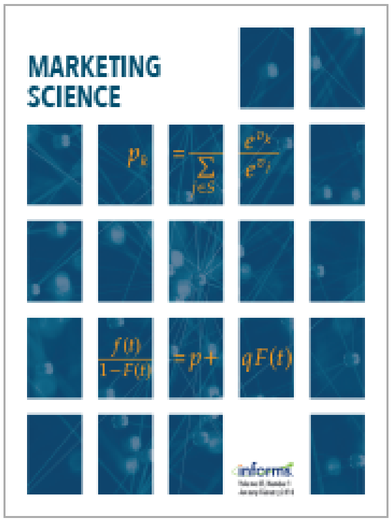


Pooling and Dynamic Forgetting Effects in Multitheme Advertising: Tracking the Advertising Sales Relationship with Particle Filters

Norris I. Bruce,

MARKETING SCIENCE
Vol. 27, No. 4, July–August 2008, pp. 659–673
ISSN 0732-2399 | EISSN 1526-548X | 08 | 2704 | 0659

DOI 10.1287/mksc.1070.0317

Pooling and Dynamic Forgetting Effects in Multitheme Advertising: Tracking the Advertising Sales Relationship with Particle Filters

Norris I. Bruce
School of Management, The University of Texas at Dallas, Richardson, Texas 75083,
norris.bruce@utdallas.edu

Firms often use a pool or series of advertising themes in their campaigns. Thus, for example, a firm may employ some of its advertising to promote price-related themes or messages and other of its advertising to promote product-related themes. This study examines the interdependence that can occur between pairs of themes in a pool (i.e., *pooling effects*), the impact of these pooling effects on the allocation of advertising expenditures, and the factors that can affect forgetting rates (or, conversely, carry-over rates) in a multitheme advertising environment. The study measures pooling, wear out, and forgetting (carry-over) effects for a campaign that uses five different advertising themes. To obtain these measures, I extend the linear Nerlove-Arrow (NA) (1962) model to a nonlinear model of advertising theme quality and goodwill and estimate the extended model using Markov chain Monte Carlo (MCMC) and particle filtering ideas. Particle filtering belongs to a class of sequential Monte Carlo (SMC) methods designed to estimate nonlinear/nonnormal state space models. Results show that forgetting (or carry-over) rates may be time varying and a function of prior goodwill (past advertising) and other advertising variables. Results show, moreover, that pooling effects can reduce theme wear out and, in turn, significantly improve advertising efficiency.

Key words: nonlinear state space model; particle filtering/smoothing; sequential Monte Carlo (SMC); sequential importance resampling (SIR); Markov chain Monte Carlo (MCMC); metropolis hastings; aggregate advertising models; pooling effects; forgetting effects
History: This paper was received August 30, 2006, and was with the author 4 months for 2 revisions; processed by Gary M. Erickson. Published online in *Articles in Advance* May 15, 2008.

1. Introduction

Firms often employ a pool or variety of advertising themes (for example, price and product ad themes) in their campaigns (Rossiter and Percy 1997). Although airing such themes concurrently can reduce theme effectiveness or quality because of interference (Kent 1991, Kent and Allen 1993, Kumar 2000, Kumar and Krishnan 2004), ad themes in a pool could, in fact, improve each other's quality if, for example, the firm alternates the themes that are aired over time (Grass and Wallace 1969). Thus, if price and product ads in a pool are aired alternately, the quality of the price ad could plausibly improve the quality of the product ad and vice versa. This interdependence between pairs of themes (or "pooling effects") could in principle forestall advertising quality wear out (Naik et al. 1998), and, in turn, improve ad theme efficiency.

Yet, ad efficiency is not only the result of wear out dynamics; it is more generally the result of the carry-over effects from past advertising. Thus carry-over effects, their magnitudes, and rates of decay over time have been the subject of numerous econometric studies (Bass and Clarke 1972, Clarke 1976, Hanssens et al. 1990). These studies traditionally employed linear distributed lag models in which the carry over rate (or, conversely, the forgetting rate) was a constant parameter, unaffected by changing factors in the advertising environment. Yet, despite the longevity of this econometric assumption, it appears that forgetting rates could indeed be time varying (Gensch and Welam 1973, Blattberg and Jeuland 1981), with past advertising (prior goodwill), competitive advertising, and copy theme all affecting the rates at which advertising decays (Ray et al. 1971) over time. It seems reasonable then to suggest that multitheme studies should also consider these dynamic features of advertising decay.

This study thus explores pooling effects and factors that could affect the rates at which advertising decays over time in a multitheme context. As a result, the substantive questions relevant to this study are as follows: How variable are the carry-over effects of different themes? How do goodwill, the result of past ad spending, and ad themes affect forgetting rates? In addition, prior research posits that stopping an ad theme or varying the executions of the same basic ad theme can forestall its wear out (Naik et al. 1998, Grass and Wallace 1969). What are, however, the differential effects of pooling and varying ad themes on the quality of each theme? Finally, what is the impact of pooling effects on the allocation of resources across themes?

This work builds on and extends the recent effort of Bass et al. (2007). These researchers proposed a linear model that linked demand to weekly ad spending, incurred in airing different themes. Their study extended (Naik et al. 1998) to measure the wear-out effects of different themes, and the implications of wear out on a firm's advertising budget. Their model, estimated using Kalman filtering and Markov chain Monte Carlo (MCMC) ideas, found that airing themes concurrently resulted in negative effects (Kent 1991, Kent and Allen 1993). The Bass et al. (2007) study, though, has several features that limit its generality. First, it models the advertising forgetting rate as a constant parameter, probably biasing the carry-over effects of past advertising on current demand. Second, it ignores the significant benefits of *pooling effects* as previously described. Third, while it is interesting to discover that airing themes concurrently could lead to negative effects, this result alone seems inadequate to explain why advertisers would pool themes in the first place. In contrast to Bass et al. (2007), this study (see Table 1) estimates a flexible model of forgetting effects, and captures, along with wear out, the effects ad themes in a pool may have on each other's quality (i.e., *pooling effects*). Furthermore, the response model in this study is both dynamic and nonlinear, and therefore it is estimated using a combination of particle filtering and MCMC procedures (cf. Doucet et al. 2001, Ristic et al. 2004, Liu and Chen 1998). Particle filtering is a flexible Bayesian inference method used to estimate nonlinear/nonnormal dynamic systems.

In summary, this paper contributes to the advertising response literature in several ways. First, it introduces a dynamic model to investigate several factors that could affect the rate at which advertising decays over time, and the magnitude of carry-over effects for different themes. Although analytical and experimental studies have considered some of these factors, to the best of my knowledge, they have not been investigated empirically either in a model that links advertising expenditures to sales, or in a multitheme context. Second, the paper examines the benefits that could arise because of the pooling and varying of multiple themes, and it considers the impact of these pooling effects on the firm's ad budgeting decisions. Evidence supporting the use of theme variation to diminish or forestall wear out emerges from several prior experimental studies (e.g., Schumann et al. 1990). Still, this paper is probably one of the first to consider pooling and theme variation in a market-based model of sales. Third, to estimate the resulting model, the paper introduces a Bayesian nonlinear/nonnormal filter, the particle filter. The method generalizes the Kalman filter/smoother algorithm for Bayesian dynamic linear models (DLMs) (Van Heerde et al. 2004, Liechty et al. 2005), and may open up opportunities for ad tracking, diffusion, and other marketing studies where microdynamics and nonlinear features might be important (e.g., eye tracking studies, Pieters and Wedel 2000, Pieters et al. 1999). Results show that advertising may be more efficient in the presence of pooling effects; and pooling effects may help reduce the negative effects of wear out and airing themes concurrently, providing some justification for why firms would pool messages. Results show, too, that advertising decay might be not only a function of when an ad is shown (Thaivanich et al. 2000) but also a function of prior goodwill and other advertising variables. This, in turn, implies that there might not be a single decay function for past advertising (see Clarke 1976), but a family of such functions (see Ray et al. 1971, p. 14). Finally, the proposed model is shown to be superior to Bass et al. (2007) in terms of numerical measures of fit and forecast performance.

The remainder of the paper unfolds as follows. Section 2 provides a brief review of the relevant streams in the advertising content and dynamic effects literatures. Section 3 develops the theoretical model. Section 4 presents details of the econometric model and the data. The data come from a major telecom provider, a monopolist in its category. The firm classifies its advertising into five themes: price offer ads, product offer ads, call stimulation ads, reconnect ads, and reassurance ads. Subsequently, §5 presents the results of the estimation and the implications of these results on theme carry over and ad budget decisions. The paper concludes with an overview of the findings, managerial implications, and limitations of the study.

Table 1 **Incremental Contributions**

Features	Naik et al. (1998)	Bass et al. (2007)	This study
Theme wear out	Single	Multiple	Multiple
Dependent measure	Awareness	Sales	Sales
Forgetting/carry-over rate	Constant	Constant	Time varying: goodwill, theme, competition
Pooling effects	No	No	Yes
State equations	Linear	Linear	Nonlinear
Estimation procedure	Kalman filter	Kalman filter	Particle filter

2. Literature Review

This study draws on several streams within the advertising content and dynamic effects literature that have investigated: (i) quality dynamics and theme effects,

(ii) time-varying forgetting or carry-over rates, and (iii) theme variation. What follows is a brief review of these relevant streams. For an extensive review, the reader should see, for example, Little (1979), Vakratsas and Ambler (1999), Tellis (2004), and Vakratsas (2005), and the references therein.

2.1. Quality Dynamics and Theme Effects

Several prior studies support the view that advertising effectiveness or quality should vary over time and across advertising themes. For example, Little (1979) argued that ad effects should vary because of changes in media and copy. Naik et al. (1998), however, in a definitive study, demonstrated that ad quality could regenerate during periods of no advertising because of forgetting, and could decline during periods of continuous advertising because of wear out (Grass and Wallace 1969, Greenberg and Suttoni 1973, Pechmann and Stewart 1990). These researchers (Naik et al. 1998) modeled two sources of wear out: copy wear out, which occurs because of the passage of time, regardless of the level of ad spending; and repetition wear out, which occurs because of the excessive levels of ad spending. Bass et al. (2007) extended Naik et al. (1998) to account for the wear-out effects and quality of different advertising themes. In a model of constant forgetting rate and no pooling effects, they found that airing themes concurrently resulted in negative effects, perhaps because of clutter; and support for the premise that emotional ad themes wear out faster than nonemotional themes (MacInnis et al. 2002). Similarly, in a market-level study of the effects of various ad appeals, Chandy et al. (2001) found that argument-based appeals (or themes) were more effective in new markets, whereas emotion-based appeals were more effective in mature markets.

