Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Correcting for Misspecification in Parameter Dynamics to Improve Forecast Accuracy with Adaptively Estimated Models

Ceren Kolsarici, Demetrios Vakratsas

MANAGEMENT SCIENCE
Vol. 61, No. 10, October 2015, pp. 2495–2513
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.2014.2027

Correcting for Misspecification in Parameter Dynamics to Improve Forecast Accuracy with Adaptively Estimated Models

Ceren Kolsarici
School of Business, Queen's University, Kingston, Ontario K7L 3N6, Canada, ckolsarici@business.queensu.ca

Demetrios Vakratsas
Desautels Faculty of Management, McGill University, Montreal, Quebec H3A 1G5, Canada, demetrios.vakratsas@mcgill.ca

Adaptive estimation methods have become a popular tool for capturing and forecasting changing conditions in dynamic environments. Although adaptive models can provide superior one-step-ahead forecasts, their application to multiperiod forecasting is challenging when the underlying parameter variation process is not correctly specified. The authors propose a methodology based on the Chebyshev approximation method (CAM), which provides a parsimonious substitute for the measurement updating process in the forecasting period, to help forecasters improve multiperiod accuracy in the case of parameter variation misspecification. In two empirical applications concerning the sales growth of new brands, CAM exhibits superior forecasting performance compared to a variety of benchmarks. CAM's properties are further explored through extensive simulations, which suggest that the proposed method is more likely to increase forecast accuracy when parameter variation is more systematic but misspecified because of uncertainty regarding its exact functional form.

Keywords: long-range forecasting; marketing; marketing mix; statistics; time series
History: Received March 2, 2012; accepted July 1, 2014, by Pradeep Chintagunta, marketing. Published online in *Articles in Advance* January 23, 2015.

1. Introduction

Markets have become increasingly dynamic, necessitating the use of models that can capture changing conditions caused by shifts in firm strategies and market response (e.g., Dekimpe and Hanssens 2010, Leeflang et al. 2009). Adaptive methods, such as the Kalman filter (KF), have been shown to provide superior alternatives to conventionally estimated methods for the study of dynamic markets (e.g., Bruce 2008, Goldenberg et al. 2009, Naik et al. 1998, Naik and Raman 2003, Neelamegham and Chintagunta 2004, Sriram et al. 2006, Xie et al. 1997). This is primarily due to their ability to capture dynamics via the formulation of state equations and allow for adjustments via measurement updating. A particular area of interest is the application of adaptive methods to demand forecasting (Neelamegham and Chintagunta 2004, Putsis 1998, Meade 1985, Engle et al. 1988), which has received notable attention by practitioners as well (e.g., EPRI 1983, Yelland et al. 2010). For example, the Electric Power Research Institute (EPRI) used an algorithm developed by Watson and Engle (1983) for a Kalman filter-based model to forecast residential demand for electricity. More recently, Yelland et al. (2010) used a dynamic linear model (DLM) to forecast demand for supply chain planning at Sun Microsystems by combining model- and judgment-based forecasts.

In multiperiod forecasting applications of adaptive methods, correct specification of the state equations representing the parameter dynamics is critical because of a lack of measurement updating, and hence the possibility of correcting for misspecification, over the forecasting horizon. However, correct specification can be challenging in the absence of theory or prior knowledge regarding parameter evolution. For example, in the case of radical innovations (e.g., electric cars or "wearable computing" such as watches or glasses), a lack of comparable, informative benchmarks could critically hinder the correct a priori specification of the state equations dynamics. The specification problem is further confounded because the states are not directly observed but rather estimated in the adaptive modeling framework. Thus, although correctly specifying dynamic parameter variation through a careful modeling approach would be the ideal option to produce high-quality multiperiod forecasts, this may not always be feasible because of the aforementioned challenges.

In the absence of a strong theoretical foundation or prior knowledge, researchers frequently adopt the random walk pattern to capture parameter dynamics (Engle et al. 1988, Neelamegham and Chintagunta 2004).

The random walk specification is a good option if one is only interested in temporally local parameter behavior and forecasts, since drastic changes in parameters are not expected in the short run (e.g., Neelamegham and Chintagunta 2004, West and Harrison 1997). In such cases, a lack of measurement updating is not fatal, and using the set of parameters estimated before the start of the forecasting window should work well for short-term forecasts. However, in dynamic environments, parameters are expected to change more drastically and systematically over time than the locally constant random walk pattern suggests. Hence a random walk-based model of parameter variation is likely to be misspecified, which will lead to inferior multiperiod forecasting performance since measurement updating is no longer available to provide a correction over the forecast window. Indeed, studies have shown that, as the horizon extends, the forecasting accuracy of adaptive methods is declining (Engle et al. 1988, Riddington 1993). The managerial implications of this shortcoming are also significant, because such forecasts are vital for managers who, for example, would like to assess the commercial potential of a new product, plan for inventory, or allocate marketing spending over time.

A solution to the problem would be to construct a parameter updating process over the forecasting period that would substitute for the measurement updating process used over the estimation period. In this paper we propose a multiperiod forecasting method for adaptively estimated models that seeks to minimize problems arising from the misspecification of parameter dynamics. We use Kalman filtering as our illustrative adaptive estimation methodology since it has been successfully employed in dynamic settings (Goldenberg et al. 2009, Neelamegham and Chintagunta 2004, Sriram et al. 2006). Specifically, we employ an extended Kalman filter with continuous state and discrete observations (henceforth, EKF(C-D); e.g., Kolsarici and Vakratsas 2010, Xie et al. 1997). Our context is new product sales growth, which represents a very challenging case because of the nonlinearity of the required diffusion model specification and constitutes possibly the most common manifestation of marketing dynamics (Leeflang et al. 2009). In addition, as previously discussed, the case of product innovations is challenging also in terms of correctly specifying the parameter path. However, our method can be applied to many different contexts such as mature markets changing because of macroeconomic factors or new competitor entry.

The proposed methodology is based on Chebyshev approximation and fully leverages the learning process stemming from the evolution of model parameters estimated during the preforecasting period (Meinhold and Singpurwalla 1983). Specifically, the method approximates the estimated parameter path with a function of time, which is then used to update the parameters in the forecasting period. In other words, the proposed approach is implemented in two separate stages. In the first stage the model is estimated by using a random walk specification for parameter variation. The estimated parameter path is expected to provide an accurate representation of the parameter dynamics since it uses measurement updating to adjust the parameter space, thus avoiding potential misspecification errors. In the second stage the methodology leverages the learning from the measurement updating during the estimation window, by approximating the estimated parameter path via Chebyshev expansion and using this approximation to produce multiperiod forecasts. In two applications, one for a new prescription pharmaceutical and another for a new hybrid car brand, the proposed method outperforms the most plausible benchmarks such as sampling from the parameter distribution (e.g., Meade 1985, Neelamegham and Chintagunta 2004) or using "off the shelf" methods and heuristics such as the last estimated parameter (Engle et al. 1988) and the average of estimated parameters. In addition, it exhibits superior forecasting performance when compared to an alternative adaptively estimated model with a simpler structure, mainly a dynamic linear model (DLM) version of the Koyck model.

We further study the properties of CAM through extensive simulations that show that it performs better when parameter variation follows a more systematic pattern but is misspecified as random walk. When, on the other hand, parameter variation follows a more random pattern, the proposed methodology does not offer an advantage over competing benchmarks. Hence our analysis provides clear boundary conditions regarding the merits of the proposed methodology. However, given that the marketing-mix parameter variation frequently exhibits systematic life-cycle-driven patterns (e.g., Bass et al. 2007, Narayanan et al. 2005), it should be noted that our methodology should improve multiperiod forecasting in a variety of applications. Because of forecasting applications of adaptive methods in practice, as already discussed, we expect the proposed method to be of value to both academics and practitioners. Furthermore, its intuitive nature, based on the need for substituting the measurement updating process during the forecast period, and its ease of application would further encourage the adoption of adaptive estimation methods for forecasting purposes.

2. Methodology

Before introducing the proposed methodology, we briefly overview the EKF(C-D) with continuous state and discrete observations, which has been shown to provide a superior alternative to competing methods in marketing problems, both in terms of its predictive performance and the insights offered (Kolsarici and Vakratsas 2010, Xie et al. 1997). EKF(C-D) is a recursive time-varying estimation technique whereby state variables are calculated iteratively through time and measurement updates. Prior means and covariances of the state variables estimated at the end of the time update are adjusted during the measurement update to get the posterior means and covariances. Although EKF(C-D) follows the same main structure as the standard Kalman filter (KF) (Harvey 1991), there are two major differences: (1) it enables continuous estimation of the process functions, as they are, without resorting to any discretization; and (2) it facilitates flexible handling of parameter dynamics that are augmented into the state vector. The general state–space formulation for EKF(C-D) is represented in the following equations:

State equation:

$$\frac{d\mathbf{y}(u)}{du} = \begin{bmatrix} \dfrac{dS(u)}{du} \\ \dfrac{d\boldsymbol{\beta}(u)}{du} \end{bmatrix} = \begin{bmatrix} f_1(S(u), \boldsymbol{\beta}(u), \mathbf{X}(u)) \\ f_2(\boldsymbol{\beta}(u), S(u)) \end{bmatrix} + \begin{bmatrix} w_1(u) \\ \mathbf{w}_2(u) \end{bmatrix}; \quad (1)$$

Measurement equation:

$$z_t = \mathbf{H}\mathbf{y}_t + v_t. \quad (2)$$

In the above equations, $\mathbf{y}$ denotes the augmented state vector consisting of mean sales, S, and the unknown parameter vector, $\boldsymbol{\beta}$; $\mathbf{X}$ is the vector of control variables such as marketing effort, u is time defined in a continuous domain, t denotes the discrete time points at which data are available, w_1 and $\mathbf{w}_2$ are process noise, z denotes the observed monthly sales, and v is the observation noise. It is assumed that $\mathbf{y}(0) \sim (\mathbf{y}_0, \mathbf{P}_0)$, the process and observation errors are white noise and not correlated (i.e., $(w_1, \mathbf{w}_2) \sim (0, \mathbf{Q})$ and $v \sim (0, R)$).

Although the EKF(C-D) algorithm iterates between the time and measurement updates, much like the standard KF, it works with the process equations in their original continuous form by integrating them over the time interval at which the data are available (i.e., $(t, t+1)$). We summarize the EKF(C-D) algorithm in the following steps:

Step 1. Using the initial values for the mean and covariance of the state, $(\mathbf{y}_0, \mathbf{P}_0)$, start the time update and calculate the state prior at time $t = 0$.

$$\hat{\mathbf{y}}^-_{t+1} = \int_t^{t+1} \frac{d\mathbf{y}}{du}\, du = \int_t^{t+1} f_y(\mathbf{y}(u), \mathbf{X}(\mathrm{u}))\, du, \quad \text{where } f_y = [f_S \;\; f_{\boldsymbol{\beta}}]'.$$

Step 2. Calculate the prior for the state covariance at time t, $\mathbf{P}^-_{t+1}$.

$$\mathbf{P}^-_{t+1} = \int_t^{t+1} \frac{d\mathbf{P}^i}{du}\, du = F(\mathbf{y}, u)\mathbf{P}^T + \mathbf{P}F^T(\mathbf{y}, u) + \mathbf{Q}, \quad \text{where } F(\mathbf{y}, u) = \left.\frac{\partial f_y[\mathbf{y}(u), \mathbf{X}(u)]}{\partial \mathbf{y}}\right|_{\mathbf{y}(u)=\mathbf{y}}.$$

Step 3. When the new observation $(t+1)$ becomes available, start the measurement update process. Compute the Kalman gain:

$$\varphi_{t+1} = \hat{\mathbf{P}}^-_{t+1}\mathbf{H}^T(\mathbf{H}\hat{\mathbf{P}}^-_{t+1}\mathbf{H}^T + \mathbf{R})^{-1}.$$

Step 4. Compute the posterior state estimates:

$$\hat{\mathbf{y}}_{t+1} = \hat{\mathbf{y}}^-_{t+1} + \hat{\varphi}_{t+1}(z_{t+1} - \mathbf{H}\hat{\mathbf{y}}^-_{t+1}). \quad (3)$$

Step 5. Calculate the posterior of the state covariance.

$$\hat{\mathbf{P}}_{t+1} = [\mathbf{I} - \hat{\varphi}_{t+1}\mathbf{H}]\hat{\mathbf{P}}^-_{t+1}.$$

Step 6. Set $t = t+1$ and go back to Step 1.

