KALMAN FILTER ESTIMATION OF NEW PRODUCT DIFFUSION MODELS

Jinhong Xie, X. Michael Song, Marvin Sirbu and Qiong Wang*

July, 1996

(The Third Revision)

* Jinhong Xie is Assistant Professor of Marketing, Marketing Department, College of Business Administration, University of Florida, Gainesville, FL 32611, Phone: (904)392-0161 x1233. X. Michael Song is Associate Professor of Marketing, Department of Marketing and Logistics, N370 North Business Complex, Michigan State University, East Lansing, MI 48824-1122, Phone: (517)353-6381. Marvin Sirbu is Professor of Engineering and Public Policy and Industrial Administration, Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA 15213, Phone: (412)268-3436. Qiong Wang is a doctoral candidate, Department of Engineering and Public Policy, Carnegie Mellon University, Pittsburgh, PA 15213, Phone: (412)268-5617.

Acknowledgment We are grateful to Dan Horsky, K. Sridhar Moorthy, Barton A. Weitz, Steven M. Shugan, Dipak C. Jain, Jehoshua Eliashberg, Kannan Srinivasan, David Braden, Roger Calantone, Ambar Rao and V. Srinivasan for their many helpful suggestions. We are also thankful to JMR editor, Vijay Mahajan and three anonymous reviewers for their insightful and constructive comments.

KALMAN FILTER ESTIMATION OF NEW PRODUCT DIFFUSION MODELS

ABSTRACT

This paper introduces a new estimation procedure, Augmented Kalman Filter with Continuous State and Discrete Observations (AKF(C-D)), for estimation of diffusion models. This method is directly applicable to any differential diffusion model without imposing constraints on the model structure or the nature of the unknown parameters. It provides a systematic way to incorporate prior knowledge about the likely values of unknown parameters, and updates the estimates when new data become available. The empirical comparisons of AKF(C-D) with five other estimation procedures demonstrate AKF(C-D)'s superior prediction performance. As an extension to the basic AKF(C-D) approach, a parallel-filters procedure is also developed for estimation of diffusion models when there is uncertainty about diffusion model structure or prior distributions of the unknown parameters.

INTRODUCTION

The desire to forecast the diffusion of new products has inspired a large body of research during the past two decades. The accurate prediction of new product diffusions is critical in designing marketing strategies for new product planning and management. Before predicting sales, diffusion model specifications must be determined and parameters must be estimated. Over the past two decades, a variety of estimation methods for estimating diffusion models have been proposed (a good review of the literature on these estimation techniques can be found in Mahajan, Muller, and Bass 1990). In their review paper, Mahajan, Muller, and Bass (1990) classify diffusion model estimation procedures into two groups: *time-invariant* estimation procedures and *time-varying* estimation procedures. *Time-invariant* estimation procedures include the conventional estimation methods such as Ordinary Least Square (OLS) (Bass 1969), Maximum Likelihood Estimation (MLE) (Schmittlein and Mahajan 1982), and Nonlinear Least Squares (NLS) (Srinivasan and Mason 1986). These estimation procedures suffer two common limitations. First, in order to obtain stable and robust parameter estimates, *time-invariant* procedures often require data to include the peak sales (Mahajan, Muller, and Bass 1990). *Time-invariant* procedures are not very helpful in forecasting a new product diffusion process because by the time sufficient data have been collected, it is too late to use the estimates for forecasting or for planning marketing strategies.

Second, although diffusion models are often expressed by a continuous differential equation, the *time-invariant* procedures can only be applied to a *discrete* form of a diffusion model or a *solution* to a diffusion model. The discrete form used to estimate diffusion models often results in biased and high variance estimates. Requiring a diffusion model to be analytically solvable limits the applicability of the estimation procedures. For example , the original Bass (1969) diffusion model is expressed by

(1) $$\frac{dn(t)}{dt} = [p + \frac{q}{m} n(t)]\,[m - n(t)]$$

where $n(t)$ is the cumulative number of adopters, p is the coefficient of external influence, q is the coefficient of internal influence, and m is the potential market size. None of the *time-invariant* procedures can directly estimate Equation (1).

In order to use OLS estimation, a discrete analog (Equation 2) must be formulated to approximate the differential Equation (1) (Bass, 1969):

(2) $$x(t) = [p + \frac{q}{m} n(t\text{-}1)]\,[m - n(t\text{-}1)] = \alpha_1 + \alpha_2 n(t - 1) + \alpha_3 n^2(t - 1), \quad t=1,2, \ldots$$

where $x(t)$ is the number of new adopters in the t^{th} interval, and

$$\text{(3)} \qquad \alpha_1 = pm, \qquad \alpha_2 = (q - p), \qquad \alpha_3 = -\frac{q}{m}.$$

The transformation of the variables p, q, and m into α_1, α_2, and α_3 is necessary to produce a linear equation suitable for OLS estimation. After obtaining OLS estimates of α_1, α_2, and α_3, one can derive the parameters in the Bass model (p, q and m) using Equation (3). Schmittlein and Mahajan (1982) demonstrate that this approach introduces a time interval bias: "This substitution causes a problem in that, as defined, [$x(t)$] will underestimate [$\frac{dn(t)}{dt}$] for time intervals before the maximum adoption rate is reached and will overestimate after that point" (Schmittlein and Mahajan 1982, p.60). Moreover, multicollinearity between the explanatory variables of Equation (2) may lead to large sampling variances of the estimated OLS coefficients, to great covariance of the estimated OLS coefficients, and to great sensitivity of the estimated coefficients to small data changes (Johnston 1984; Mahajan, Muller, and Bass 1990).

Schmittlein and Mahajan (1982) show how to use Maximum Likelihood Estimation (MLE) for estimation of the Bass model. MLE avoids using a discrete analog by estimating unknown parameters directly from the *solution* to the Bass model:

$$\text{(4)} \qquad F(t) = \frac{1 - e^{-(p+q)t}}{1 + \frac{q}{p} e^{-(p+q)t}},$$

where $F(t)$ is the cumulative fraction of adopters at time t, and $n(t) = mF(t)$. While MLE eliminates the time-interval bias present in the OLS procedure by using a continuous time model and provides forecasts which are significantly better than OLS (Schmittlein and Mahajan 1982; Mahajan, Mason and Srinivasan 1986), it is limited to diffusion models that are solvable (*i.e.*, the cumulative number of adopters can be expressed as an explicit function of time). Nonlinear Least Squares (NLS) suggested by Srinivasan and Mason (1986) produces more robust forecasts than both MLE and OLS, but NLS suffers from the same limitations as MLE and requires that the diffusion model be solvable.

Requiring a diffusion model to be solvable imposes a significant limitation on the applicability of the estimation procedures. For example, many diffusion models have been developed to study the impacts of marketing mix variables on new product diffusion (*e.g.*, Kalish 1983, 1985; Dehbar and Oren 1985, 1986; Horsky and Simon 1983; Simon and Sebastian 1987, Xie and Sirbu 1995). Incorporating marketing mix variables into diffusion models often increases the complexity of the model structure and hence causes diffusion models to have no analytical solutions. Diffusion models may also be unsolvable if diffusion patterns are not assumed to be symmetric. As pointed out by

Easingwood, Mahajan, and Muller (1983), the Bass diffusion curve is assumed to be symmetric (*i.e.* the diffusion pattern after the point of inflection is a mirror image of the diffusion pattern before the point of inflection), which may not be the case for all diffusion processes. In recent years, a new set of diffusion models, called "flexible diffusion models", have been generated to relax the assumption of a symmetric diffusion pattern. Out of the ten flexible diffusion models reviewed by Mahajan, Muller, and Bass (1990), five do not have an analytical solution.

Time-varying estimation procedures have been introduced to overcome some of these limitations of *time-invariant* procedures (Mahajan, Muller, and Bass 1990)[1]. *Time-varying* estimation procedures start with a prior estimate of unknown parameters in a diffusion model and update the estimates as additional data become available. *Time-varying* estimation procedures in the marketing science literature include the Adaptive Filter (AF) developed by Bretschneider and Mahajan (1980), the meta-analysis conducted by Sultan, Farley, and Lehmann (1990), and the Hierarchical Bayesian introduced by Lenk and Rao (1990). By incorporating prior estimates of unknown parameters and updating initial estimates as new data become available, *time-varying* estimation procedures can often provide better early forecasts. However, these *time-varying* estimation procedures are also subjected to the second major limitation of the *time-invariant* estimation procedures; that is, they can not be directly applied to differential diffusion models. For example, while AF (Bretschneider and Mahajan 1980) can update parameters dynamically based on newly obtained observations and it can be applied to models with time-varying parameters, it uses the same discrete analog as the OLS procedure. Procedures developed by Sultan, Farley, and Lehmann (1990) and Lenk and Rao (1990) are applied to the solution of the Bass model, thus they also require that the diffusion model be solvable.

Table 1 summarizes the six estimation procedures discussed in the review paper by Mahajan, Muller, and Bass (1990). It shows that three procedures are *time-invariant* methods and that all procedures require either an analytically solvable diffusion model or a discrete analog. Given that (1) an important benefit of diffusion models is to provide early forecasting of new product diffusions and (2) diffusion models are often expressed by differential equations that do not have analytical solutions, a method for use with diffusion models should have at least two desirable properties. First, to facilitate forecasts early in the product cycle, when only a few observations are available, the method should provide a systematic way of incorporating prior information about the likely values of model parameters and an updating formula to upgrade the initial estimates as additional data become available. Second, it should be directly applicable to diffusion models expressed as a differential equation for cumulative sales. It should require neither a discrete analog (*i.e.*, not require that a continuous differential equation be rewritten as a discrete time equation in a way which introduces a time interval bias), nor an analytic solution to the equation (*i.e.*, not require that cumulative sales can be written as an analytic function of *t*). However, as Table 1 shows, the existing estimation procedures either do not allow incorporation of prior information or can not be directly applied to differential diffusion models.