2.2. Time-Varying Forgetting Rates

Econometric models of advertising have traditionally assumed that advertising has two effects on demand: an instantaneous effect because of current expenditures, and a carry-over effect because of prior expenditures or goodwill.[1] Moreover, the effects of prior expenditures (i.e., carry over), the result of delayed response or repeat behavior (Tellis et al. 2000), were often discounted at some constant forgetting (carry-over) rate (cf. Bass and Clarke 1972, Clarke 1976, Hanssens et al. 1990).[2] While a constant rate might simplify model estimation, such a rate is unrelated to changing factors in the advertising environment, and thus is probably unrealistic (Bultez and Naert 1979, Pechmann and Stewart 1990, Chessa and Murre 2007). For example, Ray et al. (1971), drawing on several experimental studies, concluded that copy theme, brand awareness, and competitive advertising could all shape advertising decay. In addition, Gensch and Welam (1973) showed analytically that a fixed carry-over rate could be inconsistent with learning and forgetting theory (Chessa and Murre 2007). That is, using a Nerlove-Arrow-type (1962) formulation, their study demonstrated that the effect on demand in any given period, of advertising in the same period, might be independent of the levels of advertising in all prior periods. In a solution to this inconsistency, the authors modeled carry over as a function of prior goodwill, with the duration of carry over influenced by the intensity of competitive advertising. The carry-over (forgetting) rate was hypothesized to decrease (increase) with competition, and increase (decrease) with goodwill but at a decreasing (increasing) rate. Similarly, Blattberg and Jeuland (1981) proposed a micromodel, and aggregated across consumers and over time to produce a macromodel of the ad-sales relationship. The resulting model found analytically that the carry-over rate was not constant, but varied nonlinearly over time with prior ad spending, as hypothesized earlier in Gensch and Welam (1973).

This hypothesized impact of prior goodwill on the rate of advertising decay is a particularly appealing model feature. It suggests that the probability of carry over effects (from repeat or delayed purchases) increases with cumulative prior ad spending, but at a diminishing rate. This is consistent with an extensive body of advertising research that catalogs the positive but diminishing effects of ad repetition (see, for example, Belch 1982, Cacioppo and Petty 1979).

2.3. Pooling and Varying Ad Themes

In an environment in which firms air pools of ad themes in their campaigns, it is conceivable that some themes could affect each other. For example, airing ad themes concurrently could perhaps increase interference or clutter (Kent 1991, Kent and Allen 1993, Kumar 2000, Kumar and Krishnan 2004), which could, in turn, diminish ad quality. Yet, varying the pool of ads presented for a brand could increase overall effectiveness of the individual pieces of advertising (cf. Grass and Wallace 1969, Rossiter and Percy 1997). Thus, pooling and varying themes may improve the quality of each theme; this suggests (econometrically) a synergistic effect or interdependence among themes, which, in turn, could forestall their wear out (Naik et al. 1998). The theoretical mechanisms hypothesized to explain the effects of ad variation on wear out are available elsewhere in the literature (cf. Burke

[1] See, for example, Thaivanich et al. (2000) and Pechmann and Stewart (1990) for the behavioral justification for this assumption.

[2] The single exponential decay function implied by these models (and similar linear distributed lag forms, e.g., Mitra and Golder 2006) varies with the inclusion of additional lagged values in dependent measure, and additional lagged advertising (independent) terms (Clarke 1976). However, the carry-over parameters remain constant, unrelated to changing advertising variables.

and Srull 1988, Kumar 2000, Kumar and Krishnan 2004, Unnava and Burnkrant 1991), and therefore they are not repeated here. Still, practitioners often use two variation strategies (Schumann and Clemons 1989): cosmetic variations, which refer to "peripheral" changes to the ad, such as changes in style, format, or execution; and substantive variations, which refer to changes to the ad, such as changes in the theme or copy found within the ad.

To the best of my knowledge, there are no market studies on the effectiveness of the above variation strategies. In two laboratory experiments, however, Schumann et al. (1990) found that cosmetic variation had a greater impact on attitude when motivation to process the ad was low; and conversely, that substantive variation was more influential when motivation to process the ad was high. For ad variation to reduce wear out, the study proposed, consumers must pay greater attention and have more motivation to process advertising. The greater attention paid to emotional ads, as well as their slower wear out (Pechmann and Stewart 1990), is also the reason why emotional ads are more effective in mature market (MacInnis et al. 2002, Chandy et al. 2001). This suggests that in mature markets, *substantive variation*, which requires a greater level of attention, is probably more effective for emotional ads than for rational ads.

3. Model Development

This section develops a model to account for time-varying forgetting effects and the pooling effects of themes. The model, a generalization of the model by Bass et al. (2007), links demand to weekly ad expenditures incurred in airing different themes. To begin, therefore, I update the discrete-time Nerlove-Arrow (NA) (1962) model in Bass et al. (2007) with a time-varying forgetting rate δ_t:

$$\Delta G_t = \sum_{i=1}^{n} q_{it-1}\left(g(A_{it}) + \lambda_i \sum_{\substack{j=1\\ j\neq i}}^{n} h(A_{it}, A_{jt})\right) - \delta_t G_{t-1}, \quad (1)$$

where

$g(A_{it}) =$ a function of the ad expenditures (A_{it}) on the ith theme at time, t.[3]

$q_{it-1} =$ the quality of ith theme at time, $t-1$.

$\lambda_i \sum h(A_{it}, A_{jt}) =$ an overall measure of how the ith theme interacts with all other concurrently aired themes at time, t.[4]

$G_t =$ goodwill at time, t.

$\delta_t =$ forgetting rate at time, t.

$d_t = 1 - \delta_t$, carry-over rate at time, t.

Thus, Equation (1) assumes that goodwill G_t decays in proportion to prior goodwill G_{t-1} and is sustained by an additive function of the advertiser's expenditures $g(A_{it})$ on each theme i per week (t). The effect of this expenditure $g(A_{it})$ is moderated by an overall measure ($\lambda_i \sum h(A_{it}, A_{jt})$, negative in Bass et al. 2007) of how theme i interacts with all concurrently aired themes in week, t. In each week, furthermore, the advertiser is assumed to vary the pool of themes aired. For example, over three weeks, the advertiser may air both price and product ads in the first week while airing only price ads in the second week and perhaps only product ads in the third. This theme (substantive) variation, prior research suggests (Schumann et al. 1990), might forestall quality wear out, and mitigate the effects of interference, which could arise when themes are aired concurrently.

Next, to study factors that might explain advertising decay (Gensch and Welam 1973, Ray et al. 1971, Blattberg and Jeuland 1981), the forgetting rate (δ_t) is modeled as a function of prior goodwill and advertising variables: $F(G_{t-1}, \mathbf{Z}_{1t}, \boldsymbol{\eta})$, where the vector of variables $\mathbf{Z}_{1t} = \{$*Copy Theme* and *Competitive Ad Spending*$\}$; G_{t-1} is prior goodwill; and $\boldsymbol{\eta}$ is a vector of parameters to be estimated. To capture the effects of diminishing returns with respect to prior goodwill, a quadratic term G_{t-1}^2 is also included in the function. The variable *copy theme*, the cumulative percent of nonprice and nonproduct offer ads, controls for the likelihood that, cumulatively, ads with more emotional themes (Belch and Belch 1998, pp. 267–271) have a significant effect on δ_t. The different effects of emotional and rational themes on wear out are well known (MacInnis et al. 2002, Chandy et al. 2001). Notably, δ_t is now a function of the variable *competitive ad spending*; therefore, ad quality wear out and regeneration are directly linked to the intensity of competitive advertising.[5] Finally, because δ_t is between zero and one, I use the generalized logistic (GL) transformation (Prentice 1975, Carlin and Louis 2000):

$$\delta_t = \left\{1 + \exp(-F(G_{t-1}, \mathbf{Z}_{1t}, \boldsymbol{\eta}))\right\}^{-m}, \quad (2)$$

where $m > 0$ is a shape parameter to be estimated.

The GL model has been widely applied in dose response studies, where variability in the data cannot be explained by the standard logistic model (i.e., when $m = 1$). In this study, the GL model has the appealing property of allowing the shape parameter (m) of the forgetting rate to be estimated from the data.

[3] The estimation uses a semilog transformation: $g(A_{it}) = \ln(1 + A_{it})$ and $h(A_{it}, A_{jt}) = \ln(1 + A_{it})\ln(1 + A_{jt})$. See Bass et al. (2007) for the justifications for these functional forms, and the choice of the NA (1962) model.

[4] Because advertising effects persist beyond a week, it is presumed that multiple themes aired in a week are running concurrently.

[5] I am grateful to an anonymous reviewer for suggesting this treatment of competitive advertising.

As discussed in the literature review, the quality (q_{it}) of an ad theme can vary over time because of three factors: wear out, forgetting, and pooling effects. The following deceptively simple extension of Bass et al. (2007) captures this quality dynamic in difference equation form:

$$\Delta q_{it} = \underbrace{-a(A_{it})q_{it-1} + b(A_{it})(1-q_{it-1})}_{\text{Own Theme Effects}} + \underbrace{\sum_{\substack{j\neq i\\ j\in P_t}} \phi_{ij} q_{jt-1}}_{\text{Pooling Effects}}, \quad (3)$$

for $i = 1, 2, \ldots, n$ ad themes, where P_t is the pool of ad themes that are aired in period t, and

$$a(A_{it}) = c_i + r_i A_{it}, \quad (4)$$

$$b(A_{it}) = \delta_t(1 - I(A_{it})). \quad (5)$$

Specifically, Equations (3)–(5) assume that changes in the quality of ad theme i are a function of the repetition ($r_i A_{it}$) and copy (c_i) wear outs,[6] regeneration because of forgetting effects when the theme is not repeated $b(A_{it})$, and pooling effect φ_{ij} with ad theme $j \in P_t$, $j \neq i$. In Equation (3), wear out and forgetting are own theme effects, while pooling effects depend on the quality and the varying mixture of themes in pool P_t. In general, the pooling effects parameter ϕ_{ij} may be negative or positive: if ϕ_{ij} is negative, then ad theme j could help diminish the quality of theme i; and if ϕ_{ij} is positive, then ad theme j could help forestall wear out of theme i. Similarly, when advertising is *off* (i.e., indicator function, $I(A_{it}) = 0$), there is a rejuvenating effect of advertising that is captured by forgetting (Naik et al. 1998). Finally, when advertising is *on* (i.e., indicator function, $I(A_{it}) = 1$), ad quality changes are now influenced by the negative wear-out effect $-a(A_{it})$ and pooling effects.

Thus, Equations (1)–(5) describe a model created to investigate the pooling effects of multiple themes and the factors that might affect ad decay rates in a multitheme context. Contrary to assumptions in prior econometric studies (Bass and Clarke 1972, Clarke 1976), forgetting (or carry-over) rates are assumed to depend on the build-up of goodwill and other advertising variables. With this model, therefore, decision makers could track how carry-over rates vary with advertising variables and modify their strategies accordingly. The model, furthermore, could reveal synergies that might exist between pairs of themes; this knowledge, in turn, could suggest opportunities to improve theme efficiency (Naik and Raman 2003, Raman and Naik 2004, Naik et al. 2005). Lastly, the resulting model, also an extension of the classic NA model, is nonlinear and dynamic in the goodwill and quality state variables. This raises several new challenges for the estimation procedure that follows.