The Kalman filter updating can be viewed as an evolution of a series of "regression functions" of the state parameters on the one-step forecasting errors, where the Kalman gain in Step 3 plays the role of the "slope coefficient" and the prior estimate is the "intercept" (Meinhold and Singpurwalla 1983). The prior estimate can be viewed as an initial guess that the measurement update of the Kalman filter corrects depending on how well this guess has performed in predicting the next observation. In the case of multiperiod forecasting, this correction is not possible because of a lack of measurement updating, and the forecaster is left only with the guess. If the guess is based on the correct assumption (i.e., if the state parameter path is correctly specified), the use of prior information to update forecasts will be sufficient. For example, in the case of a random walk model of parameter variation, the natural option would be to use the last estimated parameter as the intercept estimate to forecast future values of the state parameters (Engle et al. 1988) or draw from the prior distribution (Meade 1985). In fact, using the last estimated parameter is the best forecast option if one is interested in short-term forecasts or there is no misspecification due to the random walk assumption. However, if random walk is not the correct specification, relying on the prior guess for long-term forecasting is no longer sufficient, and a substitute for the measurement correction process is necessary to improve accuracy.

To overcome the shortcomings of potential misspecification, we propose a methodology based on Chebyshev approximation (Davis 1975) that reliably updates the model parameter paths during the forecasting period.

In essence, the proposed method leverages the information in the estimation window to approximate the regression functions in Equation (3) using a polynomial expansion of time. More specifically, the Chebyshev approximation method (CAM) consists of three steps: (1) use Chebyshev polynomials to approximate each estimated parameter path, (2) employ these approximations to forecast the parameter paths in the holdout window, and (3) insert parameter forecasts into the EKF(C-D) algorithm to eventually obtain sales forecasts. Although for multiperiod forecasts one needs to eliminate the *measurement update* of the filter (because of the lack of data) and rely only on the *time update*, the proposed method essentially emulates the correction procedure of the filter's measurement update in the forecasting window through the approximations of the parameter paths in the estimation window.

There is a wide range of literature on function approximation (Achiezer 1956, Christensen and Christensen 2004, Davis 1975) that examines various methods including Taylor series, Hermite polynomials, Fourier series, and wavelet transforms. Applications of approximation theory, albeit limited, can also be found in the marketing literature. Most notably, Naik and Tsai (2000) use wavelet transforms to approximate advertising awareness using GRP data exhibiting irregular periodicity. Given the variety of functional approximation methods, we choose Chebyshev polynomials primarily for two reasons. First, the main idea behind Chebyshev approximation is very similar to that of the Taylor approximation, although the error structures are quite different. Whereas a Taylor series approximates a given function locally, leading to a rapid increase in errors as one moves away from the neighborhood of the point of approximation, CAM produces errors uniformly distributed over the entire interval (Gerald and Wheatley 2004). Second, CAM polynomials provide a nearly optimal solution where the maximum error of approximating a function is very close to the smallest possible by any other polynomial of the same degree. In other words, CAM is a rapidly converging power series that yields very accurate results with fewer terms, hence a smaller functional degree (Achiezer 1956). Moreover, CAM leads to particularly flexible functions that are useful for effectively capturing dynamic parameter paths.

Chebyshev polynomials are based on the notion that a continuous function, f, defined in an interval $[a, b]$ can be represented by a polynomial $f(x) \approx \sum_{n=0}^{\infty} G_n T_n(x)$, in such a way that at every x in $[a, b]$, the error due to approximation is controlled and distributed uniformly over the interval where $\mathbf{G}$ represents the vector of approximation coefficients estimated in the process. We use the trigonometric formulation where Chebyshev polynomials are characterized by the identity, $\mathrm{T}_n(\xi) = \cos(n\cos^{-1}(\xi))$ for ξ defined in $[-1, 1]$. Hence, the polynomials are defined by a recursive relationship expressed as follows:

$$\begin{gathered} T_0(\xi) = 1, \\ T_1(\xi) = \xi, \\ \dots \\ T_n(\xi) = \cos(n\cos^{-1}(\xi)), \\ \dots . \end{gathered} \tag{4}$$

Truncating the series expansion of the approximation at the nth term gives us the nth degree approximation for $f(\xi)$. In the context of our problem, we use CAM to approximate the parameter paths estimated in the preforecast window. Specifically, for each parameter β, the procedure applies as follows.

1. We use a linear transformation to replace each time point, t, in $[t_0, \bar{t}]$ (i.e., with t_0 being the time from which the focal parameter is defined, and $\bar{t}$ being the maximum point of the estimation window[1]) by ξ in the normalized interval $[-1, 1]$ using the transformation equation in (5):

$$\xi = \xi(t) = -1 + 2(t - t_0)/(\bar{t} - t_0). \tag{5}$$

The variables at the end of the linear transformation are stacked in the vector, $\boldsymbol{\xi}_{(\bar{t}-t_0+1)\times 1}$.

2. We produce Chebyshev expansion in (4) for the vector $\boldsymbol{\xi}$, which yields a $(\bar{t} - t_0 + 1)\times(n+1)$ matrix $\mathbf{p}(\boldsymbol{\xi})$, shown in Equation (6), of stacked row vectors of Chebyshev expansion for each element of $\boldsymbol{\xi}$:

$$\mathbf{p}(\boldsymbol{\xi}) = \begin{bmatrix} 1 & \xi(1) & \cos(2\cos^{-1}(\xi(1))) & \cdots & A' \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \xi(\bar{t} - t_0 + 1) & \cdots & \cdots & A'' \end{bmatrix} \tag{6}$$

with last column entries of the form

$$A' = \cos(n\cos^{-1}(\xi(1))), \quad A'' = \cos(n\cos^{-1}(\xi(\bar{t} - t_0 + 1))).$$

Concisely,

$$(\mathbf{p}(\boldsymbol{\xi}))_{ij} = \left[\cos(j\cos^{-1}(\xi(i)))\right], \quad 1 \le i \le \bar{t} - t_0 + 1, \quad 0 \le j \le n.$$

We then regress the vector of estimated parameter values, $\hat{\boldsymbol{\beta}}_{(\bar{t}-t_0+1)\times 1}$, on $\mathbf{p}(\boldsymbol{\xi})$ obtained in the previous step. This basic regression gives us the vector of coefficients ($\mathbf{G}$) that describe the parameter evolution with respect to the time horizon, which is expressed as follows:

$$\hat{\boldsymbol{\beta}}_{(\bar{t}-t_0+1)\times 1} = p(\boldsymbol{\xi})_{(\bar{t}-t_0+1)\times(n+1)} \mathbf{G}_{(n+1)\times 1} + \varepsilon_{(\bar{t}-t_0+1)\times 1}. \tag{7}$$

where $\boldsymbol{\varepsilon} \sim N(0, \mathbf{E})$ is the error term that captures the changes in $\hat{\boldsymbol{\beta}}$ that cannot be explained by Chebyshev approximation.

[1] For instance, in our pharmaceutical data set, which is described in detail in the data section (§3.1), direct-to-consumer advertising is not employed in the first two years of the observation period; thus, advertising parameters are not defined for that range.

We repeat Steps 1–3 for each parameter and store corresponding projections for all parameters obtained in Step 3 for the entire forecast horizon. Using this set of coefficients and the holdout Chebyshev expansion matrix, $\mathbf{p}(\boldsymbol{\xi})^F$, we produce k-step-ahead parameter forecasts ($k = 1, \ldots, K$).[2]

$$\hat{\beta}^F_{\bar{t}+k} = p(\boldsymbol{\xi})^F \mathbf{G}. \tag{8}$$

It should be noted that the method does not account for the uncertainty in the parameter estimators (left-hand side of Equation (7)). However, because of the measurement updating process, we expect the estimates to be reliable. Indeed, the confidence intervals calculated for the parameter estimates in our empirical application (not shown here), are consistently narrow, confirming our expectations. In addition, we use Monte Carlo simulations to calculate confidence intervals and thus assess the reliability of the forecasted parameter paths. Specifically, we start m independent Chebyshev algorithms by resampling from the posterior distribution of the vector of parameters, $\boldsymbol{\beta}_t$, at each t in $[t_0, \bar{t}]$ for $m = 10{,}000$ times. This enables us to obtain an empirical density of the predicted and forecasted parameter paths. We derive 95% confidence intervals for CAM using the 2.5th and 97.5th percentiles of the corresponding simulated paths. Figure 1 illustrates CAM's fit and 95% confidence bounds for various model parameters in our empirical application.

3. Empirical Application

3.1. Data

We employ data from two product categories: hybrid cars and a newly established therapeutic category of prescription pharmaceuticals. Both data sets contain monthly information on sales and marketing spending for the leading brands since their launch. In the hybrid car category we focus on Toyota Prius for the period between January 2001 and December 2008. Information includes unit sales, dollar spending on network TV, spot TV, and online advertising. The second data set pertains to a novel therapeutic class of lifestyle drugs and covers the period between March 1999 and May 2006. Because of a confidentiality agreement with the source, the brand name and category for the prescription pharmaceutical cannot be disclosed. New as well as refill prescription data along with marketing mix information on category- and brand-specific direct-to-consumer advertising and physician journal advertising (measured in dollars) and detailing (measured in units) are available. The descriptive statistics for both data sets can be found in Tables 1 and 2, respectively. Figures 2 and 3 depict sales for the Prius and the focal prescription pharmaceutical brand.

[2] Chebyshev expansion matrix for the forecasting horizon, $p(\xi)^F$, is constructed the same way as in Equation (6) for the $[\bar{t}+1, \bar{t}+k]$ window.

Both data sets are particularly suitable for testing multiperiod forecasting accuracy by using adaptively estimated diffusion models. They both concern brands in novel product categories characterized by dynamic market conditions such as growth, word of mouth, and changes in marketing spending. Furthermore, planning for the marketing mix included in the data is done over multiple periods. For example, advertising spending is allocated monthly over a typical period of 12 months. Thus, multiperiod forecasts are essential for managers making decisions for the marketing mix instruments for which information is included in our data.