The purpose of this research is to introduce a new approach to diffusion model estimation-–an Augmented Kalman Filter with Continuous State and Discrete Observations [hereafter referred as AKF(C-D)]. The procedure removes the deficiencies associated with the current estimation methods and possesses the two desirable properties outlined above. Compared with other estimation procedures, the proposed procedure also has several additional advantages. It can be used for estimating parameters that change over time (deterministic or stochastic); it explicitly incorporates observation error in the estimation process, which is ignored in other procedures; and its algorithm is straightforward and easy to implement. Furthermore, a parallel AKF(C-D) procedure can be used to overcome the uncertainty in choosing diffusion model structure and/or prior distributions of unknown parameters. Using data from several different products, we compare the predictive performance of the proposed procedure with other commonly employed estimation methods. The empirical results presented later in this paper demonstrate that AKF(C-D) has some significant advantages over other techniques.

The paper is organized as follows. In the next section, we present the AKF(C-D) estimation procedure and show how the procedure can be applied to estimate diffusion models. This is followed by an empirical comparison between the AKF(C-D) and five other estimation procedures. We then extend the procedure by introducing parallel filters for estimation of diffusion models when the model structures and the prior estimates of the parameters are uncertain. We conclude by summarizing the advantages of the proposed procedure.

INTRODUCTION OF AKF(C-D)

The Standard Kalman Filter Technique

One of the major contributions to optimal control theory, the standard Kalman filter was first developed to estimate engineering systems in the early 1960s. During the last two decades, the standard Kalman filter has also been adopted to estimate social systems (*e.g.,* Athans 1974; Morrison and Pike 1977; Duncan, Gorr and Szczypula 1993; Slade 1989; Tegene 1990, 1991).

The standard Kalman filter is a state estimation technique (i.e., it is designed to estimate state variables of a dynamic system). It is based on a probabilistic treatment of process and measurement noises. The basic form of the *discrete* Kalman filter consists of two sets of equations: *system equations*, which describe the evolution of the state variable $\mathbf{y}_k$, and *measurement equations*, which describe how the observations are related to the state of the system:

(5)[2] System equations: $\mathbf{y}_{k+1} = \mathbf{f}_k\,[\mathbf{y}_k,\ \beta,\ \mathbf{u}_k,\ t_k] + \mathbf{G}_k\mathbf{w}_k$

where $\mathbf{y}_0 \sim (\tilde{\mathbf{y}}_0,\ \mathbf{P}_0)^3,\ \mathbf{w}_k \sim (0,\ \mathbf{Q})$

(6) Measurement equations: $\mathbf{z}_k = \mathbf{H}_k\,\mathbf{y}_k + \mathbf{v}_k$

where $\mathbf{v}_k \sim (\mathbf{0},\ \mathbf{R})$

where $\mathbf{y}_k$ is the state vector, and $\mathbf{z}_k$ is the observation vector; $\{\mathbf{w}_k\}$ and $\{\mathbf{v}_k\}$ are stationary white noise processes uncorrelated with $\mathbf{y}_k$ and with each other; $\mathbf{f}_k$ is a vector function of state ($\mathbf{y}_k$), parameter vector (β), control vector ($\mathbf{u}_k$), and time t_k. $\mathbf{G}_k$ and $\mathbf{H}_k$ are known matrices, and $\mathbf{Q}$ and $\mathbf{R}$ are covariance matrices of the process and measurement noises respectively.

The purpose of the standard Kalman filter is to use the observed data ($\mathbf{z}_k$) to estimate the state variables which may be measured with noise or may not be measured directly. When the noise statistics are all Gaussian, the standard Kalman filter is known to be an optimal estimator (*i.e.* the estimator that minimizes the mean squared error of the estimate). When the noise statistics are not Gaussian, it is still the best linear estimator (Lewis 1986, p.148). The standard Kalman filter has been proved to be a powerful tool for a variety of applications in both the engineering and management science literatures (Lewis 1986; Morrison and Pike 1977; Kahl and Ledolter 1983; Tegene 1990, 1991; Meade 1985).

Can the standard Kalman filter technique be applied directly to estimate a new product diffusion process? The diffusion process of a new product can be considered as a dynamic system. A diffusion model can be viewed as a system equation where the state variable is the number of adopters. The measurement equations can be simply expressed by the observed number of adopters plus a measurement noise. Unfortunately, several difficulties prevent direct application of the standard Kalman filter to the estimation of diffusion models. First, in a standard Kalman filter model, both the system equation and the measurement equation are the same type, either discrete (Equations 5 and 6) or continuous[4]. Since diffusion models are often expressed by a continuous differential equation, while sales data are obtained at discrete time intervals, neither the discrete nor continuous Kalman filter is directly applicable in the estimation of diffusion models. Second, the standard Kalman filter treats the parameter vector β as given, but in estimation of diffusion models, the parameters (such as p, q, and m in the Bass model) are often unknown.

The discrete standard Kalman filter can be used to estimate the unknown parameters if an autoregression equation can be used to describe the diffusion process: $x_k = \sum_{i=1}^{s} \beta_i x_{k-i} + \varepsilon_k$ where x_k is the number of new adopters in the k^{th} period (Morrison and Pike 1977; Kahl and Ledolter 1983; Tegene 1990, 1991; Meade 1985). However, because it requires that the sales at a given time, x_k, be expressed as a recursive function of the parameters and previous observations, this formulation of a discrete Kalman filter cannot be applied to diffusion models expressed as a differential equation unless: (a) the differential equation is approximated by a difference equation describing x_k, possibly introducing time interval bias; or (b) the differential equation for $x(t)$ has an analytical solution and $x(t)$ can be explicitly written as a function of lagged values of x. Therefore, it is also subjected to the same limitations as other conventional estimation procedures discussed earlier.

The Proposed AKF(C-D)

To overcome the above limitations of the standard Kalman filter and to make better use of the Kalman filter technique, we introduce an AKF(C-D) by combining two ideas recently developed in the engineering literature: (1) the Extended Kalman Filter with continuous state and discrete observations, which uses discrete observations to estimate the state of a continuous system with *known* parameters (Lewis 1986); and (2) the Augmented Filter for parameter estimation, which estimates unknown parameters in a *continuous* Kalman filter model (Stengel 1986). Reviewing the technical formulations and details of the two methods is beyond the scope of this article. For a review of the two methods, readers are referred to Lewis (1986) and Stengel (1986). Included in this subsection are the basic description of the AKF(C-D) model formulation, the estimation algorithm, and a discussion of the advantages of the proposed procedure.

The AKF(C-D) Model Formulation. The AKF(C-D) model formulation for diffusion models is as follows:

(7) $$\frac{dn}{dt} = f_n\,[n(t), \mathbf{u}(t), \beta, t] + w_n$$

(8) $$\frac{d\beta}{dt} = \mathbf{f}_\beta[\beta, n(t), t] + \mathbf{w}_\beta$$

(9) $$z_k = n_k + v_k$$

where n is the cumulative number of adopters; $\mathbf{u}$ is the marketing mix variable vector; β is the unknown parameter vector; w_n and $\mathbf{w}_\beta$ are the process noise; n_k and z_k are the actual and the observed cumulative number of adopters at time t_k; and v_k is the observation noise. It is assumed that $n(0)$~(n_0, σ_{n0}) and $\beta(0)$~$(\beta_0, \mathbf{P}_{\beta 0})$, $\{w_n, \mathbf{w}_\beta\}$ and $\{v_k\}$ are white noises, $\{w_n, \mathbf{w}_\beta\}$~$(0, \mathbf{Q})$, v_k ~$(0, r)$, and $\{w_n, \mathbf{w}_\beta\}$ and $\{v_k\}$ are not correlated to each other.

Equations (7)-(9) are very general formulations of any new product diffusion processes. Equation (7) is the system equation that characterizes the diffusion rate at time t as a function of the number of current adopters (n), the marketing mix variables ($\mathbf{u}$), the diffusion parameters (β), the time (t), and a random noise (w_n). Equation (8) specifies the time-varying behaviors of unknown parameters. If the unknown parameters are constant, then $\frac{d\beta}{dt} = 0$. Otherwise, a deterministic function $\mathbf{f}_\beta$ and a stochastic component $\mathbf{w}_\beta$ are used to describe changes of unknown parameters over time. Equation (9) is the measurement equation that assumes that the number of adopters can be measured directly but may contain measurement errors, v_k. Notice that different errors involved in the estimation process can be formulated as different noises. The process noise, w_n, includes: (1) model specification errors, which could be a result of either excluding some important variables, such as prices or advertising effect, from the diffusion model, or mis-specification of the diffusion function (Srinivasan and Mason 1986); (2) sampling errors, which may occur when using the model to describe the diffusion process of a sampled group instead of the entire population. The random error in the data collected is modeled by v_k.

The AKF (C-D) Estimation Algorithm**.** Without loss of generality, we form an augmented state vector ($\mathbf{y}$) that consists of the original state (n) and the unknown parameter vector (β):

(11) $$\mathbf{y} = [n\ ,\ \beta\ ]^T.$$

Then, Equations (7)-(9) can be rewritten as

(12) $$\frac{d\mathbf{y}}{dt} = \mathbf{f}_\mathbf{y}[\mathbf{y},\ \mathbf{u},\ t] + \mathbf{w}_\mathbf{y} \qquad \text{where } \mathbf{f}_\mathbf{y} = [f_n,\ \mathbf{f}_\beta\ ]^T$$

(13) $$z_k = n_k + v_k$$

We now describe the estimation algorithm based on (12) and (13).