Table 2 Descriptive Statistics of Main Variables

	Mean GRPs	Standard deviation
Call stimulation ads	100.96	128.59
Price offer ads	31.68	73.21
Product offer ads	63.87	107.77
Reconnect ads	45.62	61.33
Reassurance ads	21.25	44.01
Competition ads	505.82	
Mean call volume (hours)	16.42 million	
Mean weighted price per minute	2.46	
Mean line capacity (number of lines)	22.63 million	

4. Empirical Analysis

This section presents the empirical analysis of the preceding theoretical model, reviewing the data, empirical model, and estimation procedure employed.

4.1. Data

The data for this study come from a major telecommunications services firm. The firm, though it faces competition in the new wireless markets, is a monopolist in the landline telecommunications business.[7] Demand data for residential telephone services are available in terms of total call time in millions of hours, aggregated across three call types: local, regional, and national. Other measures—weighted average price per minute, number of residential lines, competitive advertising—are also available to help explain demand variations over time. I aggregate all competitive advertising expenditures into a single measure. The firm, however, disaggregates its advertising into five themes: call stimulation ads, product offer ads, price offer ads, reconnect ads, and reassurance ads. Advertising expenditures for these ad themes and competitive advertising, measured in gross rating points (GRP), are available for a period of 114 weeks, during the years 1995–1997. Summary statistics are given in Tables 2 and 3.

4.2. Empirical Model

The empirical model is completed by first linking goodwill to the demand for telephone services, and, then, recasting the demand, goodwill, and ad quality equations into a nonlinear state-space formulation. The state-space approach is convenient for handling multivariate data as well as nonlinear/nonGaussian dynamic systems, and provides significant advantages

[6] Recall copy wear out occurs because of the passage of time, regardless of the frequency of advertising. In contrast, repetition wear out is a consequence of excessive frequency (Naik et al. 1998).

[7] The residential market naturally faces some competition from mobile providers because some consumer calls are substitutable between fixed and wireless providers.

Table 3 Correlation between GRPs of Different Themes

	Price ads	Product ads	Reconnect ads	Reassurance ads	Competitive ads
Call stimulation ads	−0.137	−0.242**	−0.269**	−0.014	−0.020
Price ads		−0.038	0.037	−0.157*	−0.350*
Product ads			−0.029	−0.175*	−0.024
Reconnect ads				0.0001	0.186*
Reassurance ads					0.119

*Significant at $p < 0.10$, **significant at $p < 0.01$.

over traditional time series techniques (see, for example, Durbin and Koopman 2001, pp. 51–53).

Specifically, demand for residential telephone services (y_t), at time t is modeled as a function of goodwill (G_t), a vector $\mathbf{X}_t$ with variables price and number of lines; and mean zero normally distributed error, v_t.

$$y_t = G_t + \mathbf{X}_t\boldsymbol{\mu} + v_t, \quad \text{where } v_t \sim \mathrm{N}(0, \sigma_v^2). \tag{6}$$

Price and advertising are two potential sources of endogeneity. Price endogeneity could arise because of omitted variables, and advertising endogeneity could arise if the firm chooses ad expenditures strategically. I address price endogeneity with standard instrumental variable methods,[8] using the following instruments: retail price index, number of households, consumer sentiment, household spending, competitive advertising, and number of lines. Addressing advertising endogeneity, however, poses two significant modeling challenges: (i) the coefficient (q_{it}) of the endogenous variable, advertising is time-varying and (ii) ad expenditures are observed only when the firm advertises. Both challenges can be overcome (see, for example, Kim 2006) once good instruments are available. However, I leave the study of endogeneity bias in advertising as an issue for a future study, acknowledging that this is a potential limitation of the present study.

Given the above demand model, it is now convenient to recast Equations (1)–(6) into a more compact, nonlinear state-space formulation

$$y_t = \mathbf{h}_t\boldsymbol{\theta}_t + \mathbf{X}_t\boldsymbol{\mu} + v_t, \tag{7}$$

$$\boldsymbol{\theta}_t = \mathbf{f}_t(\boldsymbol{\theta}_{t-1}) + \mathbf{u}_t, \tag{8}$$

where $t \in \{t = 1, \ldots, T\}$, $\boldsymbol{\theta}_t = [G_t \; q_{1t} \; \cdots \; q_{nt}]'$, $\mathbf{u}_t \sim \mathbf{N}[\mathbf{0}, \mathbf{Q}]$, $\mathbf{h}_t = [1 \; 0 \; \cdots \; 0]$, and $\mathbf{f}_t$ is a nonlinear function of the prior state vector, $\boldsymbol{\theta}_{t-1}$. Hence the first element of the state vector $\boldsymbol{\theta}_t$ is goodwill, and the remaining elements are (n) advertising qualities, one for each theme. Error terms, $\mathbf{u}_t$ and v_t, are assumed to be mean zero, independent normal variables. Notably, while the observation Equation (7) is linear, the system Equation (8) is nonlinear in the state variables (goodwill and quality) because the forgetting rate δ_t is a function of the prior goodwill G_{t-1}. This nonlinearity has implications for the Bayesian procedure next prescribed.

4.3. Model Estimation

Bayesian methods provide a rigorous framework for estimating the dynamic system defined by Equations (7) and (8). The approach estimates the probability density function (pdf) of the state vector $\boldsymbol{\theta}_t$ based on the information available up to time period, t. In linear normal systems (cf. Naik et al. 1998, Neelamegham and Chintagunta 2004, Van Heerde et al. 2004, Liechty et al. 2005), this pdf is normal at every period and the basic Kalman filter applies. In nonlinear (e.g., Equation (8)) or nonGaussian systems, however, there is no general (or closed-form) expression for this pdf, so it has to be approximated. Two early approximation algorithms are the extended Kalman filter (EKF) (Anderson and Moore 1979) and the grid-based (GB) filter (Bucy and Senne 1971). Even so, to estimate the pdf of $\boldsymbol{\theta}_t$, this study employs the particle filter (Doucet et al. 2001, Liu and Chen 1998). Particle filtering belongs to a class of sequential Monte Carlo (SMC) integration methods that are based on Bayesian inference. It involves the use of particles (or samples) and their associated weights to approximate the pdf cited above. Unlike EKF and GB methods, particle filters are flexible, convenient, and have been successful in a wide range of complex applications, especially in visual tracking and surveillance (Ristic et al. 2004). Moreover, because the particle filter uses independent samples, it avoids the convergence difficulties (or "stickiness") associated with traditional MCMC procedures (Gamerman 1997).

4.3.1. Particle Filtering and MCMC. More formally, then, the Bayesian estimation of the above discrete system involves first specifying priors for all model parameters, and then using particle filters to estimate the state vectors, and MCMC to estimate the hyperparameters. Let $\boldsymbol{\theta}_{0:T} = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_T\}$ and $\mathbf{y}_{1:T} = \{y_1, y_2, \ldots, y_T\}$ be the state vectors and telephone usage across the entire time series. Let $\boldsymbol{\Psi} = \{\boldsymbol{\eta}, \{\boldsymbol{\varphi}_i, \lambda_i, c_i, r_i\}_{i=1}^n\}$ be a vector of forgetting, interaction, wear out and pooling effects parameters, defined in the system Equation (8); and $\boldsymbol{\mu}$ the coefficient for the other explanatory variables that influence telephone usage, defined in the observation Equation (7). Assume, furthermore, that the prior on the observation variance σ_v^2 is inverse gamma, and

[8] Bass et al. (2007) compared two methods for incorporating price endogeneity: (i) a two-step approach that involves estimating a Bayesian regression of price on all exogenous and instrumental variables, and replacing price in the DLM with its predictive values obtained from the Bayesian regression and (ii) a fully Bayesian instrumental variable procedure given in Rossi et al. (2005). The substantive results were unchanged across both methods. Thus, to be consistent, I follow Bass et al. (2007) and report results based on method (1).

the prior on the system variance $\mathbf{Q}$, is independent inverse Wishart, with both prior specifications chosen to allow the data to dominate the results. Then, by using an indirect sampling approach, the complete joint posterior, $p(\boldsymbol{\theta}_{0:T}, \mathbf{Q}, \sigma_v^2, \boldsymbol{\Psi}, \boldsymbol{\mu} \mid \mathbf{y}_{1:T})$, can be computed by iteratively resampling conditional posteriors, $p(\boldsymbol{\theta}_{0:T} \mid \mathbf{y}_{1:T}, \sigma_v^2, \mathbf{Q}, \boldsymbol{\Psi}, \boldsymbol{\mu})$ and $p(\mathbf{Q}, \sigma_v^2, \boldsymbol{\Psi}, \boldsymbol{\mu} \mid \boldsymbol{\theta}_{0:T}, \mathbf{y}_{1:T})$. The conditional posterior, $p(\boldsymbol{\theta}_{0:T} \mid \mathbf{y}_{1:T}, \sigma_v^2, \mathbf{Q}, \boldsymbol{\psi}, \boldsymbol{\mu})$, is clearly nonlinear in the state vector $\boldsymbol{\theta}_t$, and is thus sampled using particle filters. The procedure, based on importance sampling, a Monte Carlo integration method (see, for example, Geweke 1989), provides a discrete approximation to the posterior density of the states through a set of H_s support points (or particles) $\{\boldsymbol{\theta}_{0:t}^h\}_{h=1}^{H_s}$, and their respective weights, $\{w_{0:t}^h\}_{h=1}^{H_s}$, where $w_t^h > 0$ and $\sum_{h=1}^{H_s} w_t^h = 1$. Finally, conditional on all state vectors and the data, it is straightforward to sample the hyperparameters from $p(\mathbf{Q}, \sigma_v^2, \boldsymbol{\Psi}, \boldsymbol{\mu} \mid \boldsymbol{\theta}_{0:T}, \mathbf{y}_{1:T})$, using standard MCMC techniques (Gamerman 1997, Allenby and Rossi 2003). (See the appendix for details.)

In addition, with the filtered density and a sequence of predictions in hand, one can use a Monte Carlo approach to compute forecasts that parallel those produced in linear state-space models (West and Harrison 1997). For example, $\boldsymbol{\theta}_{t+k\mid t} = E(\boldsymbol{\theta}_{t+k} \mid y_{1:t})$ denotes the expected state vector at period $t+k$ given telephone usage observed through period t. By definition, this k-step ahead forecast is given by

$$\boldsymbol{\theta}_{t+k\mid t} = \int \boldsymbol{\theta}_{t+k} p(\boldsymbol{\theta}_{t+k} \mid y_{1:t})\, d\boldsymbol{\theta}_{t+k}. \tag{9}$$

Equation (9) can be approximated as follows. First, the particle filter is used to generate random samples of size H, $\{\boldsymbol{\theta}_t^1, \boldsymbol{\theta}_t^2, \ldots, \boldsymbol{\theta}_t^H\}$ from the filtered density $p(\boldsymbol{\theta}_t \mid \mathbf{y}_{1:t})$. Second, the system Equation (8) is used, setting $\boldsymbol{\theta}_t = \boldsymbol{\theta}_t^h$, to simulate predictions $\{\boldsymbol{\theta}_{t+1}^h, \boldsymbol{\theta}_{t+2}^h, \ldots, \boldsymbol{\theta}_{t+k}^h\}$ for all $h \leq H$. For sufficiently large H, an approximation to the integral in (9) is given by (cf. Kim et al. 1998, Fleming and Kirby 2003)

$$\boldsymbol{\theta}_{t+k\mid t} \approx \frac{1}{H} \sum_{h=1}^{H} \boldsymbol{\theta}_{t+k}^h. \tag{10}$$

This step-ahead procedure is later employed to assess the forecasts of several alternative specifications.