3.2. Modeling Framework

We adopt a diffusion approach to model brand sales since we expect growth and word of mouth to play an important role in shaping the underlying dynamics. For the pharmaceutical data set, we employ a three-segment diffusion model following Kolsarici and Vakratsas (2010). Total sales result from a combination of *trials* (T), *renewals* (R), and *refill* (L) prescriptions, each expressed as follows:

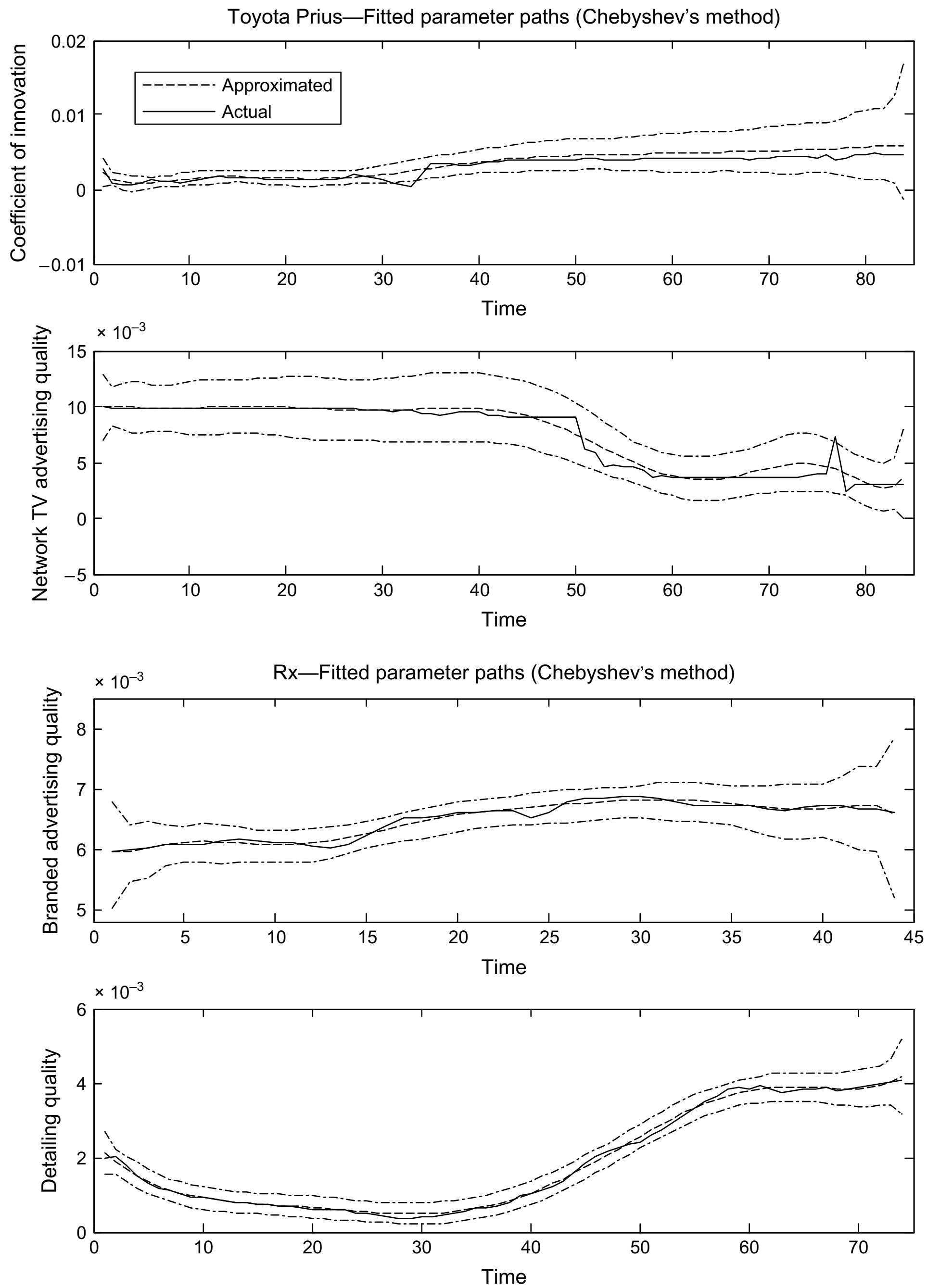
$$\begin{aligned} T(u) &= \frac{dT^c(u)}{du} = \left[p(u) + \frac{q(u)}{M}T^c(u)\right][M - T^c(u)] \\ &\quad \cdot \exp\left(\alpha(u)\sqrt{\mathbf{X}(u)}\right), \\ R(u) &= \frac{dR^c(u)}{du} = \rho_1(u)T^c(u), \\ L(u) &= \frac{dL^c(u)}{du} = \rho_2(u)T^c(u), \end{aligned}$$

where T^c, R^c, and L^c represent cumulative trials, renewals, and refills, respectively; p, q, and M are the coefficient of innovation, coefficient of imitation, and market potential, respectively; and $\mathbf{X}$ is the vector of marketing mix variables in lagged square-root form to account for carryover effects and diminishing returns to scale. For *trial* (T) prescriptions, we use the well-known Bass model formulation (Bass 1969) with separable marketing mix effects (Bass et al. 1994, Parker and Gatignon 1994). The second and third equations represent *renewal* (R) and *refill* (L) prescriptions, respectively, for which we assume that at each time point they constitute a dynamic percentage (ρ_{1t} and ρ_{2t}) of cumulative trial prescriptions up to that point. The data set contains separate information on new (i.e., trials plus renewals) and refill prescriptions, hence both ρ_{1t} and ρ_{2t} are identifiable.

Consequently, we express total sales, S, as the sum of sales from trials, refills, and renewals:

$$\begin{aligned} S(u) &= \frac{dT^c(u)}{du} + \frac{dR^c(u)}{du} + \frac{dL^c(u)}{du} + w_1(u) \\ &= \left[p(u) + \frac{q(u)}{M}T^c(u)\right][M - T^c(u)]\exp\left(\alpha(u)\sqrt{\mathbf{X}(u)}\right) \\ &\quad + \rho_1(u)T^c(u) + \rho_1(u)T^c(u) + w_1(u). \end{aligned} \tag{9}$$

Figure 1 **Sample Tracking Performance of CAM**



In the EKF(C-D) algorithm, we estimate Equation (9) in its continuous form by integrating over the interval at which data is available. For the details of the three-segment model and its estimation via the Kalman filter, the interested reader might refer to Kolsarici and Vakratsas (2010).

A more simplified model structure is used for the hybrid car brand (Prius), which omits the last two terms of Equation (9) and focuses on the extended Bass model with marketing mix effects corresponding to first-time purchases. This is a reasonable assumption given our data range and the longer purchase cycles of automobiles. Hence, hybrid car sales are expressed as

$$S(u) = \frac{dT^c(u)}{du} = \left(\left[p(u) + \frac{q(u)}{M}T^c(u)\right][M - T^c(u)] \cdot \exp\left(\alpha(u)\sqrt{\mathbf{X}(u)}\right) + w_1(u)\right). \quad (10)$$

For estimation purposes we employ EKF(C-D), introduced earlier in the methodology (§2). The

Table 1 Descriptive Statistics (Toyota Prius)

	Sales (units)	Network TV adv. ($1,000)	Spot TV adv. ($1,000)	Online adv. ($1,000)
Mean	6,971.52	325.80	234.68	84.64
Median	6,205	0	0.65	8.60
Std. dev.	5,546.36	1,683.83	486.00	221.33
Kurtosis	0.17	47.93	8.17	13.04
Skewness	0.82	6.66	2.77	3.70
Minimum	112	0	0	0
Maximum	24,009	13,842.8	2,602.6	1,082.4
Count	96	96	96	96

Table 2 Descriptive Statistics (Prescription Pharmaceutical Rx)

	Sales (units)	Branded adv. ($1,000)	Generic adv. ($1,000)	Physician journal adv. ($1,000)	Detailing (Units)
Mean	74,674.9	226.9	65.7	64.2	4,621.2
Median	74,127.0	41.0	0.0	43.1	4,466.5
Std. dev.	16,370.3	302.3	145.8	73.9	1,829.3
Kurtosis	0.19	0.25	7.10	2.09	−0.30
Skewness	−0.31	1.17	2.71	1.41	0.46
Minimum	20,600	0	0	0	998
Maximum	106,962	1,141.7	730.6	317.7	9,059
Count	86	86	86	86	86

corresponding augmented state vector can then be expressed as

$$\frac{d\mathbf{y}}{du} = \left[\frac{dT^c}{du} \quad \frac{d\boldsymbol{\beta}}{du}\right]'. \tag{11}$$

The first element of the augmented state vector in Equation (11) corresponds to the sales evolution captured in Equations (9) and (10), for each application (pharmaceutical and car), respectively; $\boldsymbol{\beta}$ is the vector of model parameters including p, q, and M for diffusion and α for marketing effectiveness (i.e., $\boldsymbol{\beta} = [p \; q \; M \; \boldsymbol{\alpha}]$). For the pharmaceutical data, the parameter vector $\boldsymbol{\beta}$ also

Figure 2 Toyota Prius Sales Evolution

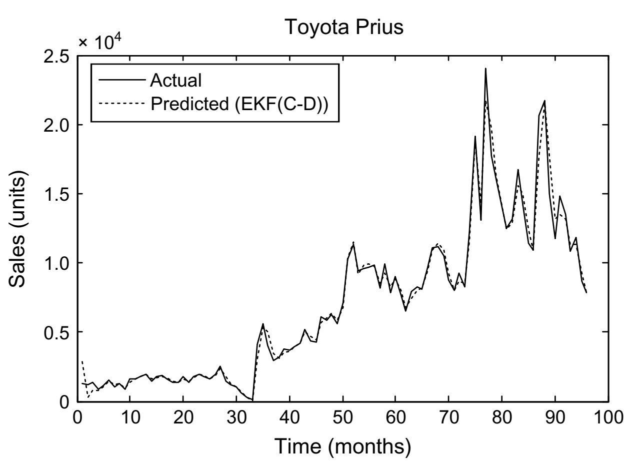
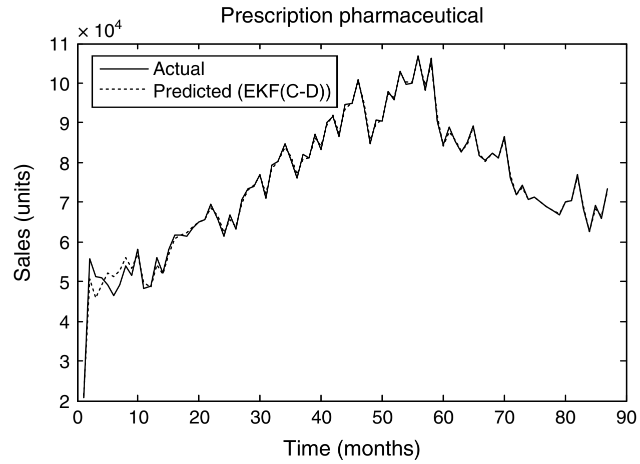


Figure 3 Prescription Pharmaceutical Sales Evolution



includes the renewal and refill rates, ρ_{1t} and ρ_{2t}, respectively. In addition, we use two observation equations for this data set, one for new and one for refill prescriptions. Following most adaptive modeling approaches, we assume that $\boldsymbol{\beta}$ follows a random walk pattern ($d\boldsymbol{\beta}/du = \sigma d\mathbf{w}_2$) because of the uncertainty regarding the exact form of parameter variation. However, potential misspecification due to the random walk assumption would present challenges in the case of multiperiod forecasting as discussed earlier. The issue of (mis)specification of the parameter variation pattern will be revisited in the empirical application and examined in detail in our simulation section when we further examine the performance of CAM versus competing methods in different scenarios.

3.3. Benchmark methods

We compare the performance of CAM to a variety of benchmark methods outlined below.