The AKF(C-D) algorithm is essentially a Bayesian updating procedure. Figure 1 provides an overview of the AKF(C-D) algorithm. The figure presents the relationships among the real market, the diffusion model, and the AKF(C-D) estimation process. To estimate a new product diffusion process in a real market, a diffusion model with unknown parameters is identified to describe the new product adoption process. Based on prior experience or knowledge, initial estimates are given for the unknown parameters. AKF(C-D) updates the parameters estimates of the diffusion model as new sales data become available. It estimates parameters and updates the state variables through two processes: a *time updating* process and a *measurement updating* process (see Figure 1). More specifically, the AKF(C-D) algorithm takes the following four steps:

1. At $t = t_0$ ($k = 0$), based on prior information, the best prior estimate of the parameter distributions ($\hat{\mathbf{y}}_0$ and $\hat{\mathbf{P}}_0$) and the noise statistics (r and $\mathbf{Q}$) are developed to initialize the filter.

2. Time update: at a given time, t_k, the diffusion model predicts sales and parameter values for the next time period (t_{k+1}) through the *time updating* process which generates an *a priori* estimate of the state[5] defined by $\hat{\mathbf{y}}^-_{k+1}$

(14) $$\hat{\mathbf{y}}^-_{k+1} = \mathrm{E}\{\mathbf{y}_{k+1}|\mathbf{z}_k\}$$

where $\mathbf{z}_k = \{z_1, z_2, \ldots, z_k\}$ includes all available observations at t_k. The corresponding error covariance matrix of the *a priori* estimate is given by

(15) $$\hat{\mathbf{P}}^-_{k+1} = \mathrm{Cov}\{\hat{\mathbf{y}}^{-\mathrm{T}}_{k+1}\, \hat{\mathbf{y}}^-_{k+1}|\mathbf{z}_k\}$$

Time updating is accomplished by integrating equations (16) and (17) over time interval (t_k,t_{k+1})

(16) $$\frac{d\mathbf{y}}{dt} = \mathrm{f}_{\mathrm{y}}(\mathbf{y}, \mathbf{u}, t)$$

(17) $$\frac{d\mathbf{P}}{dt} = \mathrm{F}(\mathbf{y},t)\mathbf{P}^{\mathrm{T}} + \mathbf{P}\mathbf{F}^{\mathrm{T}}(\mathbf{y},t) + \mathbf{Q}$$

where $$\mathbf{F}(\mathbf{y},t) = \frac{\partial \mathbf{f}_{\mathbf{y}}(\mathbf{y}(t),\mathbf{u}(t),t)}{\partial \mathbf{y}}\Big|_{\mathbf{y}(t)=\mathbf{y}}$$

and Equation (16) is the augmented system Equation (12) without process noise.

3. Measurement update: when a new observation, z_{k+1}, becomes available, the forecasting error (i.e., the difference between the observed sales, z_{k+1}, and the predicted sales, $\hat{n}^-_{k+1}$) is used to modify the estimate through the *measurement updating* process which generates the *a posterior* estimate defined by $\hat{\mathbf{y}}_{k+1}$

(18) $$\hat{\mathbf{y}}_{k+1} = \mathrm{E}\{\mathbf{y}_{k+1}|\mathbf{z}_{k+1}\}$$

where $\mathbf{z}_{k+1} = \{\mathbf{z}_k, \mathbf{z}_{k+1}\}$ and the corresponding error covariance matrix is given by

(19) $$\hat{\mathbf{P}}_{k+1} = \mathrm{Cov}\{\hat{\mathbf{y}}^{\mathrm{T}}_{k+1}, \mathbf{y}_{k+1}|\, \mathbf{z}_{k+1}\}$$

The *measurement updating* process is accomplished by equations (20) and (21)

(20) $$\hat{\mathbf{y}}_{k+1} = \hat{\mathbf{y}}^-_{k+1} + \phi_{k+1}[z_{k+1} - \hat{n}^-_{k+1}]$$

(21) $$\hat{\mathbf{P}}_{k+1} = [\mathrm{I} - \phi_{k+1}\mathbf{h}]\, \hat{\mathbf{P}}^-_{k+1}$$

where I is an identity matrix, $\mathbf{h}$ = [1, 0...0] and

(22) $$\phi_{k+1} = \hat{\mathbf{P}}^-_{k+1}\mathbf{h}^T[\mathbf{h}\hat{\mathbf{P}}^-_{k+1}\mathbf{h}^T + r]^{-1}$$

4. Go back to step 2 and iterate.

Advantages of AKF(C-D) Model and Algorithm. In the following discussion, we demonstrate some major advantages of AKF(C-D) in its model formulation and estimation algorithm.

1. *General applicability*. By specifying f_n in Equation (7) and $\mathbf{f}_\beta$ in Equation (8), the AKF(C-D) can be applied to estimate all differential diffusion models in the marketing literature. We show its usefulness and its easy application using three major types of diffusion models.

(i) *The Bass model* (Bass, 1969). Comparing the original Bass model (Equation 1) with Equation (7), we can see that the Bass model is a special case of Equation (7). The Bass model assumes that diffusion rate $\frac{dn}{dt}$ is determined by the cumulative number of adopters (n), external influence parameter (p), internal influence parameter (q), and market potential (m). The model does not include marketing mix variables ($\mathbf{u}$ = 0) nor does the diffusion rate depend explicitly on t. All parameters are constant. Mathematically, the Bass model can be written in the form of (7)-(9) by specifying:

$$\beta = (p, q, m)^T, \quad f_n = (p + \frac{q}{m} n)(m - n) \text{ and } \frac{d\beta}{dt} = 0.$$

(ii) *A diffusion model incorporating marketing mix variable* (Horsky and Simon 1983). Horsky and Simon (1983) extend the Bass model by incorporating the impact of advertising into the diffusion process. Their model is as follows

$$\frac{dn}{dt} = [\alpha + \omega\, \text{Ln}(a) + \gamma\, n(t)](m - n)$$

where a is advertising expenditures, α, ω γ, and m are parameters. This model can be estimated by AKF(C-D) by specifying Equations (7)-(9) as follows

$$\beta = (\alpha, \omega, \gamma, m)^T, \; u = a, \quad f_n = [\alpha + \omega\, \text{Ln}(a) + \gamma\, n](m - n), \text{ and } \frac{d\beta}{dt} = 0.$$

(iii) *A diffusion model with parameters changing over time* (Easingwood, Mahajan, and Muller, 1983). Easingwood, Mahajan, and Muller (1983) develop a Non-Uniform Influence (NUI) diffusion model, in which the coefficient of internal influence in the Bass model (q) is specified as a function of current penetration level:

$$q(t) = q_0[\frac{n(t)}{m}]^{\alpha-1}$$

where q_0 and α are constant. Allowing q to vary over time complicates the differential equation and the model does not have an analytical solution. As we discussed above, AKF(C-D) does not require a solvable diffusion model and we can apply the procedure directly to the NUI model by simply defining

$$\beta = (p,\ q,\ m,\ \alpha)^T, \quad f_n = (p + \frac{q}{m} n)(m - n), \qquad \frac{d\beta}{dt} = [0, \frac{dq}{dt}, 0, 0]^T$$

where $$\frac{dq}{dt} = q_0 \frac{(\alpha - 1)}{m} [\frac{n(t)}{m}]^{\alpha - 2} \qquad \frac{dn}{dt} = q_0 \frac{(\alpha - 1)}{m} [\frac{n(t)}{m}]^{\alpha - 2} (p + \frac{q}{m} n)(m - n)$$

The above examples demonstrate that AKF(C-D) can be applied directly to a variety of diffusion models. The first example shows how AKF(C-D) is directly applicable to the Bass model, the most commonly used diffusion model in the literature. The last two examples illustrate how AKF(C-D) can be used to estimate diffusion models with marketing mix variables or parameters changing over time as well as diffusion models without analytical solutions.

2. *Capability of estimating time-varying parameter*s. AKF(C-D) is a Bayesian updating process that starts with a prior estimate and updates it as additional data accumulates. We refer the procedure as a "time-varying estimation method" also because it is capable of estimating parameters that change over time. In many cases, it is unrealistic to expect diffusion parameters, such as the coefficient of internal influence, coefficient of external influence, and market potential, to stay constant throughout the entire diffusion process. These parameters change "because of the changing characteristics of the potential adopter population, technological changes, product modifications, pricing changes, general economic conditions, and other exogenous and endogenous factors." (Bretschneider and Mahajan 1980, p130). The parameters' time-varying behaviors can be captured by the AKF(C-D) procedure in two ways. First, since parameters are modeled in Equation (11) as state variables of the augmented state vector, $\mathbf{y}$, parameter values can be updated by the time-updating process if how parameters change over time is known. For deterministic changes, the value of the parameters is updated by integrating $\frac{d\beta}{dt} = f_\beta[\beta, n(t), t]$, which is part of the integration given by Equation (16). For random changes, the variance of the fluctuation is incorporated in the noise matrix, $\mathbf{Q}$, which is then used to update the variance of parameters (see Equation 17). Second, even if there is not enough knowledge to specify parameter changes in the diffusion model, AKF(C-D) is still capable of capturing these changes by adjusting to prediction error. Because the number of new adoptors in each period is a function of the diffusion parameters, changes in the parameters will be reflected in the prediction error. As shown in Equation (20), when a new observation is incorporated, the prediction error ($z_{k+1} - \hat{n}^-_{k+1}$) is used to update parameters. Therefore, information about parameter changes, which is

embedded in the prediction error, will be used to generate new parameter estimates. The use of such an adaptive approach "provides self-adaptive diffusion parameters that can adjust automatically to changing diffusion data patterns and are especially useful when causes of the variations in the diffusion parameters are not known" (Bretschneider and Mahajan 1980, p.131)[6].