4.4. Model Assessment

To investigate the importance of pooling effects and a dynamic forgetting rate, I compare the proposed model (Model 7) to six plausible alternate specifications. The first (Model 1) has a constant forgetting rate and no pooling or interaction effects ($\delta_t = \delta$, $\lambda_i = 0, \varphi_{ij} = 0$). The second (Model 2) has a constant forgetting rate ($\delta_t = \delta$) and no interaction effects ($\lambda_i = 0$), but an overall pooling effect φ_i for each theme. The third (Model 3) has a dynamic forgetting rate (δ_t) and no pooling or interaction effects ($\lambda_i = 0$, $\varphi_{ij} = 0$). The fourth (Model 4) has a dynamic forgetting rate (δ_t) and an overall pooling effect φ_i for each theme, but no interaction effects ($\lambda_i = 0$). The fifth (Model 5), proposed in Bass et al. (2007), has a constant forgetting rate, interaction effects, and no pooling effects ($\delta_t = \delta$, $\lambda_i \neq 0$, $\varphi_{ij} = 0$). The sixth (Model 6) has a dynamic forgetting rate and pooling effects φ_{ij}, but no interaction effects ($\lambda_i = 0$). Models 1, 2, and 5 are linear state-space models; Models 3, 4, 6, and 7 are nonlinear. In addition, the nonlinear models were estimated assuming two specifications for predictors $\mathbf{Z}_{1t}$ of the forgetting rate (δ_t). In the first, prior goodwill and competitive advertising are log transformed; in the second, they remain untransformed.

Table 4 **Comparison of Proposed Specifications with Alternate Specifications**

Models	Description	Log Bayes factors	MAD	RMSE
Model 1	Baseline model	115.07	0.9837	1.2237
Model 2	Aggregate pooling	42.17	0.6992	0.9035
Model 3	Forgetting	36.60	0.6325	0.8166
Model 4	Forgetting and aggregate pooling	24.84	0.5167	0.6795
Model 5	Linear Bass et al. (2007)	16.30	0.4194	0.5915
Model 6	Forgetting and pooling	12.21	0.3957	0.5713
Model 7	Proposed model	—	0.2383	0.3024

Table 4 compares the six alternatives to the proposed model using log Bayes factors. A log Bayes factor greater than two provides strong evidence in favor of the proposed model (West and Harrison 1997). For the nonlinear models (Models 3, 4, 6, and 7), I report the specifications that used the log transformed predictors for δ_t because these specifications, in terms of log Bayes factors, dominated those with the untransformed predictors. Overall, the log Bayes factors in Table 4 clearly reject all alternatives in favor of the proposed model (Model 7). Furthermore, the model with only pooling effects (Model 2) and the model with only dynamic forgetting (Model 3) dominate the model with the constant forgetting rate and no pooling or interaction effects (Model 1). Similarly, based on log Bayes factors, nonlinear Model 6 also outperforms Model 5, the linear Bass et al. (2007) model. For further evidence of model fit, I compare the predictive (in-sample) performance of all models using two measures: mean absolute deviation (MAD) and root mean square error (RMSE). These two measures in Table 4 are the lowest for the proposed model.

Finally, to compare out-of-sample forecasts of all seven specifications, I apply the step-ahead algorithm outlined in forecasting Equations (9)–(10). For each specification, telephone usage forecasts $E(y_{t+k} \mid y_{1:t})$ are generated using two sample ($t = 1{:}50$, $t = 1{:}90$) and forecast (k-step ahead) time periods ($k = 1{:}5$, $k = 1{:}10$). Table 5 reports the MAD and RMSE of these

Table 5 Step-Ahead Forecast Errors

		MAD		RMSE	
Models	Forecasts	$T=50$	$T=90$	$T=50$	$T=90$
Model 1	5-step ahead	3.4444	3.1692	3.7859	3.6536
	10-step ahead	3.6859	4.6528	4.0723	4.7042
Model 2	5-step ahead	1.7653	3.5061	2.0041	3.6331
	10-step ahead	1.8195	4.0371	2.6075	4.1175
Model 3	5-step ahead	0.7447	1.5052	0.8951	1.7785
	10-step ahead	1.2045	2.1892	1.4516	2.3226
Model 4	5-step ahead	0.8214	0.6404	0.8321	0.8749
	10-step ahead	0.8573	1.1917	1.3381	0.9011
Model 5	5-step ahead	0.7790	0.6722	0.9422	0.7159
	10-step ahead	0.9091	1.1454	1.0107	1.1650
Model 6	5-step ahead	0.6155	0.6797	0.7122	0.7253
	10-step ahead	0.6168	0.6004	0.8140	0.8374
Model 7	5-step ahead	0.4531	0.5711	0.6271	0.6154
	10-step ahead	0.5010	0.5533	0.7100	0.6577

forecasts. The evidence clearly supports the predictive performance of the proposed model (Model 7). Thus, taken together, these fit and forecast measures suggest the importance of accounting for pooling effects and a dynamic forgetting rate.

5. Results

Table 6 presents the main results from the particle filter/MCMC estimation of the proposed model. It shows estimates of the mean, standard deviation, and 95% highest probability density interval (HPDI) of each parameter. Parameter estimates for price, capacity, wear out, and interaction are comparable to those previously discussed in Bass et al. (2007); thus, that earlier discussion will not be repeated here. The primary focus here will be on results related to forgetting, carry-over, and pooling effects.

Estimates related to the forgetting rate (δ_t) are largely consistent with conjectures in Gensch and Welam (1973), experimental findings in Ray et al. (1971), and analytic findings in Blattberg and Jeuland (1981). That is, carry-over rates $d_t = (1-\delta_t)$ might not be constant as historically assumed (cf. Bass and Clarke 1972, Clarke 1976, Hanssens et al. 1990), but might be time varying and affected by several advertising variable. First, the linear and quadratic effects of prior goodwill (G_{t-1}) are significant; the linear effect is negative, and the quadratic effect is positive. Hence the probability of carry-over demand could first increase with prior ad spending, but at a diminishing rate, and then decrease when prior ad expenditures attain some level of overspending (Bultez and Naert 1979). As cited earlier, this result is consistent with an extensive body of work that catalogs the positive but diminishing effects of ad repetition (cf. Belch 1982, Cacioppo and Petty 1979). Second, the effect of competitive advertising on the forgetting rate

Table 6 Estimates Proposed Model (Model 7)

	Estimate	Standard deviation	95%	HPDI
Parameters:				
Price per minute	−9.1148	3.0618	−14.2119	−4.1179
Capacity (number of lines)	0.8174	0.2722	0.3735	1.2701
Initial goodwill, G_0	15.3740	0.3640	15.2050	15.5880
Forgetting rate model:				
Constant	−6.9199	2.0090	−10.2417	−3.6440
Prior goodwill, G_{t-1}	−8.8807	2.2455	−12.4561	−5.1493
Prior goodwill^2, $(G_{t-1})^2$	3.7985	0.9030	2.3250	5.2135
Competitive advertising	0.1261	0.2242	−0.2508	0.4881
Cumulative percentage emotional	−0.9411	0.3014	−1.0742	−0.8140
Shape (m)	3.0744	0.9391	1.6660	4.6900
Pooling effects: (ϕ_{tj})				
Call stimulation:				
Price offer	0.2518	0.4005	−0.2048	0.7589
Product offer	0.5285	0.2140	0.1640	1.0218
Reconnection	0.1418	0.1389	−0.0744	0.3642
Reassurance	0.3176	0.1245	0.1096	0.6446
Price offer:				
Call stimulation	0.1940	0.0761	0.0217	0.3934
Product offer	0.3253	0.1345	0.0284	0.6345
Reconnection	0.1153	0.1536	−0.0687	0.3949
Reassurance	0.1603	0.2076	−0.0960	0.5759
Product offer:				
Call stimulation	0.0988	0.0377	0.0675	0.1340
Price offer	0.3020	0.1240	0.1084	0.5075
Reconnection	0.1102	0.1104	−0.0198	0.2627
Reassurance	0.0072	0.1311	−0.1314	0.1667
Reconnection:				
Call stimulation	0.1788	0.1853	−0.0943	0.5127
Price offer	0.3881	0.5556	−0.3887	1.4411
Product offer	0.3330	0.3623	−0.1852	0.9757
Reassurance	0.1978	0.0891	0.0243	0.3682
Reassurance:				
Call stimulation	0.3671	0.1335	0.0160	0.7061
Price offer	0.5612	0.8549	−0.5375	1.9559
Product offer	0.6289	0.6803	−0.2460	1.7821
Reconnection	0.3340	0.1245	0.0108	0.7464
Own wear out effects:				
Call stimulation:				
Copy wear out, c	0.6017	0.1793	0.3146	0.9040
Repetition wear out, r	−0.0675	0.0235	−0.1281	−0.0190
Initial quality, q_{10}	0.1214	0.0824	0.0167	0.2635
Price offer:				
Copy wear out, c	0.9859	0.2290	0.6161	1.3742
Repetition wear out, r	−0.1040	0.0297	−0.1945	−0.0169
Initial quality, q_{20}	0.0577	0.0341	0.0090	0.1191
Product offer:				
Copy wear out, c	0.7248	0.3011	0.2111	1.2224
Repetition wear out, r	−0.0852	0.0206	−0.1549	−0.0201
Initial quality, q_{30}	0.0470	0.0374	0.0511	0.1161
Reconnection:				
Copy wear out, c	0.4662	0.1757	0.2076	0.7775
Repetition wear out, r	−0.0363	0.0115	−0.0503	−0.0284
Initial quality, q_{40}	0.1440	0.0990	0.0257	0.3192
Reassurance:				
Copy wear out, c	0.5404	0.2038	0.2494	0.9103
Repetition wear out, r	−0.0295	0.0102	−0.0434	−0.0179
Initial quality, q_{50}	0.1255	0.0829	0.0179	0.2838
Interaction: (λ_{it})				
Call stimulation	−0.0793	0.0284	−0.1247	−0.0313
Price offer	−0.0407	0.0724	−0.1471	0.0842
Product offer	−0.0742	0.0532	−0.1568	0.0129
Reconnection	−0.1002	0.0275	−0.1439	−0.0538
Reassurance	−0.1415	0.0410	−0.2106	−0.0775

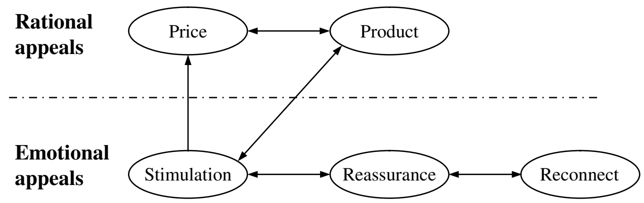
is positive but insignificant. Presumably, competitive advertising could accelerate forgetting and shorten the duration (Clarke 1976) of advertising effectiveness (cf. Assmus et al. 1984, Kent 1991, Kent and Allen 1993). However, the lack of a significant effect in this study may be a result of the firm's dominance in its category.