3.3.1. Random Parameter Draw (RPD). This method makes use of the prior estimates of the parameter means and covariances obtained at the end of each KF time update within the forecast window. Following the continuous-time extended Kalman filter assumption that system and measurement noise have an approximate normal distribution (Harvey 1991), RPD assumes that the parameter vector at time t follows the multivariate normal distribution below, where the superscript"−" refers to the *prior* estimates.

$$\hat{\boldsymbol{\beta}}_t \sim MVN(\hat{\boldsymbol{\beta}}_t^-, \hat{\mathbf{P}}_t^-). \tag{12}$$

To obtain a holdout forecast at time point t, we first sample from Equation (12) to get $\hat{\boldsymbol{\beta}}_t$. The drawn parameters, $\hat{\boldsymbol{\beta}}_t$, and the partial prior covariance, $\hat{\mathbf{P}}_t^-$, are then used as inputs to the time update process which leads to (a) the forecasted sales through Equations (9) and (10) and (b) the prior mean and covariance of the parameters at time $t+1$. Then, the same procedure is repeated for $t+1$. The distribution in Equation (12) is updated with these newly calculated "priors," $\hat{\boldsymbol{\beta}}_{t+1}^{-}$ and $\hat{\mathbf{P}}_{t+1}^{-}$. The projected parameter vector $\hat{\boldsymbol{\beta}}_{t+1}$, resulting from the draw, is used to calculate the sales forecast at $t+1$. Thus, the parameter path projection for this benchmark is based on the sampling of the parameters from their assumed distribution. In other words, RPD focuses on updating the intercept term of Equation (3) while ignoring the slope coefficient-related term, which is based on the prediction error.

It should be noted that this method is similar to the one proposed by Meade (1985) for the calculation of parameter confidence bounds. A comparable approach has also been employed by Neelamegham and Chintagunta (2004) in a Bayesian context where the authors use a hierarchical Bayesian model to investigate the time-varying impact of marketing efforts on the sales of technology products. They define the evolution of model parameters with a discrete-time (rather than continuous as is the case in our application) random walk model, and use this relationship together with its variance–covariance estimate to forecast parameter values and the resulting sales in the holdout sample. The forecast window in Neelamegham and Chintagunta (2004) is fixed and ranges from one to five months taken at two distinct time points in the estimation sample (i.e., 1.5 and 3 years into observation). Unlike Neelamegham and Chintagunta (2004), we follow an expanding window approach where our forecasts start from as early as the end of the first year of the observation period with a range of up to 24 months ahead.

3.3.2. Last Estimated Beta (LEB). This is a heuristic method that uses the EKF(C-D)-based estimated parameters obtained in the last observation before the start of the forecast window as parameter forecasts. Similarly to RPD, LEB focuses on the intercept of Equation (3) but in a deterministic manner since it relies on the last estimated parameter rather than drawing from the prior distribution as RPD does. Whereas with LEB there is essentially no parameter updating, which may be considered a disadvantage, it is also not subject to random draw error. This method is expected to work well when random walk is the true pattern of parameter variation (Engle et al. 1988, West and Harrison 1997) and in short-term forecasting where drastic changes in the parameters are not expected. It has been used by EPRI (1983), Engle et al. (1988), and Engle and Watson (1985).

3.3.3. Mean Beta (MB). This method uses the means of all previously estimated parameters to produce parameter value forecasts. MB effectively ignores the dynamics of the state parameters and can be viewed as the equivalent of using a static forecasting method. It is expected to work well when parameter variation does not exhibit any systematic pattern, in which case the estimated mean becomes the best guess.

3.3.4. Dynamic Linear Model (DLM). The final benchmark is a dynamic linear model (DLM) embedded into an Markov chain Monte Carlo (MCMC) Gibbs sampler (Gelman et al. 2004). Specifically, we use an extension of the simple first-order autoregressive advertising model (Naik and Raman 2003, Palda 1964), effectively in a Koyck model structure, which is commonly used in practice (Bucklin and Gupta 1999). The DLM algorithm recursively moves between forward filtering and backward sampling (FFBS) to get the state estimates at each time point (West and Harrison 1997). Forward filtering is essentially the same as the updating equations of the Kalman filter, and backward sampling enables retrospective analysis by which we can update the posterior belief about sales at $(t-1)$, given the information at t. The state space representation of the DLM is presented in the following set of equations:

State equation:

$$\begin{gathered} S_t = \lambda S_{t-1} + \boldsymbol{\alpha}\sqrt{\mathbf{X}_{t-1}} + w_t, \\ w_t \sim N(0, Q); \end{gathered} \tag{13}$$

Observation equation:

$$z_t = S_t + \nu_t, \quad \nu_t \sim N(0, R).$$

Here, S and z are the mean and observed sales values, respectively; λ accounts for the carryover effects; and $\mathbf{X}$ is the vector of marketing mix efforts. We assume conjugate prior distributions for all the nonstate parameters and covariance terms. More specifically, we assume inverse Gamma priors for Q and R, normal prior for λ, and multivariate normal prior for the marketing mix effectiveness vector $\boldsymbol{\alpha}$. By using a direct Gibbs sampling approach (Gelfand and Smith 1990), we iteratively resample a series of conditional posteriors to obtain the state and nonstate parameters. The estimated nonstate parameters (i.e., λ and $\boldsymbol{\alpha}$) are then used in the holdout window to forecast sales. Thus, the Koyck model in Equation (13) represents a relatively simple, yet adaptively estimated, benchmark method.

3.4. Results

We track the forecasting performance of the proposed and benchmark methods for two different forecast window sizes, 12 and 24 months, throughout the product life cycle. We start with an estimation window of $\bar{t} = 12$ months, and then we expand the window by one month at a time to allow for the gradual introduction of additional information. More specifically, for each iteration we use the first $\bar{t}$ observations for estimation (i.e., estimation window: $[1, 2, \ldots, \bar{t}]$) and forecast the next K observations (i.e., forecast window: $[\bar{t}+1, \bar{t}+2, \ldots, \bar{t}+K]$ with $K \in \{12, 24\}$ depending

Table 3 **Rx Data Mean Long-Range Forecasting Errors**

	Proposed	Benchmarks			
	CAM	RPD	LEB	MB	DLM
		12-month forecast			
MAPD	9.34	10.55	12.72	14.35	26.22
MAD	7.33E+03	8.27E+03	9.67E+03	1.07E+04	2.14E+04
MSE	8.77E+07	1.04E+08	1.46E+08	1.80E+08	7.18E+08
		24-month forecast			
MAPD	13.97	15.21	19.32	22.27	28.37
MAD	1.11E+04	1.23E+04	1.51E+04	1.72E+04	2.39E+04
MSE	2.10E+08	3.01E+08	3.85E+08	4.94E+08	8.96E+08

Note. MAPD, mean absolute percentage deviation; MAD, mean absolute deviation; MSE, mean squared error.

Table 4 **Prius Data Mean Long-Range Forecasting Errors**

	Proposed	Benchmarks			
	CAM	RPD	LEB	MB	DLM
		12-month forecast			
MAPD	46.48	69.75	67.65	70.46	76.81
MAD	1.90E+03	3.42E+03	2.55E+03	4.11E+03	8.06E+03
MSE	1.21E+07	2.38E+07	1.69E+07	2.83E+07	9.15E+07
		24-month forecast			
MAPD	66.11	72.66	76.27	78.48	85.54
MAD	3.11E+03	3.23E+03	3.63E+03	4.63E+03	8.75E+03
MSE	1.63E+07	2.27E+07	2.46E+07	3.29E+07	1.02E+08

Note. MAPD, mean absolute percentage deviation; MAD, mean absolute deviation; MSE, mean squared error.

on the length of the forecast window). Thus, as the estimation window expands by one month, the forecast window slides across the data range.

Tables 3 and 4 summarize the multiperiod forecasting performance of the proposed and benchmark methods. Since we use an expanding estimation window approach, the performance statistics reported represent the average errors across all estimation windows for 12- and 24-month forecasts. To provide a better idea of the evolution of forecasting performance with the expansion of the estimation window, we plot the mean absolute percentage deviations (MAPDs) over all estimation windows expanding from the beginning to the end of the observation period (Figures 4 and 5). A careful examination of Tables 3 and 4 suggests that the proposed method of Chebyshev approximation (CAM) outperforms all other benchmarks in both data sets and for both window lengths. This should be attributed to CAM's leveraging of the parameter evolution knowledge acquired in the preforecasting estimation period. As suggested by Figures 4 and 5, the superiority of CAM in comparison to its counterpart

Figure 4 **12-Month Forecast Error Evolution for Rx Data**

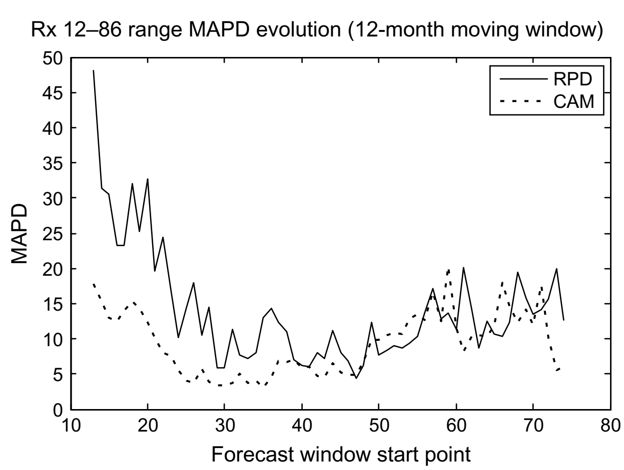
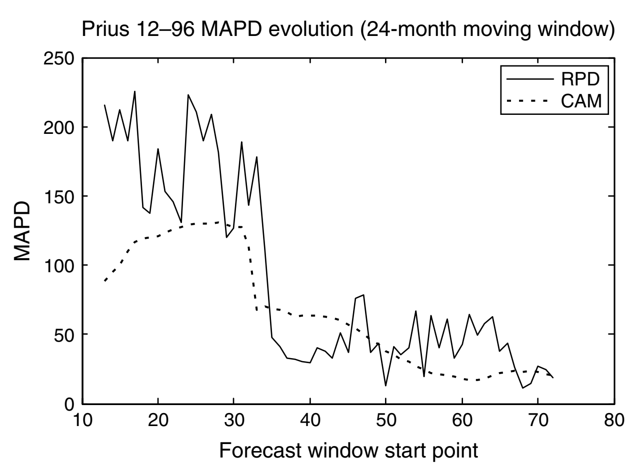


Figure 5 **24-Month Forecast Error Evolution for Prius**



RPD is particularly noticeable when the estimation windows are smaller, i.e., early in the life cycle, when there are limited data and good quality multiperiod forecasts are paramount for the manager. This is consistent with the findings of Neelamegham and Chintagunta (2004), possibly because in the latter stages of the life cycle parameters exhibit higher stability; hence differences among forecasting methods tend to diminish.

Thus, CAM appears to be a natural choice and a potent substitute for the measurement update process during the forecasting period. This should encourage researchers and practitioners who wish to combine accurate dynamic parameter estimation and short-term forecasts with good longer-term forecasts. RPD's lagging performance can be due to its higher susceptibility to the variance of the distribution and the inability to capture any emerging systematic patterns suggested by the estimated parameter path. This may partially explain the occasionally inferior forecasting performance of RPD compared to LEB (e.g., 12-month forecasts for Prius) as well as RPD's high forecasting errors in the early windows. CAM's superiority highlights the importance of generating a mechanism that could be used as a substitute for the measurement update process to capture the parameter dynamics over the forecasting period.

In addition, the overall better forecasting performance of the nonlinear Kalman filter-based models compared to a simpler DLM Koyck model challenges the prevailing view of simple models forecast better (Meade and Islam 2001), much in the spirit of the seminal study by Engle et al. (1988). Thus, a policy of using more structured models for estimation and short-term forecasts and simpler models for longer-term forecasts will not be optimal, because there are gains to be made from using more structured models for longer-term forecasts. This should further reinforce the use of sophisticated, adaptively estimated models by practitioners and researchers alike.

It should be noted that, although the mean absolute percentage deviations (MAPDs) for Prius are somewhat high, the available information is monthly, which in the context of the car category should be considered finely grained and thus susceptible to higher noise. Even so, the errors are comparable to those reported by Xie et al. (1997) and Lenk and Rao (1990) for one-step-ahead forecasts. In addition, some of the forecasts are quite early; for example, forecasts for the 12- to 24-month period correspond only to the second year of the launch of a radical innovation for which it would be difficult to obtain an accurate longer-term forecast. However, despite the challenges of the application context, the errors decrease relatively fast (third to fourth year of launch), improving the reliability of the forecasts while they are still managerially relevant (Figures 4 and 5). Hence, the consistently superior forecasting performance of the proposed methodology even under stringent conditions lends further credibility to CAM.