3. *Capability of incorporating observation noise into the estimation process.* In comparison with other estimation procedures, one important advantage of AKF(C-D) is that it explicitly acknowledges random errors in the data collected and formulates them as the observation noise. The variance of observation noise (r) is incorporated in the measurement updating process (see Equation 22). If the variance of measurement errors, r, is larger, which means the data collected is less reliable, then ϕ_{k+1} in Equation (22) will decrease. A smaller ϕ_{k+1} implies updating $\hat{\mathbf{y}}_{k+1}$ is less dependent on the prediction error (z_{k+1} - $\hat{n}^{-}_{k+1}$), so the newly obtained observation will have less impact on the parameter updating process.

The above advantages are derived from AKF(C-D)'s model formulation and estimation algorithm, which make AKF(C-D) a better estimator for diffusion model estimation in general. However, the performance of an estimation procedure is determined not only by its formulation and algorithm, but also by the data sources. In the cases where the data collected contain substantial error (sampling error or non-sampling error), these advantages in AKF(C-D)'s model formulation and estimation algorithm will be less effective. In Appendix B, we provide, through mathematical analysis and numerical simulations, a detailed discussion of AKF(C-D)'s advantages and the conditions which strengthen/weaken these advantages.

AN EMPIRICAL EVALUATION OF AKF(C-D) ESTIMATION

In this section, we empirically evaluate the predictive performance of the AKF(C-D) procedure by comparing its forecasting results with five commonly used procedures. Since most studies of the evaluation of estimation procedures use the Bass model as a basis for comparison, the Bass model has the most reported empirical results. Although one of the major advantages of AKF(C-D) is its capability to estimate more complicated diffusion models, to facilitate comparison with other estimation approaches suggested in the literature which were tested for the Bass model, we follow the literature and evaluate AKF(C-D) using the same Bass new product diffusion model.

Data, Evaluation Criteria, and Prior Estimates

Diffusion data for seven products are used including three consumer durables (room-air conditioner, color TV and clothes dryer), two types of medical equipment (ultrasound and mammography), and two educational programs (foreign language and accelerated program). Mahajan, Mason, and Srinivasan (1986) [hereafter MMS] use the same seven data sets to present a comprehensive evaluation of four commonly used diffusion model estimation methods: Ordinary Least Squares Estimation (OLS, Bass 1969), Maximum Likelihood Estimation (MLE, Schmittlein and Mahajan 1982), Non-Linear Least Squares Estimation (NLS, Srinivasan and Mason 1986), and Algebraic Estimation (AE, Mahajan and Sharma 1986). MMS uses three criteria for comparing the one-step-ahead forecasts of the four methods: Mean Absolute Deviation (MAD), Mean Squared Error (MSE), and Mean Absolute Percentage Deviation (MAPD) (see Table 2 for mathematical formulae for the three criteria). After examining the one-step-ahead forecast errors of the four estimation procedures, they conclude that NLS produces the best forecasting results. We apply AKF(C-D) to the same data sets and use the same criteria to compare AKF(C-D)'s performance with the four *time-invariant* methods reviewed in MMS's paper. We also compare AKF(C-D) with a *time-varying* method--the Adaptive Filter developed by Bretschneider and Mahajan (1980). Lenk and Rao (1990) discuss a Hierarchical Bayesian approach whose primary focus is on the development of priors to use in a Bayesian update procedure. The procedure requires making quite different assumptions about the distributions of the parameters as compared to AKF(C-D). Moreover, the actual algorithm used for updating is not specified in their paper. Consequently, we do not compare AKF(C-D) and HB here.

As discussed in the AKF(C-D) algorithm, we need prior estimates of the unknown parameters to initiate the filter. For marketing managers who are in charge of forecasting sales for a given new product, initial estimates can be constructed based on information from various sources such as marketing research results or experience with comparable products. Given that we don't have this product-specific knowledge, in this paper, we construct the prior estimates based on research results reported in the literature and on common knowledge. According to results of a meta-analysis of 213 products conducted by Sultan, Farley, and Lehmann (1990), for most Bass-type diffusion processes, the coefficient of external influence (p) is on the order of 10^{-2}, and the coefficient of internal influence (q) is on the order of 10^{-1}. Thus, we set the mean value of the prior distribution for p to be 10^{-2} and that for q to be 10^{-1} for *all seven products*. The meta-analysis also shows that the variances of p and q are on the same order as the parameter values. Thus, the variances of the prior distributions are simply set to equal the means[7]. The mean value of the prior distribution for the potential market (*m*) is set to be a percentage of the total population (or of the sample population). To be consistent with priors of *p*, *q*, we also set the variance of *m* to be its mean. All values used for the prior estimates are given in the Appendix A. If AKF(C-D) produces superior estimates with this generic approach to estimating priors, it can only perform even better with priors provided by managers with product specific knowledge.

Empirical Results

Table 2 reports a detailed comparison of one-step ahead forecasting performance between AKF(C-D) and the five methods using the three criteria. The empirical results of the four time-invariant procedures (OLS, MLE, AE, and NLS) are taken from Table 8-6 in MMS's paper. Subject to the limitation of *time-invariant* methods, no results are available before the peak for the first four methods (To obtain stable and reliable estimates, MMS uses data up to and including the peak period to estimate the diffusion model and then presents one-step-ahead forecasts in each period after the peak). Like AKF(C-D), AF starts with a priori values of unknown parameters and upgrades the initial estimates as additional data become available. To maintain the objectivity of comparisons, we initiate AF estimation with the same prior as that we used for AKF(C-D) estimation.

To highlight the comparisons, we also present Table 3. In Table 3, a "+" means that AKF(C-D) provides a better prediction than the comparison procedure does using the corresponding criterion, while a "-" indicates that the AKF(C-D) prediction is worse. The results in Table 3 suggest that, in general, AKF(C-D) provides better one-step-ahead forecasting than the other five methods. Out of the total 126 comparisons, AKF(C-D) is better in 109 cases. Compared with each method separately: AKF(C-D) is better than OLS in all 21 cases; better than MLE in 20 out of 21 cases; better than AE in 18 out of 21 cases; better than AF in 35 out of 42 cases. In comparison with NLS, which is considered the best forecasting method by MMS, AKF(C-D) is better in 15 out of 21 cases.

The result that AKF(C-D) outperforms other estimation methods can be explained by its advantages discussed earlier. First, unlike OLS and AF, which rely on discrete analogs in the estimation procedure, AKF(C-D) is applied directly to the Bass model, and thus avoids time-interval bias. Second, the diffusion parameters in the Bass model may vary over time. The rationale for believing the parameters in the Bass model vary over time and the empirical evidence of time-varying behavior of these parameters have been documented in the marketing literature (Bretschneider and Mahajan 1980). Unlike methods that presume constant parameters, (OLS, MLE, AE, and NLS), AKF(C-D) is an adaptive filter which is capable of adjusting automatically to changing diffusion data patterns even without *a priori* knowledge of how the parameters change over time. Sometimes, a *time-invariant* method such as NLS can be applied recursively to make one-step-ahead forecasts, and in that process, prediction error is used implicitly to update parameter estimates. Even in these cases, AKF(C-D) is still more efficient than other methods in using the information in the prediction error to modify parameter estimates (see Appendix B for details). Third, AKF(C-D) explicitly models observation errors as a measurement noise (v_k) and the error variance (r) is used as inputs to the *measurement updating* process (see Equations 20-22). As shown by simulations in Appendix B, the approach of explicitly considering observation errors can significantly improve AKF(C-D)'s forecasting performance over that of AF. Given the advantage of AKF(C-D) in the formulation of the estimation model and the estimation algorithm, it is not surprising that AKF(C-D) gives better overall forecasting results than other methods, as shown in Table 3.

However, as we can also see from Table 3, the performance of AKF(C-D) varies by conditions. While AKF(C-D) outperforms all of the other methods for all the three consumer durable products (total 54 comparisons), its forecasting performance with the four non-durable products is less outstanding. Although the overall performance of AKF(C-D) in forecasting these four products is still better than the competing methods in most cases (55 out of 72 comparisons), its performance is less impressive. While AKF(C-D) is still better than OLS in all 12 cases and is better than MLE in 11 out of 12 cases, it is better than AE in 9 out of 12 cases and better than AF in 17 out of 24 cases. Particularly, we found that AKF(C-D) and NLS have comparable performances (AKF(C-D) is better than NLS in only 6 out of 12 comparisons).

To understand this result, one must not only consider the difference between estimation methods in model formulation and estimation algorithm, but also the data source. Note that the data for durable goods are collected from all 50 million American households and the data for medical equipment and educational programs are collected from survey studies of 209 hospitals and 107 schools respectively. The major difference between the two types of data is that the former contains almost no sampling error but the latter is more subject to sampling error. Our empirical results show that, when sampling error in the data used for estimation is large, the advantages provided by AKF(C-D)'s model formulation and estimation algorithm are diminished. The simulation results presented in Appendix B also confirm this conclusion.

PARALLEL AKF(C-D) -- AN EXTENSION

In previous sections, we discussed the AKF(C-D) estimation procedure that estimates a new product diffusion process based on a given diffusion model and a set of given prior distributions of unknown parameters. In this section we extend the AKF(C-D) estimation to the situation where there is uncertainty in choosing model structure or prior distributions.

