shown below. We may rewrite the above ODE as follows:

\[k_3'(t) - \rho k_3(t) + G(t) = 0\]

where,

\[G(t) = \frac{\left( e^{(b+\rho)t} \int_0^1 e^{(b+\rho)s}m(s)ds - e^{(b+\rho)t} e^{(b+\rho)T} \right) \beta_1^2(t)}{4c_2(t)} + \frac{\left( e^{(b+\rho)t} \int_0^1 e^{(b+\rho)s}m(s)ds - e^{(b+\rho)t} e^{(b+\rho)T} \right) \beta_2^2(t)}{4c_1(t)}
\]

The solution to the ODE is then immediately obvious by inspection—since it always has an integrating factor, $k_3(t)$ is given by:

\[k_3(t) = \int_0^t e^{-\rho s} G(s) ds, \quad \text{where } G(t) \text{ is given by :}
\]

\[G(t) = \frac{\left( e^{(b+\rho)t} \int_0^1 e^{(b+\rho)s}m(s)ds - e^{(b+\rho)t} e^{(b+\rho)T} \right) \beta_1^2(t)}{4c_2(t)} + \frac{\left( e^{(b+\rho)t} \int_0^1 e^{(b+\rho)s}m(s)ds - e^{(b+\rho)t} e^{(b+\rho)T} \right) \beta_2^2(t)}{4c_1(t)}
\]

From the solution $V(s, t, T)$, we compute the optimal controls $u$ and $v$ as follows: find the shadow price $V_s$ from the solution, and substitute it into the expressions for the optimal controls:

\[u = e^{\theta T} s_1(t) V_s / 2c_1(t), \quad \text{and} \]

\[v = e^{\theta T} s_2(t) V_s / 2c_2(t).
\]

This yields the expressions $u^*$ and $v^*$ shown in the paper.
Appendix B. Extension to Multiple Activities

The Value function $V(s, t, T)$, defined as before, denotes the maximum profit that can be achieved by using the optimal marketing mix with 'n' instruments over the remaining horizon $[t, T]$, starting from an arbitrary state 's' at time $t$. The value function $V(s, t, T)$ satisfies the Hamilton–Jacobi–Bellman (HJB) equation (Fleming and Rishel 1975), where $V_t = \partial V/\partial s$, and $V_s = \partial V/\partial t$:

$$V_t + \max_U \{e^{-\rho t} \pi((s, U, t) + V_f(s, U, t))\} = 0. \quad (B1)$$

The function $f(s, U, t)$ is defined as: $-\partial S + \sum_{i=1}^{i=n} \beta_i(t) u_i(t).$

where $\frac{ds}{dt} = -\partial S + \sum_{i=1}^{i=n} \beta_i(t) u_i(t).$ In our problem, the vector $U$ consists of the 'n' decision variables $U = (u_1(t), u_2(t), ..., u_n(t))$ – and the instantaneous profit is $\pi(t) = m(t)s - \sum_{i=1}^{i=n} c_i(t)u_i^2$.

$$\pi(t) = m(t)s - \sum_{i=1}^{i=n} c_i(t)u_i^2 \quad (B2)$$

where $S(t)=s$ is the instantaneous sales rate at time “t.” The function $f(s, U, t)$ is defined by the right-hand side of Equation (3) in our paper. We obtain the necessary conditions for dynamic optimality by differentiating the maximand in Bellman’s equation with respect to $U$ and equating the gradient vector to zero, i.e.

$$e^{-\rho t} \pi V_U + V_f f_U = 0 \quad (B3)$$

This yields the system of ‘n’ equations, $1 \leq i \leq n$, below.

First Order Condition For $u_i : -2c_i(t)e^{-\rho t}u_i + \beta_i(t)V_s = 0. \quad (B4)$

We conclude from the above that

$$u_i = e^{\rho t}\beta_i(t)V_s / 2c_i(t). \quad (B5)$$

Substituting these expressions for $u_i$, $1 \leq i \leq n$, into the HJB equation yields the following second-order nonlinear partial differential equation for our problem.

$$V_t + \frac{1}{4} \left[4e^{-\rho t}m(t)s^2 - 4\delta s V_s + \left(\sum_{i=1}^{i=n} \frac{\beta_i(t)^2}{c_i(t)} \right) e^{\rho t} p_s^2 \right] = 0 \quad (B6)$$

In Equation (B6), the notation $\prod_{j=1}^{j=1} c_j$ denotes the product $c_1c_2...c_n$ the notation $\prod_{j=1}^{j=n} c_j$ denotes the product $c_1c_2...c_n$ in which all the $c_j$ appear except that for which $j=f$. Equation (B6) may be simplified to yield Equation (B7) in which the mathematical structure of the generalization to 'n' activities is made immediately transparent:

$$V_t + \frac{1}{4} \left[4e^{-\rho t}m(t)s^2 - 4\delta s V_s + \left(\sum_{i=1}^{i=n} \frac{\beta_i(t)^2}{c_i(t)} \right) e^{\rho t} p_s^2 \right] = 0. \quad (B7)$$

The above nonlinear partial differential equation must be solved subject to the initial condition $s=x_0$ and the boundary conditions at the terminal time $t=T$, given by the salvage value at the end of the planning horizon: $V(S, T, T) = mS\theta e^{-\rho T}$. We next solve this two-point boundary value problem, using the methodology developed in Raman (2006).

Using the method of undetermined parameters, we conjecture a polynomial (in $s$) solution with time-varying coefficients to the above PDE. After experimenting with different orders of the polynomial, we find that a quadratic does the job. Consequently, we conjecture the following solution for $V(s, t, T)$.

$$V(s, t, T) = e^{-\rho t} \left[ s^2k_1(t) + sk_2(t) + k_3(t) \right] \quad (B8)$$

Next we compute the partial derivatives $V_t$ and $V_s$ under the above conjecture, put the resulting expressions back into Equation (B7). Substituting the quadratic conjecture and the partial derivatives $V_t$ and $V_s$ from the previous step into the PDE yields a set of coupled nonlinear ordinary differential equations for the unknown coefficient functions $k_1(t), k_2(t),$ and $k_3(t)$ of the conjectured quadratic. While the equations determining function $k_1(t), k_2(t),$ and $k_3(t)$ are nonlinear and coupled, they are ordinary – not partial – differential equations.

Simultaneously solving these differential equations subject to the boundary conditions on $V(s, t, T)$ at $t=T$ yields the optimal controls $u_1(t), u_2(t), ..., u_n(t).$ We do not pursue this last step in this paper—the purpose of this appendix was to derive the optimality equation for the extension to dynamically optimal resource allocation for 'n' instruments.

References