Third, the effect of ad theme on the forgetting rate (δ_t) is negative and significant, indicating that the effects of prior ad spending become more short lived as the cumulative percentage of price and product offer ads increase. To gain some insight into this finding, one should recall the theme classification adopted in Bass et al. (2007), which though informal has support in Belch and Belch (1998, pp. 267–271). The classification suggests that price and product offer ads have more rational appeals than call stimulation, reassurance, and reconnect ads. This third finding, therefore, suggests that rational ads are more short lived. This seems consistent with prior studies that suggest rational ads wear out faster than emotional ads and are less effective in mature markets (e.g., residential telephone services) (Chandy et al. 2001).

Estimates of the pooling effects (ϕ_{ij}) are all positive, and several are significant. This implies that pooling and varying ad themes (substantive variation) can forestall the wear out of individual ad themes; analogous to the manner in which varying the executions (cosmetic variation) of an ad theme may forestall its wear out (Unnava and Burnkrant 1991). Pooling estimates (ϕ_{ij}) also vary widely. For example, the effect of product offer on call stimulation is 0.5285, while the effect of call stimulation on product offer is 0.0988. Similarly, the effect of price offer on reassurance is 0.5612, while the effect of reassurance on price offer is 0.1603. Based on the above theme classification, these results reveal an important asymmetry. For any pair of emotional ad theme (i) and rational ad theme (j), the effects of theme (j) on theme (i) is greater than the effects of theme (i) on theme (j), ($\phi_{ij} > \phi_{ji}$). This asymmetry suggests that wear out reduction because of pooling and varying of ad themes might differ across emotional and rational messages. I shall return to this question presently.

Still, beyond reducing wear out, advertisers could benefit from airing multiple ad themes because purchase decisions are often based on multiple criteria. For example, many purchases are not based solely on price—product features are often just as important. Moreover, purchases decisions are often made because of both rational and emotional motives (Belch and Belch 1998, pp. 271–272).[9] Figure 1 shows

Figure 1 **Pooling Effects Significant at the 95% Level**



schematically pooling effects (ϕ_{ij}) that are positive and significant at the 95% level, and suggests theme synergies the firms may exploit, when scheduling advertising. The arrows in Figure 1 point in the direction of the effect. For example, call stimulation ads improve the effectiveness of price offer ads, but not vice versa. Other effects are interdependent, in that each theme directly improves the others effectiveness. Thus the ad theme pairs—price offer and product offer, reassurance and reconnect, reassurance and call stimulation—and product offer and call stimulation—are interdependent.

Finally, to return to the question of why pooling effects might benefit emotional appeals more than rational appeals, it is necessary to recall that emotional appeals are more effective in mature markets (Chandy et al. 2001). The effectiveness of emotional appeals is related to their greater impact on memory—e.g., greater attention and recall (Olney et al. 1991, Thorson and Page 1988). Thus, in mature markets, theme (substantive) variation, more effective at higher levels of attention (Schumann et al. 1990), can have a greater effect on emotional themes. This, in turn, implies that, along with wear-out rates, pooling effects (ϕ_{ij}) can differ across emotional and rational ad themes. Table 7 examines this implication. It reports average pooling effects, repetition wear out, and number of executions for each advertising theme. Note that reassurance had the highest average pooling effect. Moreover, the average pooling effect is larger for emotional ads than it is for rational ads; this suggests that the strategy of pooling and varying ad themes (substantive variation) could be more effective for emotional ads. Consider also the number of executions and the repetition wear out for each theme in Table 7. Note that as the number of executions increases, repetition wear out decreases. Rational ads, with the highest number of executions, have the smallest wear-out rates; and therefore might benefit more from cosmetic variation.

[9] According to Ogilvy and Raphaelson (1982, p. 18), "Few purchases of any kind are made for entirely rational reasons. Even a purely functional product such as laundry detergent may offer what is now called an 'emotional end benefit'—say, the satisfaction of seeing one's children in bright, clean clothes. In some product categories, the rational element is small. These include soft drinks, beer, cosmetics, certain personal-care products, and most fashion products. And who hasn't experienced the surge of joy that accompanies the purchase of a new car?"

Table 7 Comparison of Wear Out and Pooling Effects Parameters

	Number of executions	Repetition wear-out parameter	Average pooling effects parameter
Rational appeal:			
Price offer ads	60	−0.1040	0.1987
Product ads	22	−0.0852	0.1296
Average	41	−0.0946	0.1641
Emotional appeal:			
Call stimulation ads	43	−0.0675	0.3099
Reassurance ads	17	−0.0363	0.4728
Reconnect ads	16	−0.0295	0.2744
Average	25.3	−0.0444	0.3528

5.1. The Implied Carry Over Effects of Themes

As cited in the introduction, advertising theme efficiency depends more generally on the decay in carry-over effects from prior goodwill. It is therefore worthwhile to explore the impact of a dynamic carry-over rate d_t on demand attributable to past thematic advertising. An implied measure of this demand can be derived by substituting the goodwill Equation (1) recursively into the demand Equation (7) to obtain the nonlinear distributed lag form (see Gensch and Welam 1973, Hitsch 2006)

$$\begin{aligned} y_t^a &= f(\mathbf{a}_t) + d_t f(\mathbf{a}_{t-1}) + d_t d_{t-1} f(\mathbf{a}_{t-2}) \\ &\quad + \cdots + \prod_{k=0}^{J} d_{t-k} f(\mathbf{a}_{t-J-1}) \\ &= \underbrace{f(\mathbf{a}_t)}_{\text{Current}} + \underbrace{\sum_{j=1}^{J}\left(\prod_{k=0}^{j} d_{t-k}\right) f(\mathbf{a}_{t-j-1})}_{\text{Carry-over}}, \end{aligned} \tag{11}$$

where

$$f(\mathbf{a}_t) = \sum_{i=1}^{n} q_{it-1}\left(g(A_{it}) + \lambda_i \sum_{\substack{j=1\\ j\neq i}}^{n} h(A_{it}, A_{jt})\right) \quad \text{and}$$

$$d_t = 1 - \delta_t.$$

From Equation (11), the implied *carry-over* demand in week (t) is a separable nonlinear function of the prior ad spending. It is also noteworthy to consider the set of lag weights (S_{it}) applied to prior ad spending for a given theme, i: $S_{it} = \{d_t q_{it-2}, d_t d_{t-1} q_{it-3}, d_t d_{t-1} d_{t-2} q_{it-4}, \ldots\}$. These weights, unlike those in early studies (cf. Bass and Clarke 1972, Clarke 1976), are time varying; this means carry-over also depends on the period an ad is aired (Thaivanich et al. 2000), as well as on the variables that may moderate wear out and forgetting (δ_t). Moreover, these time-varying weights imply that there is no single exponential decay function for past advertising as implied in early studies (e.g., Clarke 1976), but a family of such functions, in this case, one for each period ($t \geq 2$) and theme (see Ray et al. 1971, p. 14). With Equation (11), it is possible to derive implied estimates of the carry-over by theme per week, using mean values of the state variables and hyperparameters (see the appendix).

Table 8 Implied Carry-Over by Advertising Themes

	Model 7 (δ_t)		Model 5 (δ)	
Ad themes	$\lambda_i = 0$	$\lambda_i \neq 0$	$\lambda_i = 0$	$\lambda_i \neq 0$
Call stimulation	2.0507	1.0382	2.1610	1.0865
Price offer	1.4767	0.9731	1.4917	1.3038
Product offer	1.3932	0.6336	1.8195	1.3583
Reconnect	2.2845	0.7007	2.6328	1.1929
Reassurance	2.0878	0.6199	3.8752	1.3547

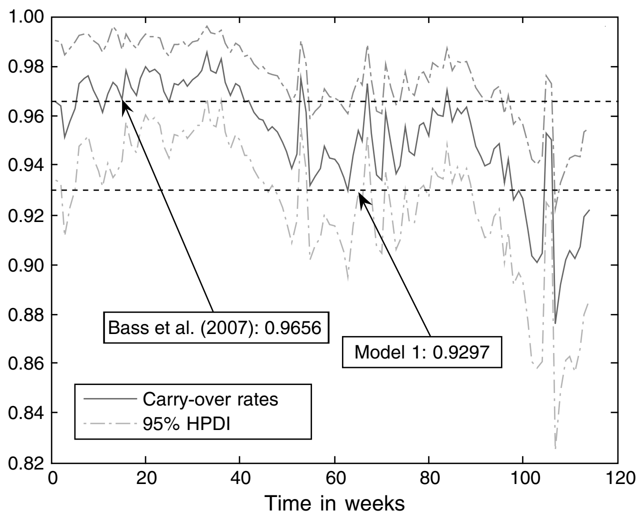
Table 8 reports the total carry-over demand ($t = 2, 3, \ldots, T$) for each theme (divided by total theme spending), based on the estimates in Model 7 (the full model) and Model 5 (Bass et al. 2007), with and without interaction effects (λ_i). These interaction effects were found to be negative, and more so for emotional ads. There are several interesting observations in Table 8. First, carry-over varies considerably across themes. For example, in the proposed model, when interaction is set to zero ($\lambda_i = 0$), product offer has the lowest carry-over (1.3932), while reconnect has the highest (2.2845). Moreover, in both models (with $\lambda_i = 0$), price and product offer (rational themes) have the lowest carry-over, analogous to results in prior studies that suggest that rational ads wear out faster than emotional ads. Conversely, when interaction is not set to zero ($\lambda_i \neq 0$) in the proposed model, reassurance has the lowest (0.6199), and call stimulation the highest (1.0382). Thus, in a multitheme context, where some themes are aired concurrently, quality wear out alone probably provides an inaccurate summary of ad efficiency.

Second, the carry-over measure for each theme is higher in Model 5, suggesting a bias if one assumes the forgetting rate is constant. The size of bias averages 48% and ranges from 5% to 119%. Additional support for this result emerges from Figure 2—plots of the mean and 95% interval for the dynamic carry-over rate d_t across the 114 weeks. Here, the constant carry-over rate in Model 5 exceeds the mean of d_t (0.9550), inflating the effects prior ads on current demand.