Various diffusion models have been developed in the past two decades. Models often differ from each other in terms of the model structure (*e.g.* how should price be incorporated into the diffusion model; should it influence market potential, hazard rate, or both) and assumptions about their parameters (*e.g.*, are parameters constant or varying over time). For a given product, a manager or researcher may have uncertainties about choosing a model for describing the underlying diffusion process from competing models in the marketing literature. Furthermore, s/he may also have uncertainty in constructing prior estimates of parameters because information from different sources may suggest different initial estimates. For example, when developing a prior estimate for the market potential in the Bass model, prior estimates suggested from a survey may be different from test-market results. In this section, we show how a parallel AKF(C-D) procedure can be used to construct forecasts when there are multiple alternatives.

The process of parallel AKF(C-D) estimation is shown in Figure 2. Suppose that L alternative models are considered. The models differ from each other either in system equations (i.e. the diffusion model structures), or in prior estimates of unknown parameters. L filters that correspond to the L alternatives are used in parallel. At the beginning of the estimation, each filter is assigned an initial weight $\omega_i(0)$ based on the researcher/manager's preference (where $\omega_i(0)>0$, $i = 1, 2, \ldots L$ and $\sum_{i=1}^{L} \omega_i(0) = 1$). If no preference is given for any particular filter, then the initial weight will be the same for all filters (i.e. $\omega_i(0)= \frac{1}{L}$). Following the AKF(C-D) estimation algorithm discussed in Section 2, each filter does *time-updating* and *measurement updating* independently. The combined forecast is the weighted sum of L forecast results from the L filters. The weight assigned to each filter is adjusted dynamically according to the filter's forecasting performance. If filter i provides relatively better forecasting, its weight will increase, and thus in the next period, its forecast result will have a stronger impact on the combined forecast. But, if filter *i's* forecasting error is relatively large compared with other filters, its weight will be reduced. When the weight of a filter is reduced to 0, this filter is eliminated from the estimation process (see the Appendix C for a more detailed discussion of this process).

As an example, we apply the parallel AKF(C-D) filters procedure to estimate the diffusion process of room air conditioners. Assume that the Bass model is considered an appropriate model but there are two alternative prior estimates of the market potential m: 10% or 50% of the total population. We use two parallel AKF(C-D)s that differ only in the prior estimate of m. At the beginning of the estimation, equal weights are given to both filters, $\omega_1(0)= \omega_2(0)= 0.5$. The results are presented in Figure 3 (Figure 3-a shows four curves: observed sales and three predicted sales; Figure 3-b shows how the weights of the two filters change over time). From Figure 3-a, we find that the prediction results of filter 2 are consistently better than those of filter 1. From Figure 3-b, we can see that the filter 2's weight increases and that filter 1's weight decreases accordingly. Eventually, filter 1's weight reduces to 0 and it is eliminated from the estimation process. The results suggest that filter 2 is a better model. This example demonstrates that, if there is uncertainty about the prior distributions, a marketing manager can start with several possible prior distributions and the parallel AKF(C-D) will eventually select a "best" model.

CONCLUSION

In this paper, we introduce a new diffusion model estimation procedure (AKF(C-D)) and provide conditions under which the new procedure has a superior predictive performance. Our empirical results suggest that, in many cases, the AKF(C-D) provides better predictive performance than the five commonly used methods. In summary, AKF(C-D) has the following advantages.

First, AKF(C-D) is a very general estimation approach that does not require any constraints on the model structure or on the nature of the unknown parameters. It can be directly applied to any differential diffusion model without requiring the diffusion model to be replaced by a discrete analog or requiring that the diffusion model have an analytical solution. It can be used to estimate both constant parameters and parameters changing over time (both deterministic or stochastic changes). Although many diffusion models have been developed in marketing, only limited empirical results have been reported in the literature. Among others, the specific requirements imposed on model structure by existing estimation approaches make it difficult to empirically test many of the diffusion models in the marketing literature. The general applicability of AKF(C-D) makes it possible to empirically test any differential diffusion models.

Second, AKF(C-D) is a Bayesian estimation procedure. By incorporating prior information in the estimation process and updating the estimate adaptively, AKF(C-D) can provide better forecasts from the very early stages of the diffusion process. As a Bayesian approach, AKF(C-D) uses any available information about prior distributions of the parameters and incorporates them explicitly into the initial distributions of the unknown parameters. Any qualitative procedures (e.g., focus groups) or quantitative procedures (e.g., Hierarchical Bayes and Meta-analysis) that produce more refined prior values for the parameters can be used in conjunction with AKF(C-D) to improve the final performance of the forecasts.

Third, the empirical results show that AKF(C-D) is capable of providing overall superior prediction. Compared with other procedures, three advantages in its model formulation and estimation algorithm make AKF(C-D) a better estimator: (1) Although the data are typically collected at discrete time intervals, AKF(C-D) assumes continuous state evolution and updates the state variables accordingly; it thus avoids the time interval bias problem incurred when continuous models are converted to their discrete equivalents. (2) AKF(C-D)'s model formulation make it capable of estimating parameters with time-varying behavior, with or without *a priori* knowledge of how the parameters change over time. And (3) the AKF(C-D) method accounts explicitly for possible noise during the data collection process. Despite its advantages in model specification and estimation algorithm, AKF(C-D)'s forecasting superiority can be compromised if the data used for estimation contains significant sampling error.

Fourth, the algorithm is very straightforward and easy to implement. When AKF(C-D) is used to estimate parameters changing over time, one can simply modify equation (8), without changing the fundamental algorithms. In the case of nonstationary noise processes, all that is necessary is to replace the covariance matrices r and $\mathbf{Q}$ with r_t and $\mathbf{Q}_t$.

Fifth, when multiple model structures or prior distributions are considered, the parallel AKF(C-D) procedure can be used to deal with the uncertainty. Starting with multiple filters reflecting different diffusion model specifications or initialized with different prior distributions, it is possible to converge rapidly on a "best" model.

Appendix A: Prior Estimates

To initiate AKF(C-D), one needs to provide prior distributions of unknown parameters. These prior estimates can be obtained by conducting a marketing survey or by using previous experience with similar products (see Sultan, Farley, and Lehmann (1990) and Lenk and Rao (1991) for discussion of more sophisticated ways of generating prior estimates). In Table A-1, we present prior estimates of unknown parameters used for AKF(C-D) estimation in this paper.

1. Prior distribution. We make use of results of the meta analysis conducted by Sultan, Farley, and Lehmann (1990) which suggests that for most Bass-type diffusion processes, the coefficient of external influence (p) is on the order of 10^{-2}, and the coefficient of internal influence (q) is on the order of 10^{-1}. We let $E(p_0)=0.01$, and $E(q_0)=0.1$ as prior estimates of the means of p and q for all seven products. The mean value of the prior distribution for the potential market (m) is set to be a percentage of the total population. The prior mean of m, the number of potential adopters, is given as a percentage of total American households in 1960. Given that a higher percentage of the population will adopt color TVs than room air conditioners or clothes dryers, we set the mean value of m to be 40% of the total households for room air conditioners and clothes dryers, but 80% for color TV. For both medical equipment and educational programs, the prior mean of m is set to be 2/3 of the number of hospitals/schools being surveyed. Without product-specific knowledge, we simply set $var(p_0)=E(p_0)$, $var(q_0)=E(q_0)$, and $var(m_0)=E(m_0)$.

2. Noise statistics. We also need to determine the variances of process noise and observation noise. Given that the number of adopters of durable goods numbers in the millions (around 50 million total households in the 50s), while the number of adopters for medical equipment and educational programs numbers in the hundreds (209 hospitals and 107 schools), we set the variance of process noise for all three durable goods to be 10^5, and the process noise for all four non-durable goods to 5. As for observation noise, as discussed in the paper, it can be ignored for the medical equipment and educational programs, but has to be considered for the durable goods. Accordingly, we set the standard error of the observation noise for all three durable goods to be 10% of the observed number of adopters and we assume there is no observation noise in the four survey data sets.

3. To examine the robustness of the estimation, we conducted sensitivity analysis of prior estimates for three durable goods. For the prior estimates shown in table A-1, we increase the values of the initial estimate by as much as 100% and decrease them by as much as 75%. The results are consistent with what we report in this paper.

Appendix B: A Discussion of Advantages of AKF(C-D)

The empirical study in this paper demonstrates that AKF(C-D) provides better one-step-ahead forecasting results than other procedures. We note in the paper that AKF(C-D) achieves that superiority because of three advantages: 1) the procedure is applied directly to diffusion models instead of using a discrete analog; 2) the procedure is more capable of following parameters that change over time; and 3) the procedure explicitly considers observation error. We also discussed how the sampling error will affect the superiority of AKF(C-D). Clearly, it is always better to apply an estimation procedure to a diffusion model instead of its discrete analog. However, it may be less intuitive why AKF(C-D) is more capable of following parameters that change over time; how explicitly considering observation error will improve AKF(C-D)'s forecasting results; and why sampling error will reduce the superiority of the AKF(C-D) procedure. The following formal analysis and numerical simulations further illuminate these issues.