Two further observations from the empirical findings are worth discussing: the lagging performance of LEB despite the random walk assumption, and the pattern of parameter variation estimated from the data and illustrated in Figure 1. The random walk assumption suggests that the parameter path is determined by the previous value of the parameter, positioning LEB as the best forecasting method since it relies on the last estimated value. However, its lower forecasting accuracy does not appear to validate the random walk specification. This is indeed corroborated by the empirical findings that indicate, as shown in Figure 1, that parameter paths may follow more systematic patterns such as cyclical or unimodal. Such patterns are not uncommon in the literature on communications effects where cyclical (Bruce 2008, Naik et al. 1998), decreasing (Narayanan et al. 2005, Osinga et al. 2011) and nonmonotonic (Kolsarici and Vakratsas 2010, Vakratsas and Kolsarici 2008) paths have been reported. We further pursued this issue by calculating scale-adjusted variances of the sales prediction errors, effectively representing the observation errors, and those of the first differences of the estimated parameters, effectively representing the estimated errors of the random walk specification. If random walk is the correct specification, then the variance of the first differences of the estimated parameters should be much smaller than the observation error variance (West and Harrison 1997). In both empirical applications, the scale-adjusted variances of the first differences of the estimated parameters were similar to those of the sales prediction errors, suggesting that the locally constant random walk model may not be the appropriate specification.

The possibility of parameter variation misspecification coupled with the superior performance of CAM in the empirical application tends to support our prior arguments regarding the significance of CAM due to its ability to substitute the measurement update process during the forecasting period when state parameters follow a pattern more systematic than the assumed random walk. In the following section we further investigate the properties of CAM.

4. Validation: Exploring Boundary Conditions Through Simulations

We use simulations to provide a more comprehensive assessment of CAM properties and investigate boundary conditions concerning its performance versus competing methods. We focus on nonlinear (diffusion model) specifications since they were shown to be clearly superior to simpler (linear) ones such as DLM. Since the main contribution of CAM is improving forecasting quality through parameter updating in the absence of measurement information, the pattern of parameter variation becomes the focal manipulation of the simulations. We also examine the issue of the choice of polynomial degree (n) for CAM. We further benchmark CAM by considering an alternative method of parameter variation approximation based on cubic splines (CS), which provides another flexible way of capturing dynamic parameter paths. Instead of approximating the whole parameter path with a single polynomial, CS uses multiple third-degree polynomials fitted in a piecewise manner over the estimation window. We first provide the details of the simulation manipulations, then proceed with discussion of the choice of n, and continue with the forecast accuracy analysis.

4.1. Simulation Manipulations

We examine an extensive set of parameter paths that cover the entire spectrum of variation ranging from systematic to purely random (Winer 1979). More specifically, we consider the following data-generating process (DGP) scenarios.

1. *Sinusoidal*: This is a systematic cyclical variation pattern, representing the case of "extreme" systematic variation. Similar patterns of parameter variation have been reported in Naik et al. (1998) and Bass et al. (2007) in the context of dynamic advertising effects reflecting wear-out and restoration phenomena.
2. *Quadratic*: This is a systematic pattern that allows for possible nonmonotonic variation over time due to changes in preferences and the entry of new consumer groups, e.g., early versus late adopters (Vakratsas and Kolsarici 2008).
3. *Autoregressive* (AR(1)): This is a sequential pattern combining systematic and random variation.
4. *Random Walk* (RW): This is similar to AR(1) in terms of the combination of systematic and random variation; however it is more conducive to LEB as a forecasting method because of its dependence on the previous parameter value.
5. *Random Coefficients* (RC): This is a purely random pattern that lacks any systematic variation component.

Thus, the above scenarios exhaust all three parameter variation possibilities discussed in Winer (1979): (1) systematic (sinusoidal and quadratic); (2) sequential (AR(1) and RW), which combines systematic and random variation; and (3) random (RC).

In terms of the sales-generating process we use the simplified diffusion model of Equation (10). Without loss of generality we limit the number of marketing mix variables to two in order to facilitate interpretation and exposition of the results. Much like in the context of our empirical study, we focus on communications-related marketing variables and choose coefficients that produce elasticities in the range of reported meta-analytic values (e.g., Sethuraman et al. 2011). Similarly, we fix the monthly diffusion parameters of innovation, imitation, and market potential and set them to 0.003, 0.025, and 1e09, respectively, following the meta-analytic values reported in Sultan et al. (1990). We simulate 120 monthly data points to ensure the presence of both pre- and postpeak sales periods. Figure 6 depicts the simulated sales values for each DGP in comparison to a baseline that reflects the effect of the diffusion parameters only. Except for the deviations due to the difference in the draws and the patterns of parameter variation, the sales patterns across the different DGP scenarios are comparable. Table 5 summarizes

Figure 6 Stimulated Sales Data

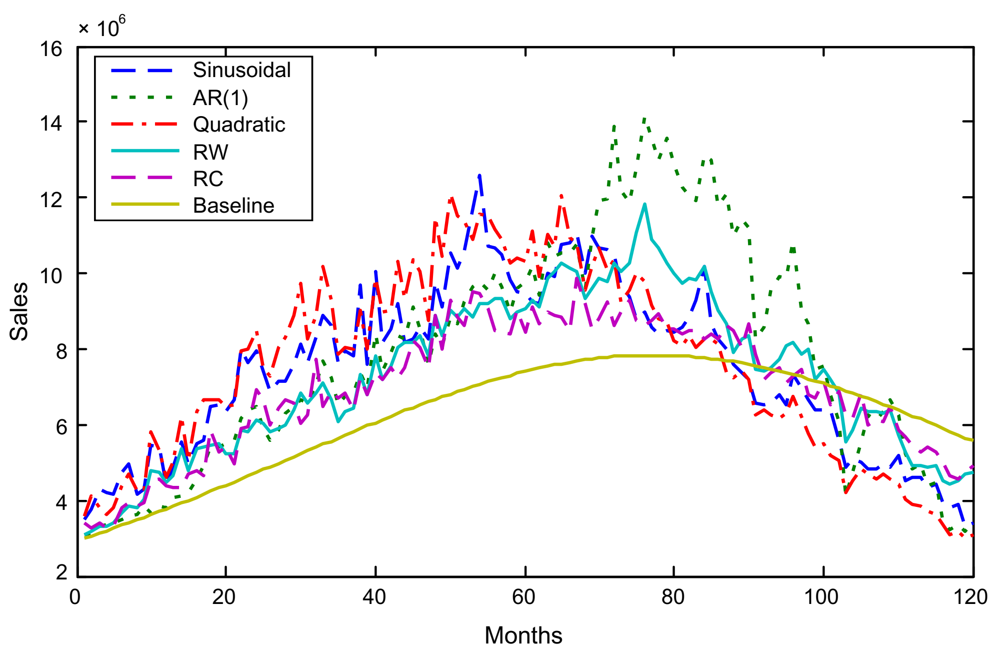
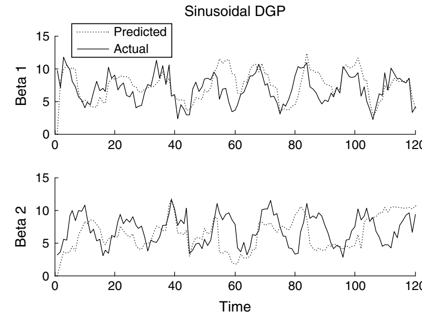
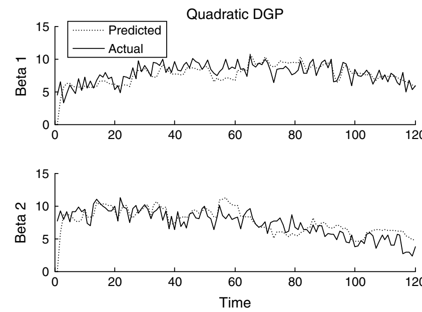
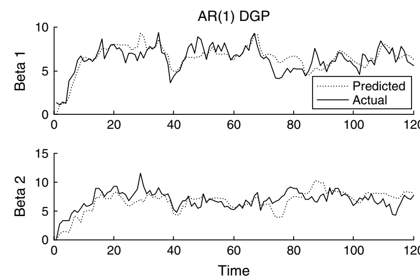
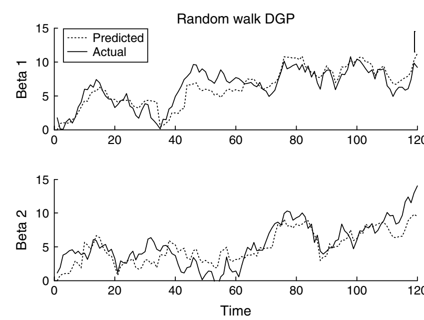
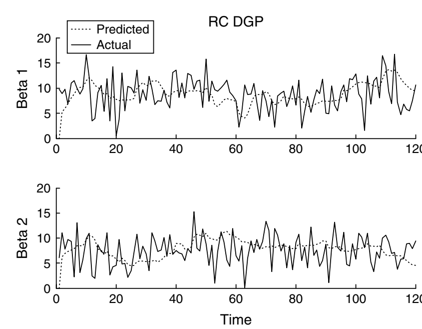


Table 5 Simulation Summary

	Data-generating process (DGP)	Formulation and hyperparameters	Simulated parameter paths
Systematic variation	Sinusoidal	$\boldsymbol{\beta}_t = \mathbf{a}_0 + \mathbf{a}_1 \sin(\mathbf{a}_2(t + \mathbf{a}_3))$ $\mathbf{a}_0 = [7, 6]$ $\mathbf{a}_1 = [3, 3.2]$ $\mathbf{a}_2 = [0.4, 0.3]$ $\mathbf{a}_3 = [0, 12]$	
	Quadratic	$\boldsymbol{\beta}_t = \mathbf{a}_3(\mathbf{a}_0 + \mathbf{a}_1 t^2 + \mathbf{a}_2 t) + \mathbf{e}$ $\mathbf{a}_0 = [1e04, 1.8e04]$ $\mathbf{a}_1 = [-2, -1]$ $\mathbf{a}_2 = [250, 20]$ $\mathbf{a}_3 = [5e–4, 5e–4]$	
Sequential variation	First-order autoregressive	$\boldsymbol{\beta}_t = \mathbf{a}_0 + \mathbf{a}_1 \boldsymbol{\beta}_{t-1} + \mathbf{e}_t$ $\mathbf{a}_0 = [2, 1.8]$ $\mathbf{a}_1 = [0.7, 0.8]$ $\boldsymbol{\beta}_0 = [0, 0]$	
Sequential variation	Random walk	$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \mathbf{e}_t$ $\boldsymbol{\beta}_0 = [0, 0]$	
Random variation	Random coefficient	$\boldsymbol{\beta}_t \sim N(\mathbf{a}_0, \mathbf{a}_1)$ $\mathbf{a}_0 = [9, 7]$ $\mathbf{a}_1 = [3, 3.5]$	

the simulation manipulations including the formulations and resulting parameter paths along with the EKF(C-D) predictions based on the entire set of observations. The errors for all DGPs are assumed to follow the standard normal distribution (i.e., $e_t \sim N(0,1)$). The EKF(C-D) estimation is carried by using the random walk assumption for the parameter paths to examine the consequences of potential misspecification.