5.2. Multitheme Budgeting with Pooling Effects

A major goal of this study is to show the implications of pooling effects on the ad planner's budget allocation problem. Thus, in this section, I reconsider the budgeting problem over time, across the five advertising themes, and under two scenarios: pooling effects (Model 7) and no pooling effects (Model 5). In each scenario, price offer and product offer themes

Figure 2 Time-Varying Carry-Over Rate, $d_t = 1 - \delta_t$



had the two highest copy wear out but the lowest repetition wear-out rates. With pooling effects, however, call stimulation and product offer ads could increase the quality of price ads, and reassurance themes had the highest average pooling effects. How might these results affect the budget allocation decisions and demand for telephone services? It turns out that with pooling effects (Model 7), advertisers can exploit synergies among ad themes to increase advertising efficiency (Raman and Naik 2004, Naik et al. 2005).

To show this increased efficiency, I develop a model that reallocates the total advertising GRPs (b_t) in each period, across the five themes to maximize the total expected demand (y_t) over $T = 114$ weeks. That is, with estimates of the state vectors from the particle filter $\{\boldsymbol{\theta}^h_{0:T}, w^h_{0:T}\}_{h=1}^{H_s}$, and hyperparameters $\{\boldsymbol{\Psi}, \boldsymbol{\mu}, \mathbf{Q}, \sigma^2_v\}$ at their mean values, I solve the following optimization problem, F1:

$$\max_{A_{11}..A_{1m}..A_{T1}....A_{Tm}} \sum_{t=1}^{T}\sum_{h=1}^{H} w^h_{t-1} E(y_t \mid \boldsymbol{\theta}^h_{t-1})$$

$$\text{st} \quad \sum_{i=1}^{n} A_{it} \le b_t, \quad A_{it} \ge 0 \quad t = 1, \ldots, T,$$

where $E(y_t \mid \boldsymbol{\theta}^h_{t-1})$ is the one-step-ahead forecast at the particle $\{\boldsymbol{\theta}^h_{t-1}\}$, and A_{it} the GRP for ad theme (i) at period t.[10] Using multiple start values, I select solutions to F1 that give allocations that represent improvements in advertising efficiency.

[10] The program (F1) is solved in SNOPT, routines developed by Stanford Optimization Laboratory. SNOPT, called within MATLAB, can solve large-scale optimization problems (Gill et al. 2002). The purpose of F1 is to suggest implications of the pooling effects parameters, and the solutions should not be interpreted as *optimal* in the strategic (or game-theoretic) sense.

Table 9 Comparison of GRP Allocations (0000s)

		No pooling effects†		Pooling effects††	
Ad themes	Actual	Model 5	Change (%)	Model 7	Change (%)
Call stimulation	1.1509	0.5815	−49.5	0.6575	−42.9
Price offer	0.3611	0.0914	−74.7	0.4079	13.0
Product offer	0.7281	0.3218	−55.8	0.3895	−46.5
Reconnect	0.5201	1.0962	110.7	0.6958	33.8
Reassurance	0.2422	0.9115	276.3	0.8517	251.7

†2.0% increase in total hours. ††4.4% increase in total hours.

Table 9 reports one such allocation, along with the reallocation based on Model 5 (Bass et al. 2007). Recall, without pooling effects (Model 5), the reallocation shifts resources markedly away from both price and product ads, i.e., ads with the high copy wear-out rates. This reallocation created an additional 35.82 million hours of calling time, a 2.0% increase over the actual demand. With pooling effects, however, the solution to (F1) would increase ad expenditures on reconnect, reassurance, and now price ads, while decreasing expenditures on the other two themes. Reassurance had the highest percentage (251.7%) increase in allocation and product offer had the highest percentage (46.5%) decrease. This reallocation of advertising would create an additional 83.13 million hours of calling time, a 4.4% increase over the actual demand. Therefore, in the presence of pooling effects, increased expenditures on price offer and reassurance ads could lead to greater telephone use (i.e., greater ad efficiency). Recall that reassurance had the highest average pooling effect and price ad benefits from the quality of call stimulation and product ads. Thus, not only do pooling effects benefit emotional ads, but pooling effects might also mitigate the negative wear-out effects of rational, price offer ads; as a result, rational messages can be more effective when combined with emotional messages (Belch and Belch 1998, Ogilvy and Raphaelson 1982).

6. Conclusion

This study investigates the factors that could affect the rates at which advertising decays over time, the pooling effects of different themes, and the implications of these effects on the allocation of advertising resources. In doing so, it extends the linear NA (1962) model employed in Bass et al. (2007), to a nonlinear model of advertising quality and goodwill. The extended model is estimated using a combination of MCMC and particle filtering ideas. Particle filtering belongs to a class of SMC methods developed to estimate nonlinear/nonGaussian dynamic systems. The results should be of interest to advertisers and integrated marketing communications managers who are concerned with synergies when scheduling multitheme advertising.

The study offers several interesting insights related to the issues of pooling themes and advertising decay. First, estimates related to advertising decay are consistent with conjectures in Gensch and Welam (1973), experimental findings in Ray et al. (1971), and analytic findings in Blattberg and Jeuland (1981). That is, forgetting (carry-over) rates could be time varying and affected by prior goodwill and copy theme. Furthermore, I demonstrate that a constant rate could bias the demand attributable to past advertising. Time-varying rates, on the other hand, imply that there is no single decay function for prior advertising, but a family of decay functions, in this study, one for each period and theme. Ad decay, then, may be a far more complex phenomenon than econometric studies in marketing have traditionally assumed. Second, pooling effects were positive, suggesting that pooling and varying ad themes could forestall wear out, analogous to the manner in which varying the executions of ads for a fixed theme could forestall the wear out of that theme (Unnava and Burnkrant 1991). Therefore, advertisers in multitheme contexts could use multiple executions of the same theme or vary themes to forestall wear out. Rational themes had on average a higher number of executions and lower repetition wear outs, which could mean that airing multiple executions of same theme (i.e., cosmetic variation) is more effective for rational ads. Emotional themes, however, had on average higher pooling effects, which suggests that pooling and varying ad themes (i.e., substantive variation) could be more effective for emotional ads (Schumann and Clemons 1989). Finally, the results of the budget reallocation experiment suggest that advertising is more efficient when pooling effects are present, and pooling effects could mitigate the higher wear-out rates of rational themes.

Yet, this work has several potential limitations. For example, the study treats advertising as an exogenous variable. Presumably, advertising should be treated as an endogenous variable as assumed for price, if managers allocate advertising in a strategic manner. However, given that the estimation uses weekly data, it is perhaps much less likely that endogeneity will be a major factor (Leeflang et al. 2000, p. 382). Still, advertising endogeneity in this state-space model can be addressed in the future if good instruments are available (e.g., Kim 2006). The study could perhaps also have modeled competition more fully, for example, modeling the dynamic effectiveness of competitive advertising. Yet, the firm analyzed here is a monopolist in its category, and so dynamic competition was not modeled explicitly. The NA (1962) model, however, has been commonly extended in the differential game literature to model dynamic advertising competition (e.g., Jorgensen and Zaccour 2004). Thus, given data on competitor's demand, one could draw on these extensions to model dynamic competition more fully and formally.

A future direction for this research could be to extend the model to investigate the nonlinear dynamics of forgetting and satiation at the consumer level. This could, however, pose an interesting computational problem, if the number of consumers is large. Because, then, the state space would also be large, and the particle filter could become unwieldy. Thus, one might have to use procedures to reduce the size of the state space. Finally, particle filters have been widely applied to eye tracking research, where nonlinearity and microdynamics are important (Perez et al. 2004), and thus could find applications in the nascent area of visual marketing (e.g., eye tracking, eye movement studies; Pieters and Wedel 2000, Pieters et al. 1999).

Acknowledgments
The author thanks the editor-in-chief, the area editor, two anonymous reviewers, and seminar participants at University of Michigan, University of Pennsylvania, University of Texas at Dallas, and the Marketing Dynamics Conference at University of California, Los Angeles for their helpful comments and suggestions. The author also thanks Sumit Majumdar, who kindly provided the data for this study. The usual disclaimer applies.

Appendix

This appendix provides an overview of the posterior sampling algorithm. The algorithm combines MCMC and particle filtering/smoothing methods (cf. Doucet et al. 2001, Ristic et al. 2004, Liu and Chen 1998, Godsill et al. 2004). The sampling scheme adopts a *particle* approach because our model belongs to a class of dynamic systems in which the observation Equation (7) is linear, but the system Equation (8) is nonlinear in the state variables. The algorithm should be regarded as a nonlinear/nonGaussian version of the Kalman filter/smoother algorithm for Bayesian DLMs (cf. Carter and Kohn 1994, Fruhwirth-Schnatter 1994). The appendix begins by briefly motivating the nonlinear filtering problem, and then provides a general description of a particle/MCMC solution.

A1. Nonlinear Filtering Problem

Let $\boldsymbol{\theta}_{0:T} = \{\boldsymbol{\theta}_0, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_T\}$ be the relevant state vectors, and let $\mathbf{y}_{1:T} = \{y_1, y_2, \ldots, y_T\}$ be telephone usage observed across the entire data set $t \in \{t = 1, \ldots, T\}$. Recall the first element of this state vector $\boldsymbol{\theta}_t$ is goodwill, and the remaining elements are (n) advertising qualities, one for each ad theme. Suppose further that the most recent values of the hyperparameters $\boldsymbol{\Psi} = \{\boldsymbol{\eta}, \{\boldsymbol{\varphi}_i, c_i, r_i, \lambda_i\}_{i=1}^n\}, \sigma_v^2, \mathbf{Q}$ and $\boldsymbol{\mu}$ are available but excluded from description of the state densities to simplify the exposition. The goal, then, is to estimate the full conditional posterior distribution of the states $p(\boldsymbol{\theta}_{0:T} \mid \mathbf{y}_{1:T})$ and, recursively, the filtered (or marginal) states $p(\boldsymbol{\theta}_t \mid \mathbf{y}_{1:t})$.

The full conditional posterior is given by Bayes rule

$$p(\boldsymbol{\theta}_{0:T} \mid \mathbf{y}_{1:T}) = \frac{p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta}_{0:T}) p(\boldsymbol{\theta}_{0:T})}{\int p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta}_{0:T}) p(\boldsymbol{\theta}_{0:T})\, d\boldsymbol{\theta}_{0:T}}. \qquad (12)$$

But the filtered states are obtained, recursively, via the standard one-step-ahead prediction and updating steps, (12) and (13)

$$\textit{Prediction}: \quad p(\boldsymbol{\theta}_t \mid \mathbf{y}_{1:t-1}) = \int p(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}) p(\boldsymbol{\theta}_{t-1} \mid \mathbf{y}_{1:t-1})\, d\boldsymbol{\theta}_{t-1}, \tag{13}$$

$$\textit{Update}: \quad p(\boldsymbol{\theta}_t \mid \mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t \mid \boldsymbol{\theta}_t) p(\boldsymbol{\theta}_t \mid \mathbf{y}_{1:t-1})}{\int p(\mathbf{y}_t \mid \boldsymbol{\theta}_t) p(\boldsymbol{\theta}_t \mid \mathbf{y}_{1:t-1})\, d\boldsymbol{\theta}_t}. \tag{14}$$

Equation (12) and recursive relationships in Equations (13) and (14) form the foundation for a conceptual Bayesian solution to all state-space systems. The integrals in (12) to (14), however, are unavailable analytically in the case of nonlinear/nonGaussian dynamic systems. This study therefore employs a Monte Carlo integration approach, the particle filter.