The capability of following parameters that change over time

As has been previously documented in the literature (Bretschneider and Mahajan 1980), it is unrealistic to assume that in the Bass model, p, q, and m stay constant for all time. These parameters are influenced by different time-varying factors, and are likely to change over time. However, in many cases, how parameters vary over time may not be easy to specify in the diffusion model. Under such circumstances, the change of parameters can only be captured indirectly: since the number of new adopters is determined by the diffusion parameters, the prediction error (defined as observation - prediction) contains some information on parameter changes. As a result, the change of parameters can be followed in an estimation process if the prediction error is used as feedback in updating parameter estimates. By definition, all adaptive filters, including AF and AKF(C-D), use the prediction error as feedback to modify parameter estimates. The prediction error can also be used implicitly by some *time-invariant* estimation methods in making one-step-ahead forecasts when those methods are being applied recursively to estimate parameters with each newly available observation. In the following, the advantage of AKF (C-D) over *time-invariant* methods is examined first by comparing parameter updating formulas between AKF(C-D) and NLS and then by consideration of numerical simulation results. To reduce the complexity of the problem, we assume p, q in the Bass model are known constants (p=0.01, q=0.1), and focus our analysis only on one unknown parameter, the market potential.

AKF(C-D) estimation. In the AKF(C-D) procedure, when a new observation becomes available, the parameter m is updated through the measurement updating process. Applying Equations (20-22), AKF(C-D) updates m by the following equation:

(B-1) $$\hat{m}_k = \hat{m}_{k-1} + \frac{p_{12}}{p_{11} + r}(z_k - \hat{n}_k^-)$$ (note: since $\frac{dm}{dt} = 0$, $\hat{m}_k^- = \hat{m}_{k-1}$ for all k)

where $\hat{m}_{k-1}$ is the optimal estimate of m given observations up to time t_{k-1}, $\hat{n}_k^-$ and z_k are the predicted and the observed accumulated number of adopters at time t_k. p_{11} is the variance of $\hat{n}_k^-$, r is the observation noise, and p_{12} is the covariance between the prediction of n and the estimate of m.

NLS estimation. Following Srinivasan and Mason (Srinivasan and Mason, 1986), the formulation of the Bass model in NLS estimation is:

(B-2) $$x_k = m\Delta F_k$$

where x_k is the number of new adopters in period [k-1,k), and

(B-3) $$\Delta F_k = \frac{1 - e^{-(p+q)t_k}}{1 + \frac{q}{p}e^{-(p+q)t_k}} - \frac{1 - e^{-(p+q)t_{k-1}}}{1 + \frac{q}{p}e^{-(p+q)t_{k-1}}}$$

Since p and q are known, ΔF_k is an exogenous variable that changes with k. Given $\Delta z_1, \Delta z_2, \ldots, \Delta z_k$ as observations of $x_1, x_2, \ldots, x_k$ (z_k is the accumulated number of adopters by time k), the optimal estimator for m using the NLS procedure is:

(B-4) $$\hat{m}_k = \frac{\sum_{i=1}^{k} \Delta z_i \Delta F_i}{\sum_{i=1}^{k} \Delta F_i^2}$$

To facilitate the comparison, we rewrite (B-4) in a recursive form:

(B-5) $$\hat{m}_k = \hat{m}_{k-1} + \frac{\Delta F_k}{\sum_{i=1}^{k} \Delta F_i^2}(\Delta z_k - \hat{x}_k)$$

where $\hat{x}_k$ is the one-step-ahead forecast of x_k.

(B-6) $$\hat{x}_k = \hat{m}_{k-1}\Delta F_k$$

In both Equations (B-1) and (B-5), the estimate of m at time t_k is expressed as the sum of the estimate of m at time t_{k-1} and the weighted prediction error at t_k:

(B-7) $$\hat{m}_k = \hat{m}_{k-1} + g\,\delta$$

where g is the weight assigned to the feedback and δ is the prediction error (measured in incremental or accumulated number of adopters). A major difference between these two formulas is the weight assigned to the feedback. In AKF(C-D), the weight is a function of different error variances and covariance:

(B-8) $$g_{\text{AKF}} = \frac{p_{12}}{p_{11} + r}$$

(B-8) implies that the weight will be larger when there is a strong correlation between forecasting error and parameter estimation error (i.e. p_{12} is larger), and be smaller if the variance in forecasting errors is larger (i.e. p_{11} is larger), or the observation is less reliable (r is larger). In NLS estimation, the weight parameter is a decreasing function of the number of observations, k, regardless of estimation errors:

(B-9) $$g_{\text{NLS}} = \frac{\Delta F_k}{\sum_{i=1}^{k} \Delta F_i^2}$$

(B-9) implies that when k becomes large enough, the prediction error, which contains information resulting from parameter changes, has little influence in updating parameters. Consequently, the method fails to closely follow the time-varying behavior of diffusion parameters.

The conclusion of the above mathematical analysis is confirmed by our numerical simulations. To demonstrate the advantage of AKF(C-D) in estimating parameters that change over time, we generate a series of data based on the Bass model with time-varying market potential:

(B-10) $$\hat{m}_k = \hat{m}_{k-1} + 0.4 + r_k, \qquad k=1,2,\ldots \qquad m_0=100$$

where 0.4 is the deterministic increase of m from time t_{k-1} to t_k. r_k is the normally-distributed random variable with mean 0 and standard error of 5 percent of m_k. We apply both AKF(C-D) and NLS to estimate simulated diffusion data series assuming there is no prior knowledge on how m changes over time (set $\mathbf{f}_\beta = 0$ in AKF(C-D)). Figure B1-a compares the results from estimating m using both NLS and AKF(C-D). As k increases, the estimated value of m by NLS is significantly smaller than the true value of m, while the estimated value of m by AKF(C-D) can closely follow the change in its true value. Since AKF(C-D) follows the change of parameter better than NLS, it should not be surprising that AKF(C-D) generates a better one-step-ahead forecast result than NLS, as shown in Figure B1-b.

The advantage of explicitly considering observation error

An important feature of the AKF(C-D) algorithm is its explicit consideration of the observation error in the estimation process. This feature provides AKF(C-D) some advantages in diffusion model estimation for two reasons.

First, it allows a better use of market data based on its reliability. As shown in Equation (B-8), the weight of the feedback for AKF(C-D) procedure is a decreasing function of the variance of the observation noise, r. As a result, if the observation contains a larger error (r is larger), then the parameter estimates depends less on observations, and if observation contains a smaller error (r is smaller), then the observation becomes more important in parameter updating.

Second, explicitly considering observation error in the estimation process improves the estimation of n_k by reducing its error variance, which can be proved as follows.

For a procedure that does not consider observation error (e.g., AF), at t_k, its best estimate of n_k is simply the observed number of adopters:

(B-11) $$\hat{n}_k = z_k$$

The variance of the estimation error can be calculated as

(B-12) $$E[(\hat{n}_k - n_k)^2] = E[(z_k - n_k)^2] = E[(n_k + v_k - n_k)^2] = E(v_k^2) = r$$

In AKF(C-D) estimation, however, the observation error is explicitly considered and the estimate is taken as the weighted sum of the observation and the previous prediction of n_k. Based on Equation (20) we have

(B-13) $$\hat{n}_k = \frac{r}{p_{11} + r}\hat{n}_k^- + \frac{p_{11}}{p_{11} + r} z_k$$

where $\hat{n}_k^-$ is the prediction of n_k made at time t_{k-1}. The variance of estimation error by AKF(C-D) can be calculated as:

(B-14) $$E[(\hat{n}_k - n_k)^2] = E[(\frac{r}{p_{11} + r}\hat{n}_k^- + \frac{p_{11}}{p_{11} + r} z_k - n_k)^2] = E[(\frac{r}{p_{11} + r}\hat{n}_k^- + \frac{p_{11}}{p_{11} + r} v_k - \frac{r}{p_{11} + r} n_k)^2]$$

$$= E\{[\frac{r}{p_{11} + r}(\hat{n}_k^- - n_k) + \frac{p_{11}}{p_{11} + r} v_k]^2\} = \frac{p_{11}}{p_{11} + r} r$$

Comparing (B-12) and (B-14), we notice that

(B-15) $$\frac{p_{11}}{p_{11} + r} \le r \qquad \text{for all } p_{11} \ge 0 \text{ and } r \ge 0$$

(B-15) indicates that by explicitly modeling the observation noise in the estimation process, the AKF(C-D) procedure leads to a smaller error variance than procedures that do not consider observation noise.

The advantage of explicitly considering observation noise can also be demonstrated by numerical simulations. Based on the Bass model, we generate two series of data as observations of a diffusion process: one with no observation noise and the other one contains a normally distributed random noise with mean 0 and a standard error that equals 10% of true incremental sales. Starting from the same prior, we apply both AF and AKF(C-D) to the simulated diffusion data. Figures B2-a and B2-b presents the estimated market potential by AF and AKF(C-D) using these two series of data.

Comparing the two figures we see that in the absence of observation noise (Figure B2-a), starting from an underestimated prior (initial value of m is set at 80 while the true value of m is 100), both methods are capable of converging to the true value of the parameter. However, when the data is contaminated by a noise (Figure (B2-b), the parameter estimate of AKF(C-D) keeps intact while that of AF has been carried away by the observation noise.

Another finding of our simulation study is that, whether explicitly considering observation noise will give AKF(C-D) a significant advantage depends on what estimation procedure it is compared with. To further investigate the effect of AKF(C-D)'s advantage of explicitly considering observation error, we have also conducted a similar simulation to compare AKF(C-D) and NLS. We did not find a significant influence of observation noise on the superiority of AKF(C-D). This can be explained by the fact that NLS generates parameter estimates that minimize the sum of squared prediction error for all available data. As a result, the current prediction error does not have much influence on parameter updating, so the observation error does not affect the parameter estimate much. Therefore, the advantage of explicitly considering observation error also becomes less significant.

Finally, to incorporate observation error in the estimation procedure, one needs to know the value of the variance of observation noise, which may be difficult to obtain. In our empirical study of three durable products, we assume the standard error of observation at each time is 10% of incremental sales at that time, which seems to work quite well. However, this is just a rough simplification, a more through approach to adaptively calculating variance of the observation noise has been suggested by Stengel (see Stengel, 1986). Because of the complexity of that approach, we do not included it in this paper.