In the case of cubic splines as an alternative to CAM in approximating parameter variation, we use the following approach. First, in the interpolation stage, for each parameter β we fit a function $\hat{f}(t)$, which minimizes the following error measure,

$$\sum_{t=1}^{\bar{t}}[\beta(t)-f(t)]^2+\lambda\int_S\left(\frac{\partial^2 f(u)}{du}\right)^2 du, \qquad (14)$$

where n is the estimation window length, S is the defined range for the integration such that $[1,n]\subseteq S\subseteq\mathbb{R}$, and λ is the smoothing parameter that balances smoothness of the approximation and its accuracy (Green and Silverman 1994). We then move to the extrapolation stage and evaluate the smoothing spline, $\hat{f}(t)$, in the forecast window to calculate the parameter forecasts.

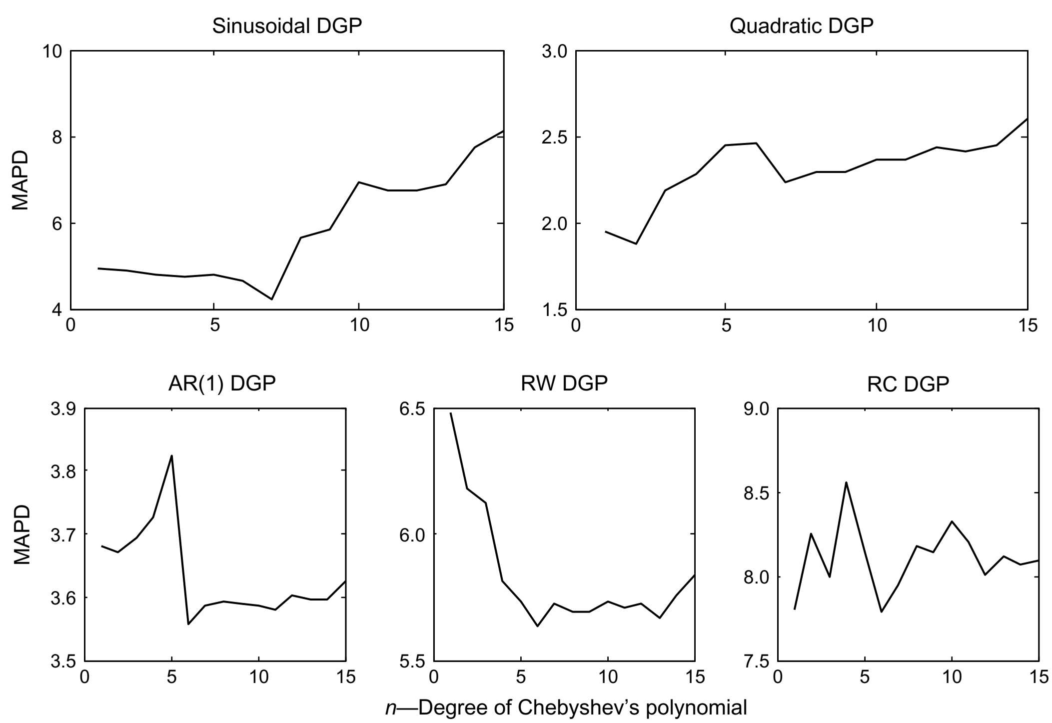
It should be noted that other alternative methods such as vector autoregression (VAR) were also considered. Although VAR appears to be a plausible alternative to CAM, it requires many different steps such as endogeneity (e.g., Granger causality) and evolution (e.g., augmented Dickey–Fuller, Kwiatkowski–Phillips–Schmidt–Shin (KPSS) unit root, and cointegration) tests to determine its specification (VAR, VAR in differences, and vector error-correction (VEC)) before implementation. Hence, contrary to CAM, it critically lacks application ease, which is a desirable property for any forecasting method. Nevertheless, simulation results (not reported here) suggest that VAR performs very well in cases of random variation but is dominated by CAM in cases of systematic variation.

4.2. Choice of Degree of Polynomial Expansion

To provide guidance regarding the choice of CAM's degree of polynomial expansion, we examine its relationship to forecast accuracy via a sensitivity analysis. Specifically, we use all the data available for up to two years before the end of the observation window, vary n over the range [1, 15] and track CAM's forecast accuracy for all DGP scenarios over the entire 24-month window. Figure 7 illustrates CAM's forecast accuracy as it relates to the degree of polynomial expansion for a 96-month estimation window. Although we choose MAPD as the accuracy measure, other metrics such as MAD and MSE follow essentially similar patterns.

A careful examination of Figure 7 suggests that CAM generally achieves optimal forecast accuracy (minimum MAPD) efficiently, in other words, with a relatively small polynomial degree. This is consistent with CAM's parsimony property, which is discussed in the methodology section. However, in the cases of systematic

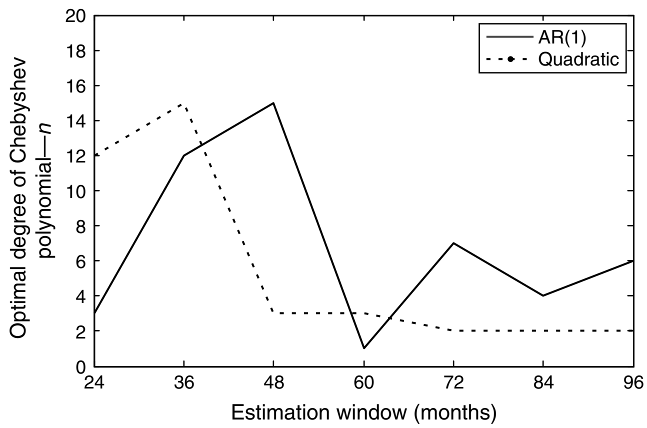
Figure 7 **24-Month Forecast Accuracy vs.** *n*



variation (sinusoidal and quadratic), forecast accuracy is much more sensitive to the degree of polynomial. Specifically, forecast accuracy in such cases deteriorates considerably for polynomial degrees past the optimal, as indicated by the much higher corresponding MAPDs. Hence, when variation is systematic, CAM appears to have a global optimal n that can be conclusively identified through forecast accuracy since nonoptimal polynomial degrees lead to severe overfitting. In other words, CAM is essentially "self-penalized" for the use of a higher-degree polynomial in cases of systematic variation. This reveals a very appealing property of CAM that can be used to discern whether the underlying pattern of variation is systematic or not. It is also worth noting that, for the quadratic scenario, the optimal CAM degree is $n = 2$, which matches the degree of the actual variation pattern. This suggests that CAM can accurately recover the actual path in cases of systematic variation. In cases of more random variation, forecast accuracy is less sensitive to the polynomial degree, and when variation is purely random (RC) it fittingly exhibits a random pattern. Hence, the choice of optimal polynomial degree in such cases is less clear but also less critical, suggesting that the potential for accuracy gains with CAM is higher when variation is systematic. This argument will be further supported in the next section.

The choice of polynomial degree is also influenced by the amount of data available to the forecaster or the length of the estimation window. For example, in shorter estimation windows the optimal polynomial degree may exhibit variation due to the evolving nature of the parameter path and the lack of sufficient information regarding its direction and pattern. However, as more information becomes available and the estimation window increases, the optimal polynomial degree should exhibit convergence, more so in the case of systematic variation due to the higher sensitivity of forecast accuracy to the polynomial degree. In other words, the choice of optimal n should become more conclusive as the estimation window expands, particularly for systematic variation patterns. Indeed, Figure 8 shows how, in the case of quadratic variation, the optimal n converges to 2 as the estimation window expands. In the case of AR(1) the optimal polynomial

Figure 8 *n* **vs. Estimation Window Length for 24-Month Forecasts**



degree does not exhibit a similarly clear pattern because of lower sensitivity of forecast accuracy.

4.3. Forecast Accuracy Analysis

We produce different monthly forecasts using a method similar to the empirical application. Specifically, for each DGP scenario, we forecast sales for up to 24 months out starting with an estimation window of two years. We then progressively expand the estimation window by one month and produce 1- to 24-month forecasts until the end of the eighth year (96 months) so we can project a full 24-month forecast at all times. In other words, we keep expanding the estimation window in monthly increments from month 25 to month 96 or for a series of 72 months. This results in $24 \times 72 = 1{,}728$ different forecasts for each DGP scenario. The average errors across all estimation windows for 12- and 24-month forecasts are summarized in Table 6 for all parameter variation scenarios. Similar to the last section's results, we report MAPDs since they provide a scale-free forecasting accuracy measure. However, much like in the empirical application, the results for the other accuracy measures essentially follow the same pattern. A careful examination of Table 6 leads to the following observations:

(a) CAM outperforms competing methods when parameter variation is more systematic (sinusoidal, quadratic).

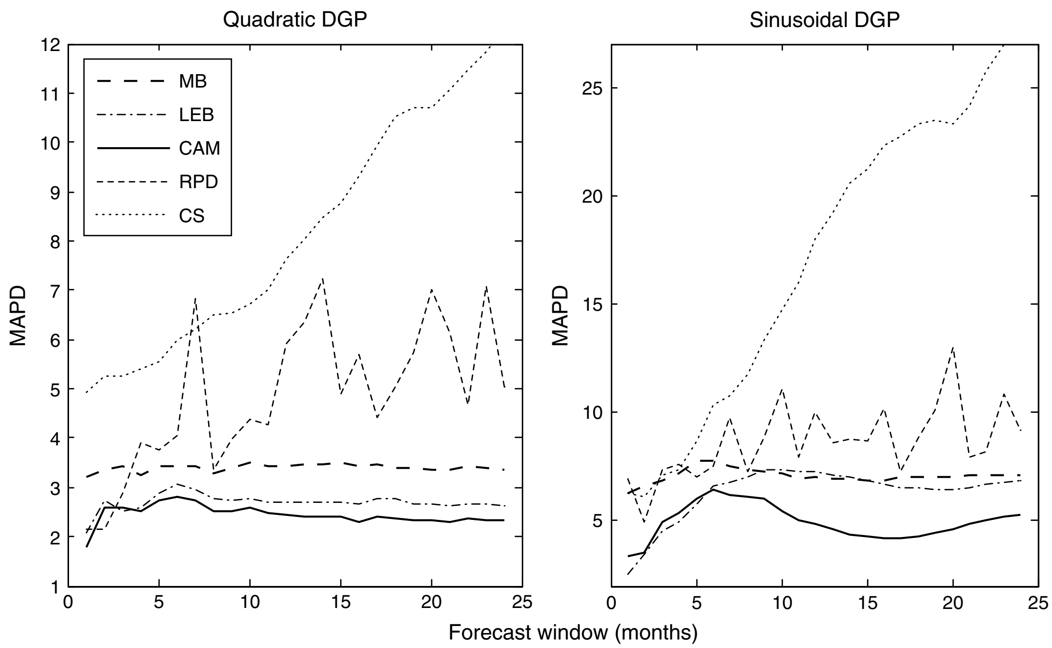
(b) LEB is the best forecasting method when parameter variation follows a random walk.

Table 6 Simulation Results: MAPDs

	12-month						24-month					
	CAM	LEB	MB	RPD	CS	Accuracy[a] improvement (%)	CAM	LEB	MB	RPD	CS	Accuracy improvement (%)
Sinusoidal	**4.88**	7.29	7.06	10.02	20.4	**31**	**5.27**	6.89	7.12	9.16	64.99	**23**
Quadratic	**2.44**	2.72	3.43	5.91	6.54	**10**	**2.32**	2.61	3.35	5.04	43.15	**11**
AR(1)	**3.18**	**3.18**	3.9	6.17	4.75	**0**	**2.91**	2.97	3.48	6.78	9.4	**2**
RW	4.99	**4.73**	6.67	5.22	11.31	**5**	5.96	**5.15**	7.26	5.86	7.45	**12**
RC	6.36	6.2	**6**	6.25	8.56	**3**	6.54	6.7	**6.02**	7.71	29.74	**8**

[a]Percentage forecasting accuracy improvement achieved by the best method over the second-best method.