A2. Particle Filtering/Smoothing

Sequential importance sampling (SIS) forms the basis for several Monte Carlo filters, including the particle filter, which I now briefly describe. SIS extends the basic importance sampling results to the sequential (or time-varying) case. (For background information, see Geweke 1989, Gelfand and Smith 1990.) A full description is also available in, for example, Doucet et al. (2001). To give an overview here, let $\{\boldsymbol{\theta}^h_{0:T}, w^h_T\}_{h=1}^{H_s}$ be a random measure that characterizes the full conditional posterior $p(\boldsymbol{\theta}_{0:T} \mid \mathbf{y}_{1:T})$, where $\{\boldsymbol{\theta}^h_{0:T}, h = 0, 1, \ldots, H_s\}$ is a set of support points (or particles) with associated weights $\{w^h_T, h = 0, 1, \ldots, H_s\}$; $\boldsymbol{\theta}_{0:T}$ and $\mathbf{y}_{1:T}$ are as previously defined. Suppose the particles and weights are chosen based on importance sampling ideas with the weights normalized such that $\sum_{h=1}^{H_s} w^h_T = 1$, where $w^h_T > 0$. Then, formally, an approximation to the full conditional posterior $p(\boldsymbol{\theta}_{0:T} \mid \mathbf{y}_{1:T})$ is given by

$$p(\boldsymbol{\theta}_{0:T} \mid \mathbf{y}_{1:T}) \approx \sum_{h=1}^{H_s} w^h_T \hat{\delta}(\boldsymbol{\theta}_{0:T} - \boldsymbol{\theta}^h_{0:T}), \tag{15}$$

where $\hat{\delta}(\cdot)$ is the *dirac* delta measure.

Now, consider a recursive procedure to approximate the filtered states defined by (13) and (14), i.e., $p(\boldsymbol{\theta}_t \mid \mathbf{y}_{1:t}) \propto p(\mathbf{y}_t \mid \boldsymbol{\theta}_t) p(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}) p(\boldsymbol{\theta}_{t-1} \mid \mathbf{y}_{1:t-1})$. Suppose at time $t-1$, samples and weights are available to approximate the distribution $p(\boldsymbol{\theta}_{t-1} \mid \mathbf{y}_{1:t-1})$; then at time t, one can approximate $p(\boldsymbol{\theta}_t \mid \mathbf{y}_{1:t})$ by augmenting the existing samples with a new set of samples of the state vector at time, t. More precisely, if the importance density (see Geweke 1989) is chosen to factorise as follows:

$$q(\boldsymbol{\theta}_t \mid \mathbf{y}_{1:t}) = q(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, \mathbf{y}_{1:t}) q(\boldsymbol{\theta}_{t-1} \mid \mathbf{y}_{1:t-1}), \tag{16}$$

then the new particles $\{\boldsymbol{\theta}^h_t\}_{h=1}^{H_s}$ are drawn from $\boldsymbol{\theta}^h_t \sim q(\boldsymbol{\theta}^h_t \mid \boldsymbol{\theta}^h_{t-1}, \mathbf{y}_{1:t})$, where the weights and the filtered states are as follows (Doucet et al. 2001, 1999; Ristic et al. 2004):

$$w^h_t \propto w^h_{t-1} \frac{p(\mathbf{y}_t \mid \boldsymbol{\theta}^h_t) p(\boldsymbol{\theta}^h_t \mid \boldsymbol{\theta}^h_{t-1})}{q(\boldsymbol{\theta}^h_t \mid \boldsymbol{\theta}^h_{t-1}, \mathbf{y}_{1:t})},$$

$$\text{normalized to } w^h_t \to \frac{w^h_t}{\sum_{l=1}^{H_s} w^l_t}, \quad \text{and} \tag{17}$$

$\{\boldsymbol{\theta}^h_t, w^h_t\}_{h=1}^{H_s}$ is a discrete approximation of $p(\boldsymbol{\theta}_t \mid \mathbf{y}_{1:t})$.

The major problem with the SIS particle filter is the degeneracy phenomenon, where after a certain number of recursive steps, all but one particle have neglible weights. Several versions of the particle filter have been developed to address this problem (Doucet et al. 2001). Even so, I use two standard and convenient methods mentioned in the SMC literature: (i) resampling and (ii) choice of the importance function.

First, a common measure of degeneracy is the effective sample size

$$H_{\mathit{eff}} = \frac{1}{\sum_{h=1}^{H_s} (w^h_t)^2}, \tag{18}$$

where w^h_t is the normalized weights obtained in (17). If H_{eff} is small and $H_{\mathit{eff}} < H_s$, then degeneracy is indicated. So to reduce degeneracy, resampling with replacement is done if H_{eff} is below some threshhold. Second, degeneracy is also reduced by the choice of the importance function. Therefore the particles will be sampled from the *optimal* importance function, $p(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, \mathbf{y}_{1:t})$, which is available analytically because here the observation Equation (7) is linear. Ristic et al. (2004) show that this importance function $q(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, \mathbf{y}_t) = p(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, \mathbf{y}_t)$ is optimal in the sense that it minimizes the variance of the importance weights. I choose sample size $H_s = 1{,}000$ and a threshold of $0.8H_s$ values observed in the literature.

At this point, we have a mechanism to obtain a particle (or discrete) representation of the filtered densities $p(\boldsymbol{\theta}_t \mid \mathbf{y}_{1:t})$. Smoothed estimates can also be derived based on this discrete representation. For example, Godsill et al. (2004) provide such a method for smoothing computations, as well as for generating random samples from the above filtered particles. First, recall from Godsill et al. (2004) that smoothing is based on the factorization

$$p(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t+1}, \mathbf{y}_{1:T}) \propto p(\boldsymbol{\theta}_t \mid \mathbf{y}_{1:t}) f(\boldsymbol{\theta}_{t+1} \mid \boldsymbol{\theta}_t). \tag{19}$$

That means, by using weighted resampling ideas (Geweke 1989), one can resample backward using the filtered particles and the revised weights

$$w^h_{t\mid t+1} = \frac{w^h_t f(\hat{\boldsymbol{\theta}}_{t+1} \mid \boldsymbol{\theta}^h_t)}{\sum_{h=1}^{H_s} w^h_t f(\hat{\boldsymbol{\theta}}_{t+1} \mid \boldsymbol{\theta}^h_t)}. \tag{20}$$

In other words, for backward smoothing, all one needs to do is reweight the filtered sample according to (20).

The following, then, is the psuedocode for the particle filtering/smoothing algorithm just briefly decribed. As noted above, it should be regarded as a nonlinear/nonGaussian version of the Kalman filter/smoother algorithm for Bayesian DLMs (cf. Carter and Kohn 1994, Fruhwirth-Schnatter 1994).

A2.1. Simulation for $\boldsymbol{\theta}_{0:T}$

Forward Filtering: Sampling Importance Resampling.

(1) For $h = 1, \ldots, H_s$ Sample: $\boldsymbol{\theta}^h_0 \sim p(\boldsymbol{\theta}_0)$, $w^h_0 = 1/H_s$. Set $t \to 1$.

(2) For $h = 1, \ldots, H_s$ Sample: $\boldsymbol{\theta}^h_t \sim q(\boldsymbol{\theta}_t \mid \boldsymbol{\theta}_{t-1}, \mathbf{y}_t)$

(3) For $h = 1, \ldots, H_s$, Update weights:

$$w^h_t = w^h_{t-1} \frac{p(\mathbf{y}_t \mid \boldsymbol{\theta}^h_t) p(\boldsymbol{\theta}^h_t \mid \boldsymbol{\theta}^h_{t-1})}{q(\boldsymbol{\theta}^h_t \mid \boldsymbol{\theta}^h_{t-1}, \mathbf{y}_t)},$$

$$\text{and normalize them: } w^h_t \to \frac{w^h_t}{\sum_{l=1}^{H_s} w^l_t}.$$

(4) If $H_{eff} < 0.8H_s$ resample with replacement from the set $\{\boldsymbol{\theta}_t^h\}_{h=1}^{H_s}$, where w_t^h is the probability of resampling the state $\boldsymbol{\theta}_t^h$. Reset the weights $w_t^h = 1/H_s$.

(5) Set $t \to t+1$, repeat Step 2 until end of time period (T). Filtered estimates of the full posterior are obtained from $\{\boldsymbol{\theta}_{0:T}^h, w_{0:T}^h\}_{h=1}^{H_s}$.

Backward Sampling: (See Godsill et al. 2004.)

(1) Choose $\hat{\boldsymbol{\theta}}_T = \boldsymbol{\theta}_T^h$ with probabilty w_T^h.

(2) For $t = T-1$ to 1.

- Calculate $w_{t|t+1}^h \propto w_t^h f(\hat{\boldsymbol{\theta}}_{t+1} \mid \boldsymbol{\theta}_t^h)$ for each $h = 1, \ldots, H_s$.
- $\hat{\hat{\boldsymbol{\theta}}}_t = \boldsymbol{\theta}_t^h$ with probabilty $w_{t|t+1}^h$.

(3) The results are draws $\hat{\boldsymbol{\theta}}_{1:T} = \{\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2, \ldots, \boldsymbol{\theta}_T\}$ from the full conditional posterior, $p(\boldsymbol{\theta}_{0:T} \mid \mathbf{y}_{1:T}, \sigma_v^2, \mathbf{Q}, \boldsymbol{\Psi}, \boldsymbol{\mu})$.

A3. Sampling from $p(\mathbf{Q}, \sigma_v^2, \boldsymbol{\Psi}, \boldsymbol{\mu} \mid \boldsymbol{\theta}_{0:T}, \mathbf{y}_{1:T})$

Conditional on all the states and data $\{\boldsymbol{\theta}_{0:T}, \mathbf{y}_{1:T}\}$, the dynamic system (Equations (7) and (8)) becomes a multivariate (and nonlinear) system with parameters $\boldsymbol{\Psi} = \{\boldsymbol{\eta}, \{\boldsymbol{\varphi}_i, c_i, r_i\}_{i=1}^n\}$ and $\boldsymbol{\mu}$, and variance components $\{\mathbf{Q}, \sigma_v^2\}$. Therfore, applying MCMC procedures to estimate $p(\mathbf{Q}, \sigma_v^2, \boldsymbol{\Psi}, \boldsymbol{\mu} \mid \boldsymbol{\theta}_{0:T}, \mathbf{y}_{1:T})$ is relatively straightforward and is not detailed in this appendix (see Gamerman 1997, Allenby and Rossi 2003). Notably, the forgetting rate parameter vector $\boldsymbol{\eta}$ enters the multivariate system nonlinearly, and thus is estimated using the Metropolis-Hastings procedure described in Carlin and Louis (2000, pp. 155–157).