The influence of sampling error

We concluded from our empirical study that the prediction performance of an estimation procedure not only depends on its model formulation and algorithm, but also on the data source. More specifically, we argued that AKF(C-D)'s superiority becomes less significant when the estimate is based on sample data. The diffusion pattern of a sample may differ from the diffusion pattern of the entire population described by the underlying diffusion model. Thus, using sample data will increase process noise. The sampling error weakens AKF(C-D)'s superiority in forecasting achieved through its advantages in model formulation and estimation algorithm. When sampling error is relatively large compared to other error sources such as time-interval bias, parameter time-varying behavior, or observation noise, AKF(C-D) does not exhibit significant superiority in one-step-ahead prediction over other methods.

By modeling sampling error as process noise, Figure B3 shows how sampling error affects the relative forecasting performance of AKF(C-D) and NLS (same parameters values are used as in Figures B-1 and B-2). We define the relative performance of AKF(C-D) with regard to NLS as τ

(B-16) $$\tau = \frac{\text{MAPD(NLS)}}{\text{MAPD(AKF)}}$$

where MAPD is mean absolute percentage deviation. $\tau > 1$ indicates AKF(C-D)'s performance is superior to NLS, and the larger the τ the more advantage AKF(C-D) has over NLS. Figure B3 confirms our argument that although AKF(C-D) is able to provide better forecasts than NLS ($\tau > 1$), its superiority in prediction power decreases when sampling error becomes large.

Appendix C: Parallel AKF(C-D) Procedure

Assume L AKF(C-D)s are used in parallel to accommodate L choices over model structure and/or prior estimates. At the beginning of the estimation, a weight, $\omega_i(0)$, is assigned to filter i (i=1,2, … L), where $\omega_i(0) \geq 0$ and $\omega_1(0) + \omega_2(0) + .. + \omega_L(0) = 1$. Suppose that, at time t_{k-1}, the time updating result of filter i is $\hat{n}_i(k)$ (i=1,2, … L), then a combined forecast is constructed as the weighted sum of the L forecasting results:

(C-1) $$\hat{n}(k) = \omega_1(k)\,\hat{n}_1(k) + \omega_2(k)\,\hat{n}_2(k) + .. + \omega_L(k)\hat{n}_L(k)$$

When the observed sales, z_k, becomes available, measurement updating (as we described in the paper) will be conducted for each individual filter, and at the same time, the weight assigned to each filter is adjusted as follows:

(C-2) $$\omega_i(k+1) = \frac{\Pr_i(k)\omega_i(k)}{\sum_{l=1}^{L} \Pr_l(k)\omega_l(k)}$$ where

(C-3) $$\Pr_i(k) = \frac{1}{\sqrt{2\pi * \sigma}} Exp\{-\frac{1}{2 * \sigma}[\frac{z_k - \hat{n}_i(k)}{z_k}]^2\}$$

Notice that for each filter i, $\omega_i(k+1)$ decreases as the mean percentage forecasting error ($\frac{z_k - \hat{n}_i(k)}{z_k}$) increases. σ is the standard error of the mean percentage forecasting error, in our case, we set $\sigma = 1000$. Therefore, a filter that made poor predictions in previous rounds will be given a small weight in current forecasting. From (C-1) and (C-2), when the weight of a filter is reduced to 0, then this filter is eliminated from the estimation process.

Footnote

1 Following Mahajan, Muller, and Bass (1990) we use the terms *time-invariant* and *time-varying* to classify estimation methods. In the marketing literature, the term "time-varying" has been used to refer to two different categories of estimation methods: (1) estimation methods which start with a prior and update the prior as additional evidence accumulates (Bayesian updating procedures), and (2) methods which can estimate models with parameters changing over time. Here, we use the term *time-varying* by the first meaning, even though the method we introduce can apply to both categories.

2 In the rest of this article, the upper case bold letters denote matrices. Lower case bold letters denote column vectors. Italic lower case letters denote scalar variables and parameters.

3 The notation $\mathbf{y}_0 \sim (\tilde{y}_0, \mathbf{P}_0)$ indicates $\mathbf{y}_0$ is a random vector with expectation $E[\mathbf{y}_0] = \tilde{y}_0$ and covariance matrix $Cov[\mathbf{y}_0, \mathbf{y}_0] = \mathbf{P}_0$. Expectations and covariance are always unconditional unless otherwise indicated.

4 A continuous Kalman Filter is given as:

System equations: $\frac{d\mathbf{y}}{dt} = \mathbf{f}(\mathbf{y}, \beta, \mathbf{u}, t) + \mathbf{G}\mathbf{w}$ $\quad \mathbf{y}(0) \sim (\tilde{y}_0, \mathbf{P}_0),$ $\quad \mathbf{w} \sim (\mathbf{0}, \mathbf{Q}).$

Measurement equations: $\mathbf{z} = \mathbf{H}\mathbf{y} + \mathbf{v}$ $\quad \mathbf{v} \sim (\mathbf{0}, \mathbf{R}).$

5 If the parameters are time-varying and if one knows how the parameters change with time: $\frac{d\beta}{dt} \neq 0$, then the model will predict both sales and parameter values for the next period. If $\frac{d\beta}{dt} = 0$, then the model will predict sales only.

6 The prediction error may also be used implicitly by some *time-invariant* estimation methods when they are applied recursively to estimate parameters with each newly available observation. In Appendix B, we prove that AKF(C-D)'s updating formula allows it more efficiently use the prediction error as feedback in updating parameters than other methods. Simulation results also domenstrate AKF(C-D)'s advantage in following parameter changes.

7 Since the purpose of our empirical analysis is to provide a comparison between AKF(C-D) and other estimation methods in a general setting, we prefer to use a generic rule to construct priors rather than using product-specific knowledge. While setting prior variance to equal to prior mean is not a sophisticated way to set prior variances, it is a generic rule that is easy to apply.

Table 1

Summary of Procedures to Estimate the Bass Model

Estimation Procedure	Reference	Estimation Equation	Limitation
Ordinary Least Squares	Bass (1969) Heeler and Hustad (1980)*	$x(t)=\alpha_1(t\text{-}1)+\alpha_2(t\text{-}1)n(t\text{-}1)+\alpha_3(t\text{-}1)n^2(t\text{-}1)+\varepsilon(t)$	- Requires a discrete analog - Time-invariant
Maximum Likelihood Estimation**	Schmittlein and Mahajan (1982)	$L(x_i)=[1-F(t_{T-1})]^{x_T}\prod_{i=1}^{T-1}[F(t_i)-F(t_{i-1})]$	- Requires analytical solution - Time-invariant
Non-linear Least-Square Estimation	Srinivasan and Mason (1986)	$x(t_i)=m[F(t_i)-F(t_{i-1})]+\varepsilon(t)$	- Requires analytical solution - Time-invariant
Bayesian Updating in Meta Analysis	Sultan, Farley, and Lehmann (1990)	$x(t_i)=m[F(t_i)-F(t_{i-1})]+\varepsilon(t)$	- Requires analytical solution
Adaptive Filter	Bretschneider and Mahajan (1981)	$x(t)=\alpha_1(t\text{-}1)+\alpha_2(t\text{-}1)n(t\text{-}1)+\alpha_3(t\text{-}1)n^2(t\text{-}1)+\varepsilon(t)$	- Requires a discrete analog
Hierarchical Bayes	Lenk and Rao (1990)	$\frac{x(t)}{m}=c[F(t_i)-F(t_{i-1})]+\varepsilon(t)$	- Requires analytical solution

$x(t)$: Sales in period $[t\text{-}1,t)$

$n(t)$: The cumulative sales up to time t.

$F(t_i)$: The cumulative fraction of adopters at time t_i, where $F(t_i) = \dfrac{1-e^{-(p+q)t_i}}{1+\frac{q}{p}e^{-(p+q)t_i}}$, i=1,T

m: The size of the population of potential adopters.

c: Potential adoption rate.

T: The final period of estimation.

α_1, α_2, α_3: $\alpha_1=pm$, $\alpha_2=q\text{-}p$, and $\alpha_3=-\frac{q}{m}$, where p and q are coefficients of internal and external influence as defined in the Bass model.

*: They consider OLS produces a biased result, so they make some adjustments of estimated parameters by using empirical equations.

**: Define the likelihood function as $L(x_i)$.

Table 2

Prediction Error of AKF(C-D) and Other Methods

Period	Criterion	Method	Room Air Conditioners	Color Televisions	Clothes dryers	Ultrasound	Mammography	Foreign language	Accelerated program
Before Peak	MAD	AF	524.3	2,714.4	368.2	5.9	5.1	1.5	3.6
		AKF(C-D)	261.0	971.0	211.0	4.8	2.7	1.0	2.8
	MSE	AF	429,032	10,048,812	188,493	68.2	59.3	4.8	39.1
		AKF(C-D)	104,000	1,092,000	63,000	35.3	14.0	2.0	28.9
	MAPD	AF	51.27	71.3	47.5	50.1	40.5	48.5	83.1
		AKF(C-D)	40.1	33.0	35.2	56.5	27.8	50.1	85.5
After Peak	MAD	OLS	791.3	2,523.0	401.7	a	a	a	a
		MLE	454.7	1,218.5	316.0	5.1	6.1	3.7	2.6
		AE	794.3	340.0	487.3	5.8	11.7	1.2	4.3
		NLS	334.3	1,083.8	296.0	4.8	4.2	1.0	2.5
		AF	558.3	3,835.4	103.6	17.0	2.9	2.4	5.0
		AKF(C-D)	55	241.0	77.0	4.6	3.1	2.1	2.2
	MSE	OLS	648,993	7,099,978	171,583	a	a	a	a
		MLE	220,117	1,726,315	107,309	33.4	51.6	24.1	10.4
		AE	639,474	175,136	255,973	46.1	141.7	1.6	23.5
		NLS	129,326	1,548,310	97,662	24.3	33.0	1.9	9.2
		AF	330,247	14,914,101	24,000	289.9	13.6	6.3	44.0
		AKF(C-D)	4,000	59,000	9,000	38.6	15.4	7.8	5.4
	MAPD	OLS	48.3	46.5	30.3	a	a	a	a
		MLE	27.7	22.5	24.0	46.3	133.5	235.1	96.9
		AE	49.3	7.5	39.0	61.8	416.0	36.7	70.8
		NLS	20.3	19.0	22.3	41.5	62.0	66.0	39.0
		AF	33.9	72.5	7.4	123.6	103.6	85.8	61.3
		AKF(C-D)	3.6	4.58	5.92	27.49	77.46	133.2	53.3

[a] OLS yielded an incorrect sign for these parameters.