Figure 9 **Multiperiod Forecast Accuracy**



(c) MB is the best forecasting method in the purely random variation scenario (RC).

(d) AR(1) variation appears to represent the "boundary" case. CAM and LEB tie for best in the 12-month case, but CAM slightly edges LEB in the 24-month forecast.

(e) CAM dominates cubic splines (CS) across all parameter variation scenarios.

The previous points clearly suggest that CAM achieves higher forecast accuracy when parameter variation patterns are more systematic. This is consistent with the previous section's observation on the higher sensitivity of forecast accuracy to CAM's polynomial degree, hence the potential for more accuracy gains, in the cases of systematic variation. In these cases, CAM also maintains its advantage over the entire 24-month forecasting Horizon, suggesting low accuracy variance (Figure 9). This should be attributed to CAM's ability to maintain its parameter forecast accuracy throughout the forecasting horizon. To check this, we calculated the variance of CAM's parameter forecast accuracy and found it to be low in both cases of systematic variation, where CAM outperforms the competing methods, and much lower than in the cases of the random walk and random coefficients, where CAM underperforms (see Figure 10). Thus, the variance of CAM's parameter forecast accuracy provides an indicator of its forecasting performance and can serve as a diagnostic for its use. It should be noted that the variance of the parameter forecast accuracy in the empirical application is also low (see Figure 11 for the Prius case). Furthermore, the sensitivity of CAM to the polynomial degree and the accuracy improvement attained in the empirical application are similar to

Figure 10 **CAM Average Parameter Forecast Accuracy**

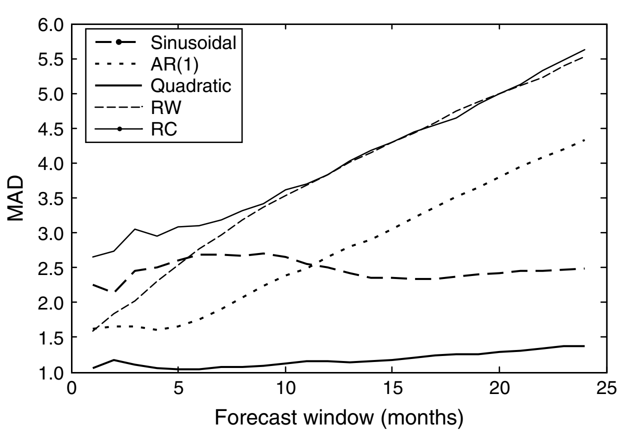
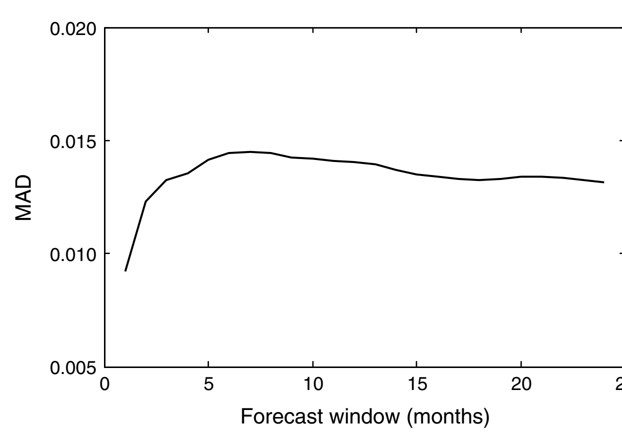


Figure 11 **CAM Average Parameter Forecast Accuracy—Prius**



those reported for the systematic variation simulation scenarios. Thus, conditions in the empirical application are indicative of systematic variation, lending validity to our findings.

AR(1) appears to represent the boundary condition for CAM's forecasting superiority relative to its benchmarks, since for this case CAM is in a virtual tie with LEB before further dropping in terms of accuracy when patterns become more random than AR(1) (RW, RC). Thus, when the pattern transitions toward more randomness, with AR(1) as the boundary case, CAM loses its accuracy. This is also effectively captured in Table 6 where as one moves down the rows to patterns of increasing randomness (i.e., from sinusoidal to RC), the highest accuracy (denoted in bold) shifts from the left-most column (CAM) to LEB (for RW), and eventually to MB (RC). The better performance of LEB in the random walk scenario is in line with the expectation that it should be the best forecasting option in the absence of misspecification. The performance of RPD in this case is also notable and according to expectations since it draws on the previous parameters' (prior) distribution, in other words, the intercept of the Kalman filter update, but it is more susceptible to the error of the draw just as in the empirical application. Finally, in the scenario of entirely random variation (RC), MB does better than competing methods also in line with expectations. The lagging performance of RPD in the simulations, in contrast to the empirical application, is noteworthy but should be attributed to the more controlled nature of the simulations environment. For example, in real-world applications, parameters may exhibit more local deviations from known distribution behavior than in a simulated environment, a condition which should be more conducive to RPD. Hence, forecasters should not discount RPD as a potential method, especially in noisy environments. It is worth noting that across all scenarios the best performing methods maintain and frequently improve their accuracy over competing methods as the forecasting horizon increases. Hence, performance over a longer forecasting horizon seems to provide a reliable criterion for the choice of the most accurate method.

Given that the state parameter variation specification used in the simulations is also random walk, to mirror the empirical setup, the simulation results suggest that CAM handles misspecification better when the true pattern of parameter variation tends to be more systematic. This is a nontrivial advantage since the random walk is a locally constant model and parameter variation in dynamic environments is expected to exhibit patterns with more systematic and frequently drastic changes, as our empirical application suggested. However, forecasters may still opt to employ the random walk specification due to lack of prior knowledge regarding the exact form of parameter variation. Hence the likelihood of misspecification due to the random walk assumption in such environments is high, underlining the importance of CAM which improves multiperiod forecasts by providing a surrogate for the measurement updating process over the forecasting horizon. Our empirical application and simulations suggest that the combination of a random walk model and a CAM forecasting method can provide forecasters with an advantage in the presence of uncertainty regarding the functional form of parameter variation. We further discuss this important implication in the concluding section below.

Finally, the cubic spline (CS) method fails to match CAM's accuracy across all parameter variation scenarios and both forecast window lengths. Particularly in the case of the larger forecast window (24 months) under systematic variation, CS forecast accuracy exhibits significant decline (i.e., MAPDs increase 3- to 8-fold when the forecast window increases from 12 to 24 months). Our results confirm the role of splines as a method mainly for interpolation, smoothing, and local approximation rather than for long-range forecasting, further supporting relevant literature (e.g., Green and Silverman 1994, pp. 11–28; Hardle 1990, pp. 56–63; Ruppert 2011). CAM's superior performance over cubic splines also provides further validity to the choice of approximation method, suggesting that such a choice is critical in forecasting performance. Thus, the simulations suggest that CAM's desirable properties, discussed in the methodology section, appear to have a discernible effect on multiperiod forecasting performance.

5. Extensions

In this section we examine two extensions of our approach: the case where parameter variation is caused by another variable because of the presence of interaction effects, and the case of forecasting time-varying covariates. We investigate the first case via an additional simulation, whereas for the second case we provide an illustration of how it can be accommodated in our state–space formulation.

5.1. Alternative Causal Mechanism of Parameter Variation

In the empirical application as well as in the simulations we assume that parameters vary solely as a function of time. Such an assumption is strongly rooted in a life-cycle-based rationale, according to which response to marketing mix variables evolves over time as a result of underlying factors such as consumer awareness and learning, as well as changing market composition (e.g., more late than early adopters). However, a rich body of literature suggests that response to marketing mix may be influenced by other managerial actions, essentially pointing to an interaction effect (e.g., Desiraju et al.

Table 7(a) **Additional Check**

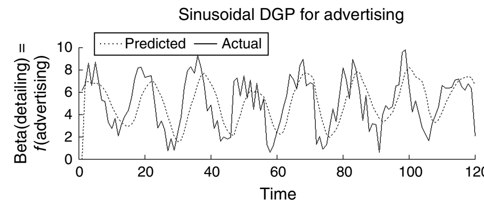
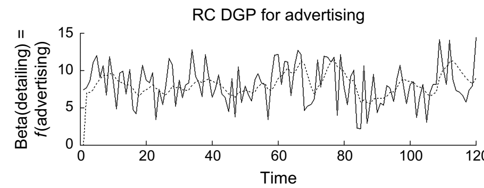
Data generating process (DGP)	Formulation and hyperparameters	Simulated parameter paths
Systematic variation Sinusoidal	$A_t = a_0 + a_1 \sin(a_2(t + a_3)) + e_{1t}$ $\beta_\tau = b_0 + b_1 A_t + e_{2t}$ $(a_0, a_1, a_2, a_3) = (-3, 1.75, 0.4, 0)$ $(b_0, b_1) = (10, 1.7)$	
Random variation Random coefficient	$A_t \sim N(a_0, a_1)$ $\beta_\tau = b_0 + b_1 A_1 + e_t$ $(a_0, a_1) = (2, 1.5)$ $(b_0, b_1) = (3.75, 2)$	

Table 7(b) **Results for Additional Simulations: MAPDs**

	12-month					24-month				
	CAM	LEB	MB	CS	Accuracy[a] improvement (%)	CAM	LEB	MB	CS	Accuracy improvement (%)
Sinusoidal	**2.02**	6.56	4.00	9.05	50	**3.46**	4.22	3.85	18.04	**10**
RC	6.29	6.00	**5.51**	7.46	8	5.38	4.76	**4.59**	9.25	**4**

[a]Percentage forecasting accuracy improvement achieved by the best method over the second-best method.

2004, Gatignon and Hanssens 1987, Naik and Raman 2003, Narayanan et al. 2004, Parsons and Abeele 1981). In fact, one of the frequently studied interactions is that between physician journal advertising and detailing in the marketing of pharmaceutical products, which is the context of one of our empirical application settings. An interaction between physician journal advertising and detailing suggests that the temporal pattern of response to detailing depends on the pattern of physician journal advertising. Hence, when interaction effects are in play, the causal mechanism of parameter variation is the underlying pattern of another variable rather than time itself. To further explore this, we conduct a new set of simulations where parameter variation is a function of another variable. To ensure better control of the simulation conditions, we focus our attention on a model where sales are a function of a single variable (detailing) and the variation of the corresponding parameter also depends on a single variable (physician journal advertising). Table 7(a) summarizes the modeling and fit details for this alternative parameter evolution mechanism.

We focus on two types of variable-induced parameter variation: (a) systematic (sinusoidal) and (b) random (RC). Similarly to previous simulations, we expect that when the variable exhibits systematic variation the proposed method will outperform benchmarks, but not when variation is more random. The results of this new simulation are exhibited in Table 7(b). Consistent with our expectations, CAM outperforms benchmarks in the systematic scenario whereas MB is preferred in the random variation case. Hence, similar to the purely temporal case, the proposed method is favored in cases of systematic variation. Given that the variables affecting parameter variation frequently reflect managerial actions, such as the case of physician journal advertising, the likelihood of a systematic variation scenario is high.

5.2. Forecasting Time-Varying Covariates

Our multiperiod forecasting methodology can be easily extended to contexts where forecasting the evolution of time-varying covariates may be a more important concern than forecasting parameters, especially when they are not entirely controlled by the forecaster such as in the cases of macroeconomic factors and consumer sentiment. It can also be applied in the context of more mature markets than the ones we examined in this study with a potentially more simplified model structure.