References

Allenby, G., P. Rossi. 2003. Bayesian statistics and marketing. *Marketing Sci.* **22**(3) 304–328.

Anderson, B. D., J. B. Moore. 1979. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, NJ.

Assmus, G., J. U. Farley, D. R. Lehmann. 1984. How advertising affects sales: Meta-analysis of econometric results. *J. Marketing Res.* **21**(February) 65–74.

Bass, F. M., D. G. Clarke. 1972. Testing distributed lag models of advertising effects. *J. Marketing Res.* **9**(3) 386.

Bass, F. M., N. Bruce, B. P. S. Murthi, S. Majumdar. 2007. Wearout effects of different advertising themes: A dynamic Bayesian model of the ad-sales relationship. *Marketing Sci.* **26**(2) 179–195.

Belch, G. E. 1982. The effects of television commercial repetition on cognitive response and message acceptance. *J. Consumer Res.* **9** 56–96.

Belch, G. E., M. A. Belch. 1998. *Advertising and Promotion: An Integrated Marketing Communications Perspective*. McGraw-Hill, Boston.

Blattberg, R. C., A. P. Jeuland. 1981. A micro-modeling approach to investigate the advertising-sales relationship. *Management Sci.* **27**(9) 988–1004.

Bucy, R. S., K. D. Senne. 1971. Digital synthesis of nonlinear filters. *Automatica* **7** 287–298.

Bultez, A. V., P. A. Naert. 1979. Does lag structure really matter in optimizing advertising expenditures. *Management Sci.* **25**(5) 454–465.

Burke, R., T. K. Srull. 1988. Competitive inference and consumer memory for advertising. *J. Consumer Res.* **15** 55–67.

Cacioppo, J. T., R. E. Petty. 1979. Effects of message repetition and position on cognitive response, recall and persuasion. *J. Personality Soc. Psych.* **37** 97–109.

Carlin, B. P., T. A. Louis. 2000. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall, New York.

Carter, C., R. Kohn. 1994. On Gibbs sampling for state space models. *Biometrika* **81** 541–553.

Chandy, R., G. Tellis, D. MacInnis, P. Thaivanich. 2001. When to say when: Advertising appeals in evolving markets. *J. Marketing Res.* **38**(November) 399–414.

Chessa, A. G., J. M. J. Murre. 2007. A neurocognitive model of advertisement content and brand name recall. *Marketing Sci.* **26**(1) 130–141.

Clarke, D. G. 1976. Econometric measurement of the duration of advertising effect on sales. *J. Marketing Res.* **13**(November) 345–357.

Doucet, A., N. de Freitas, N. Gordon, eds. 2001. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York.

Durbin, J., S. J. Koopman. 2001. *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford, UK, 51–53.

Fleming, J., C. Kirby. 2003. A closer look at the relation between GARCH and stochastic autoregressive volatility. *J. Financial Econometrics* **1**(3) 356–419.

Fruhwirth-Schnatter, S. 1994. Data augmentation and dynamic linear models. *Time Ser. Anal.* **15** 183–202.

Gamerman, D. 1997. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall, London, 124–132.

Gelfand, A. E., A. F. M. Smith. 1990. Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 972–985.

Gensch, D., U. P. Welam. 1973. An optimum budget allocation model for dynamic interactive market segments. *Management Sci.* **20**(2) 179–190.

Geweke, J. 1989. Bayesian inference in econometric models using Monte Carlo integration. *Econometrica* **57** 1317–1339.

Gill, P. E., W. Murray, M. Suanders. 2002. *User's Guide for SNOPT Version 7, Software for Large-Scale Nonlinear Programming*. http://www.sbsi-sol-optimize.com/manuals/SNOPTManual.pdf.

Godsill, S., A. Doucet, M. West. 2004. Monte Carlo smoothing for nonlinear time series. *J. Statist. Assoc.* **99**(465) 156–168.

Grass, R., W. H. Wallace. 1969. Satiation effect of TV commercials. *J. Advertising Res.* **1** 1–13.

Greenberg, A., C. Suttoni. 1973. TV commercial wearout. *J. Advertising Res.* **13** 47–54.

Hanssens, D., L. J. Parsons, R. L. Shultz. 1990. *Market Response Models, Econometric and Time Series Analysis*. Kluwer Academic Publishers, Norwell, MA.

Hitsch, G. J. 2006. An empirical model of optimal dynamic product launch and exit under demand uncertainty. *Marketing Sci.* **25**(1) 25–50.

Jorgensen, S., G. Zaccour. 2004. *Differential Games in Marketing*. Kluwer Academic Publishers, Norwell, MA.

Kent, R. J. 1991. Competitive versus noncompetitive television advertising. *J. Advertising Res.* **33** 40–46.

Kent, R. J., C. Allen. 1993. Does competitive clutter in television advertising interfere with recall and recognition of brand names and ad claims? *Marketing Lett.* **4**(2) 175–184.

Kim, C.-J. 2006. Time-varying parameter models with endogenous regressors. *Econom. Lett.* **91** 21–26.

Kim, S., N. Shephard, S. Chib. 1998. Stochastic volatility: Likelihood inference and comparison with ARCH models. *Rev. Econom. Stud.* **65** 361–393.

Kumar, A. 2000. Memory interference in advertising a replication and extension. *J. Consumer Res.* **30** 602–611.

Kumar, A., S. Krishnan. 2004. Interference effects in consumer memory for advertising: The role of brand familiarity. *J. Consumer Res.* **30** 602–611.

Leeflang, P., D. Wittink, M. Wedel, P. Naert. 2000. *Building Models for Marketing Decisions*. Kluwer Academic Publishers, Boston.

Liechty, J. C., D. K. H. Fong, W. S. DeSarbo. 2005. Dynamic models incorporating individual heterogeneity: Utility evolution in conjoint analysis. *Marketing Sci.* **24**(2) 285–293.

Little, J. D. C. 1979. Aggregate advertising model: The state of the art. *Oper. Res.* **27**(4) 629–667.

Liu, J. S., R. Chen. 1998. Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.* **93**(443) 1032–1043.

MacInnis, D. J., A. G. Rao, A. M. Weiss. 2002. Assessing when increased media weight of real-world advertisements helps sales. *J. Marketing Res.* **39**(November) 391–407.

Mitra, D., P. N. Golder. 2006. How does objective quality affect perceived quality? Short-term effects, long-term effects, and asymmetries. *Marketing Sci.* **25**(3) 230–247.

Naik, P. A., K. Raman. 2003. Understanding the impact of media synergy in multimedia communications. *J. Marketing Res.* **40**(4) 375–388.

Naik, P. A., M. K. Mantrala, A. G. Sawyer. 1998. Planning media schedules in the presence of dynamic advertising quality. *Marketing Sci.* **17**(3) 214–235.

Naik, P. A., K. Raman, R. S. Winer. 2005. Planning marketing-mix strategies in the presence of interaction effects. *Marketing Sci.* **24**(1) 25–34.

Neelamegham, R., P. Chintagunta. 2004. Modeling and forecasting the sales of technology products. *Quant. Marketing Econom.* **2**(3) 195–232.

Nerlove, M., K. Arrow. 1962. Optimal advertising policy under dynamic conditions. *Economica* **29**(May) 129–142.

Ogilvy, D., J. Raphaelson. 1982. Research on advertising techniques that work and don't work. *Harvard Bus. Rev.* **60**(4) 14–18.

Olney, T., M. Holbrook, R. Batra. 1991. Consumer responses to advertising: The effects of ad content, emotions, and attitude toward the ad on viewing time. *J. Consumer Res.* **17**(March) 440–453.

Pechmann, C., D. W. Stewart. 1990. Advertising repetition: A critical review of wear-in and wear-out. MSI Report 90-106, Cambridge, MA.

Perez, P., J. Vermaak, A. Blake. 2004. Data fusion for visual tracking with particles. *Proc. IEEE* **92**(3) 495–513.

Pieters, R., M. Wedel. 2000. Eye fixation on advertisements and memory for brands: A model and findings. *Marketing Sci.* **19**(4) 297–312.

Pieters, R., E. Rosbergen, M. Wedel. 1999. Visual attention to repeated print advertising: A test of scanpath theory. *J. Marketing Res.* **36**(4) 424–438.

Prentice, R. L. 1975. Discriminating among some parametric model. *Biometrika* **62**(3) 607–614.

Raman, K., P. A. Naik. 2004. Long-term profit impact of integrated marketing communications program. *Rev. Marketing Sci.* **2**(Article 8).

Ray, M. L., A. G. Sawyer, M. Strong. 1971. Frequency effects revisited. *J. Advertising Res.* **11**(1) 14–20.

Ristic, B., S. Arulampalam, N. Gordon. 2004. *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Artech House Publishers, Boston.

Rossi, P., G. Allenby, R. McCullogh. 2005. *Bayesian Statistics and Marketing*. Wiley Series in Probability and Statistics, Hoboken, NJ.

Rossiter, J., L. Percy. 1997. *Advertising Communications and Promotion Management*. McGraw-Hill, New York.

Schumann, D., D. Clemons. 1989. The repetition/variation hypothesis: Conceptual and methodological issues. *Adv. Consumer Res.* **16** 529–534.

Schumann, D., R. Petty, D. Clemons. 1990. Predicting the effectiveness of different strategies for advertising variation: A test of the repetition-variation hypotheses. *J. Consumer Res.* **17** 192–201.

Tellis, G. J. 2004. *Effective Advertising*: *Understanding When, How, and Why Advertising Works*. Sage Publications, London.

Thaivanich, P., R. K. Chandy, G. J. Tellis. 2000. Which ad works, when, where, and how often? Modeling the effects of direct television advertising. *J. Marketing Res.* **27**(February) 32–46.

Thorson, E., T. Page. 1988. Effects of product involvement and emotional commercials on consumers' recall and attitudes. S. Hecker, D. Stewart, eds. *Nonverbal Communications in Advertising*. Lexington Books, Lexington, MA, 111–126.

Unnava, H. R., R. Burnkrant. 1991. Effects of repeating varied ad executions on brand name memory. *J. Marketing Res.* **28**(November) 406–416.

Vakratsas, D. 2005. Advertising response models with managerial impact: An agenda for the future. *Appl. Stochastic Models Bus. Indust.* **21** 351–361.

Vakratsas, D., T. Ambler. 1999. How advertising works: What do we really know? *J. Marketing* **63**(January) 26–43.

Van Heerde, H. J., C. Mela, P. Manchanda. 2004. The dynamic effect of innovation on market structure. *J. Marketing Res.* **41**(2) 166–184.

West, M., J. Harrison. 1997. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York.