We provide an illustration of our state–space formulation where sales are determined by the firm's pricing decision and the unemployment rate that is a macroeconomic indicator not controllable by the firm. We allow price elasticity and the unemployment rate to change over time while we restrict the effect of the unemployment rate on sales to be time invariant. We choose a discrete formulation for the sake of simplicity;

however, a continuous version is easily implementable following the approach in our empirical application.

State equation:

$$\begin{bmatrix} S_t \\ \alpha_t \\ UE_t \end{bmatrix} = \begin{bmatrix} f(\alpha_{t-1}P_{t-1}, \eta\, UE_{t-1}) \\ \alpha_{t-1} \\ UE_{t-1} \end{bmatrix} + \begin{bmatrix} w_{1t} \\ w_{2t} \\ w_{3t} \end{bmatrix}, \quad \mathbf{w}_t \sim N(0, \mathbf{Q}), \tag{15}$$

where the state vector $(S_t, \alpha_t, UE_t)'$ consists of sales, price coefficient, and the net unemployment effect, and P_t denotes price all at time t. The sales model can assume any linear or nonlinear form depending on the context and information available to the modeler. Similarly, the evolution of the price coefficient and the unemployment rate that follow a random walk pattern in Equation (15) can easily take on different forms such as autoregressive.

Measurement equation:

$$\begin{bmatrix} z_t^s \\ z_t^{UE} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} S_t \\ \alpha_t \\ UE_t \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix}, \quad \mathbf{v}_t \sim N(0, \mathbf{R}), \tag{16}$$

where the measurement vector $(z_t^S, z_t^{UE})'$ reflects observed sales and unemployment rate.

The state–space model in (15) and (16) can easily be estimated with the Kalman filter for the observation window. Multistage forecasts of sales can then be calculated by using the proposed method. More specifically, the modeler can use the CAM procedure outlined at the end of the methodology section to estimate and then forecast the price coefficient and unemployment rate paths in the holdout window.

6. Discussion, Summary, and Conclusion

Markets have become increasingly dynamic, necessitating the use of adaptive estimation methods. Although such methods have reduced data requirements and lead to improvements in one-step-ahead forecasts, their use in multiperiod forecasting can be challenging when the exact pattern of state parameter variation is unknown and hence the potential for misspecification is high. Nevertheless, managers still need to rely on multiperiod forecasts to make budgeting, inventory planning, and supply chain decisions. Our study aims at helping forecasters overcome problems arising from the potential misspecification of the state parameter variation by proposing a methodology based on Chebyshev approximation (CAM). The methodology essentially leverages the knowledge acquired from the parameter evolution during the estimation period through a functional approximation, producing a substitute for the measurement updating process in the forecasting period.

Our empirical application in two contexts, sales of a new car brand and a new prescription pharmaceutical, both in newly established product classes, shows that the proposed methodology outperforms methods based on the estimated parameter distributions or other relevant heuristics. We also find that more structured nonlinear new product growth models do better than a simpler DLM version of Koyck's model. Thus, similarly to the seminal work of Engle et al. (1988) we challenge the prevailing view that simple models do better in terms of forecasting (e.g., Meade and Islam 2001).

We further explore the properties of CAM by conducting extensive simulations that examine scenarios representing all three types of parameter variation (Winer 1979): (a) systematic, (b) sequential, and (c) random. The simulation findings suggest that CAM does better when parameter variation is more systematic. Given that, frequently, adaptively estimated models such as the Kalman filter typically deploy the random walk specification because of a lack of prior knowledge on the exact pattern of parameter variation, our simulation findings suggest that CAM can help forecasters in the presence of parameter variation uncertainty. Specifically, forecasters uncertain of the parameter variation pattern may use the following plan to obtain good multiperiod forecasts: First, test the random walk parameter variation assumption by comparing the state (parameter) and observation (sales) equation errors, in a manner similar to the empirical application. If the random walk assumption is not violated, then they can proceed by using the LEB method. If the random walk model is misspecified, then they can apply CAM and check whether (a) forecasting accuracy exhibits low variance over a long horizon and (b) the optimal polynomial degree of CAM can be conclusively identified (i.e., optimal forecasting accuracy is sensitive to the choice of polynomial degree with nonoptimal degrees leading to severe overfitting). If both (a) and (b) are satisfied, they should use CAM forecasts and confirm their choice by testing it against benchmarks over a long forecast horizon.

Future work should consider comparisons of CAM and other approximation methods to combinations of model- and judgment-based forecasts such as the ones proposed by Yelland et al. (2010). Although managerial judgment forecasts were not available for our empirical application, such extensions would increase the adoption of forecasting methodologies in practice, which understandably relies not only on analytics but also on managerial intuition and heuristics.

Acknowledgments

This research benefited substantially from comments and suggestions by participants of presentations at the University of California at Davis, the Marketing Dynamics Conference at New York University, the Marketing Science

Conference in Cologne, Germany, and the EMAC Conference in Copenhagen, Denmark. The authors are also grateful to the department editor, the area editor, and the review team for their constructive comments, suggestions, and support throughout the review process. The authors acknowledge financial support from the Social Sciences and Humanities Research Council of Canada (SSHRC) [Grant 410-2011-0728].

References

Achiezer NI (1956) *Theory of Approximation*. Hyman CJ, translator. (Frederick Ungar Publishing, New York).

Bass FM (1969) New product growth for model consumer durables. *Management Sci. Ser. A-Theory* 15(5):215–227.

Bass FM, Krishnan TV, Jain DC (1994) Why the Bass model fits without decision variables. *Marketing Sci.* 13(3):203–223.

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

Bruce NI (2008) Pooling and dynamic forgetting effects in multitheme advertising: Tracking the advertising sales relationship with particle filters. *Marketing Sci*. 27(4):659–673.

Bucklin RF, Gupta S (1999) Commercial use of UPC scanner data: Industry and academic perspectives. *Marketing Sci.* 18(3):247–273.

Christensen O, Christensen KL (2004), *Approximation Theory: From Taylor Polynomials to Wavelets* (Birkhäuser, Basel).

Davis PJ (1975) *Interpolation and Approximation* (Dover Publications, New York).

Dekimpe MG, Hanssens DM (2010) Time-series models in marketing: Some recent developments. *J. Res. Management* 6(1):24–29.

Desiraju R, Nair H, Chintagunta P (2004) Diffusion of new pharmaceutical drugs in developing and developed nations. *Internat. J. Res. Marketing* 21(4):341–357.

Engle RF, Watson M (1985) The Kalman filter: Applications to forecasting and rational expectations models. Bewley TF, ed. *Advances in Econometrics Fifth World Congress*, Vol. 1 (Cambridge University Press, Cambridge, UK), 245–284.

Engle RF, Brown SJ, Stern G (1988) A comparison of adaptive structural forecasting methods for electricity sales. *J. Forecasting* 7(3):149–172.

EPRI (1983) Weather normalization of electricity sales. EA-3143, Final Report of Project 1922-1: Prepared by Cambridge Systematics, Inc. and QUERI (Electric Power Research Institute, Palo Alto, CA).

Gatignon H, Hanssens DM (1987) Modeling marketing interactions with application to salesforce effectiveness. *J. Marketing Res.* 24(3):247–257.

Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* 85(410): 398–409.

Gelman A, Carlin JB, Stern HS, Rubin DB (2004) *Bayesian Data Analysis* (Chapman and Hall/CRC, Boca Raton, FL).

Gerald CF, Wheatley P (2004) *Applied Numerical Analysis* (Pearson/Addison-Wesley, Boston).

Goldenberg J, Lowengart O, Shapira D (2009) Zooming in: Self-emergence of movements in new product growth. *Marketing Sci.* 28(2):274–292.

Green PJ, Silverman BW (1994) *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach* (Chapman and Hall, London).

Härdle W (1990) *Applied Nonparametric Regression* (Cambridge University Press, Cambridge, UK).

Harvey AC (1991) *Forecasting, Structural Time Series Models and the Kalman Filter* (Cambridge University Press, Cambridge, UK).

Kolsarici C, Vakratsas D (2010) Category- versus brand-level advertising messages in a highly regulated environment. *J. Marketing Res.* 47(6):1078–1089.

Leeflang PSH, Bijmolt THA, van Doorn J, Hanssens DM, van Heerde HJ, Verhoef PC, Wieringa JE (2009) Creating lift versus building the base: Current trends in marketing dynamics. *Internat. J. Res. Marketing* 26(1):13–20.

Lenk PJ, Rao AG (1990) New models from old: Forecasting product adoption by hierarchical Bayes procedures. *Marketing Sci.* 9(1): 42–53.

Meade N (1985) Forecasting using growth curves: an adaptive approach. *J. Oper. Res. Soc.* 36(12):1103–1115.

Meade N, Islam T (2001) Forecasting the diffusion of innovations: Implications for time series extrapolation. Armstrong JS, ed. *Principles of Forecasting: A Handbook for Researchers and Practitioners* (Kluwer Academic, Norwell, MA), 577–595.

Meinhold RJ, Singpurwalla ND (1983) Understanding the Kalman filter. *Amer. Statistician* 37(2):123–127.

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

Naik P, Tsai C-L (2000) Partial least squares estimator for single-index models. *J. Royal Statist. Soc.: Ser. B* 62(4):763–771.

Naik P, Mantrala MK, Sawyer AG (1998) Planning media schedules in the presence of dynamic advertising quality. *Marketing Sci.* 17(3):214–235.

Narayanan S, Desiraju R, Chintagunta PK (2004) Return on investment implications for pharmaceutical promotional expenditures: The role of marketing-mix interactions. *J. Marketing* 68(4): 90–105.

Narayanan S, Manchanda P, Chintagunta PK (2005) Temporal differences in the role of marketing communication in new product categories. *J. Marketing Res.* 42(3):278–290.

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

Osinga EC, Leeflang PSH, Srinivasan S, Wieringa JE (2011) Why do firms invest in consumer advertising with limited sales response? A shareholder perspective. *J. Marketing* 75(1): 109–124.

Palda KS (1964) *The Measurement of Cumulative Advertising Effects* (Prentice Hall, Englewood Cliffs, NJ).

Parker P, Gatignon H (1994) Specifying competitive effects in diffusion models: An empirical analysis. *Internat. J. Res. Marketing* 11(1): 17–39.

Parsons LJ, Abeele PV (1981) Analysis of sales call effectiveness. *J. Marketing Res.* 18(1):107–113.

Putsis WP (1998) Parameter variation and new product diffusion. *J. Forecasting* 17(3–4):231–257.

Riddington GL (1993) Time varying coefficient models and their forecasting performance. *Omega* 21(5):573–583.

Ruppert D (2011) Nonparametric regression and splines. *Statistics and Data Analysis for Financial Engineering* (Springer, New York), 579–596.

Sethuraman R, Tellis GJ, Briesch RA (2011) How well does advertising work? Generalizations from meta-analysis of brand advertising elasticities. *J. Marketing Res.* 48(3):457–471.

Sriram S, Chintagunta PK, Neelamegham R (2006) Effects of brand preference, product attributes, and marketing mix variables in technology product markets. *Marketing Sci.* 25(5):440–456.

Sultan F, Farley JU, Lehmann DR (1990) A meta-analysis of applications of diffusion models. *J. Marketing Res.* 27(1):70–77.

Vakratsas D, Kolsarici C (2008) A dual-market diffusion model for a new prescription pharmaceutical. *Internat. J. Res. Marketing* 25(4):282–293.

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

Winer RS (1979) Analysis of the time-varying effects of advertising: The case of Lydia Pinkham. *J. Bus.* 52(4):563–576.

Xie JH, Song XM, Sirbu M, Wang Q (1997) Kalman filter estimation of new product diffusion models. *J. Marketing Res.* 34(3): 378–393.

Yelland PM, Kim S, Stratulate R (2010) A Bayesian model for sales forecasting at Sun Microsystems. *Interfaces* 40(2):118–129.