[b] $\text{MAD} = \frac{1}{K}\sum_{k=1}^{K} |x(k) - \hat{x}(k)|$, $\quad \text{MSE} = \frac{1}{K}\sum_{k=1}^{K} [x(k) - \hat{x}(k)]^2$, $\quad \text{MAPD} = \frac{1}{K}\sum_{k=1}^{K} \frac{100|x(k) - \hat{x}(k)|}{x(k)}$

where K is the number of forecasts, $x(k)$ is the observed number of incremental adopters in the kth time interval (t_{k-1}, t_k), and $\hat{x}_k$ is the predicted value of $x(k)$.

Table 3

Comparison between AKF(C-D) and Other Methods

Period	Method	Criterion	room air conditioner	color TV	clothes dryer	ultrasound	mammo-graphy	foreign language	accelerated program
Before Peak	AF	MAD	+	+	+	+	+	+	+
		MSE	+	+	+	+	+	+	+
		MAPD	+	+	+	-	+	-	-
After Peak	OLS	MAD	+	+	+	+	+	+	+
		MSE	+	+	+	+	+	+	+
		MAPD	+	+	+	+	+	+	+
	MLE	MAD	+	+	+	+	+	+	+
		MSE	+	+	+	-	+	+	+
		MAPD	+	+	+	+	+	+	+
	AE	MAD	+	+	+	+	+	-	+
		MSE	+	+	+	+	+	-	+
		MAPD	+	+	+	+	+	-	+
	NLS	MAD	+	+	+	+	+	-	+
		MSE	+	+	+	-	+	-	+
		MAPD	+	+	+	+	-	-	-
	AF	MAD	+	+	+	+	-	+	+
		MSE	+	+	+	+	-	-	+
		MAPD	+	+	+	+	+	-	+

A "+" indicates that AKF(C-D) provides a better prediction than that of corresponding procedure based on the corresponding criteria. A "-" means that AKF(C-D) prediction is worse.

Table A-1

Values used in AKF(C-D) Estimation

Prior Distribution (Mean = Variance)	*p*	All seven products:	0.01	
	q	All seven products:	0.1	
	m	room air conditioner and clothes dryer:	2×10^7	(40% of household)
		color TV:	4×10^7	(80% of household)
		medical equipment:	140	(2/3 of samples)
		education programs:	67	(2/3 of samples)
Noise Statistics	Process noise (Variance)	All durables:	10^5	
		All non-durables:	5	
	Observation noise (Standard Deviation)	All durables:	10% of observation	
		All non-durables:	0	

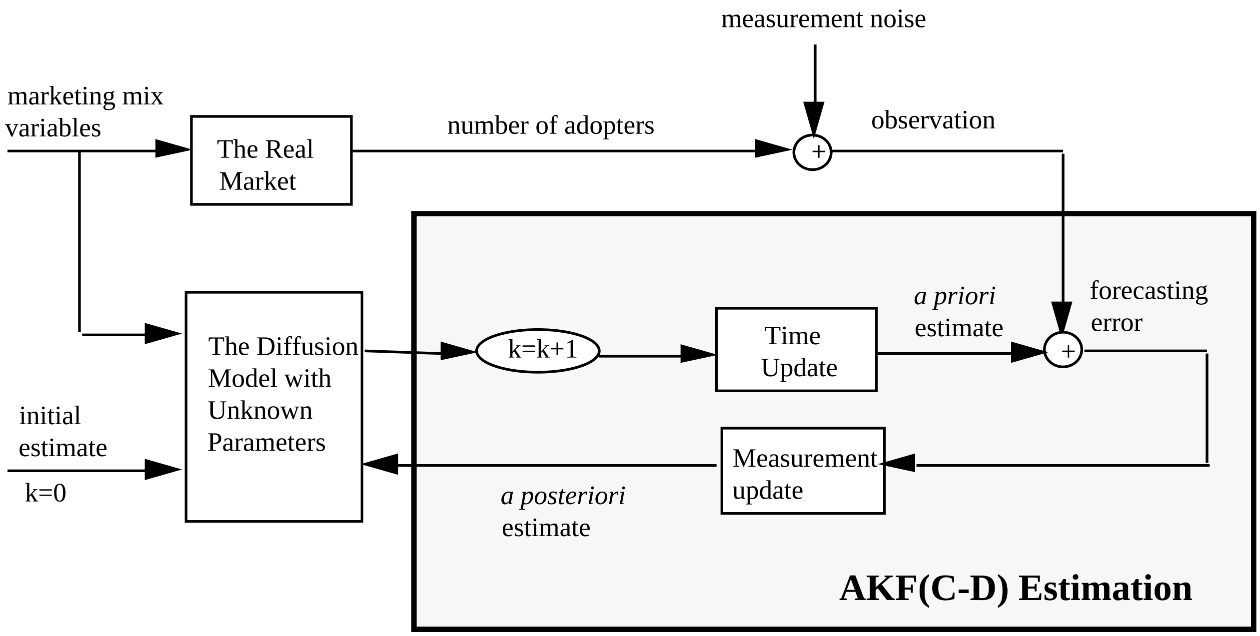


Figure 1

AKF(C-D) Estimation of Diffusion Models

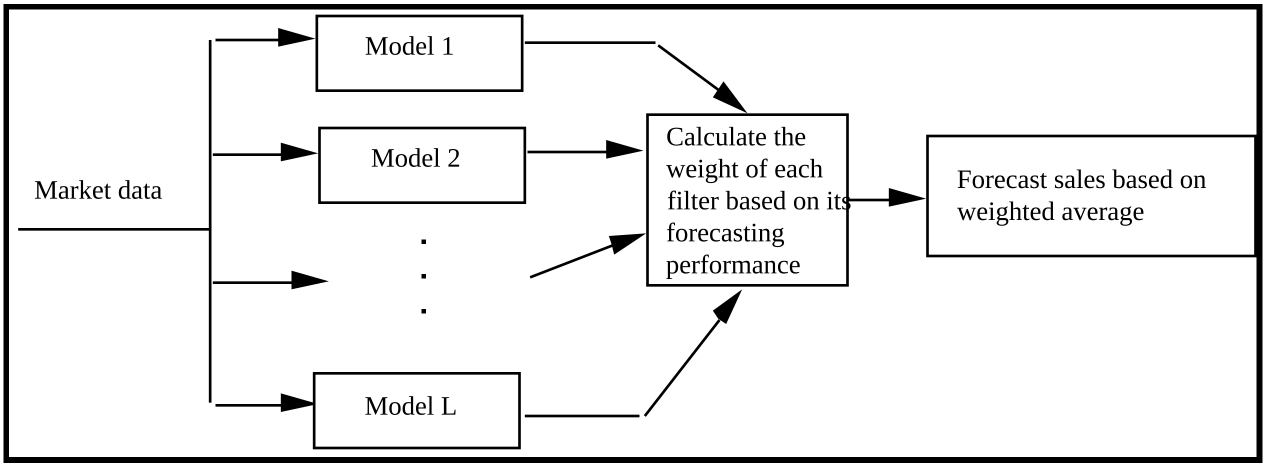
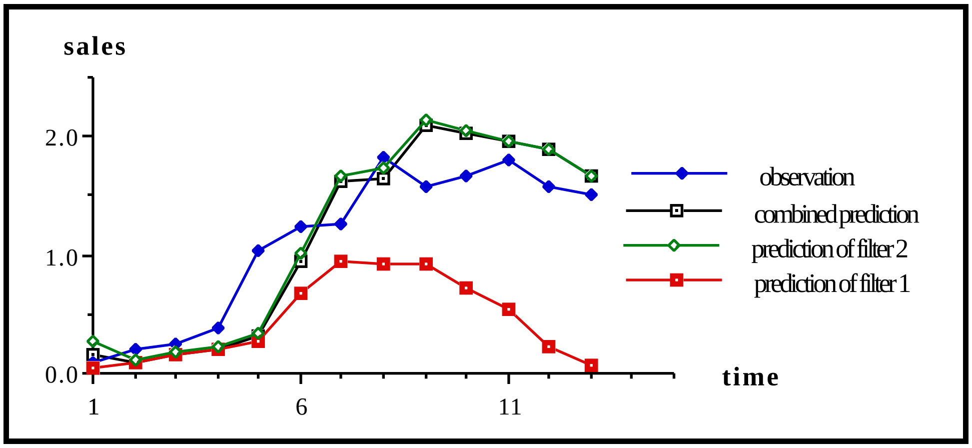
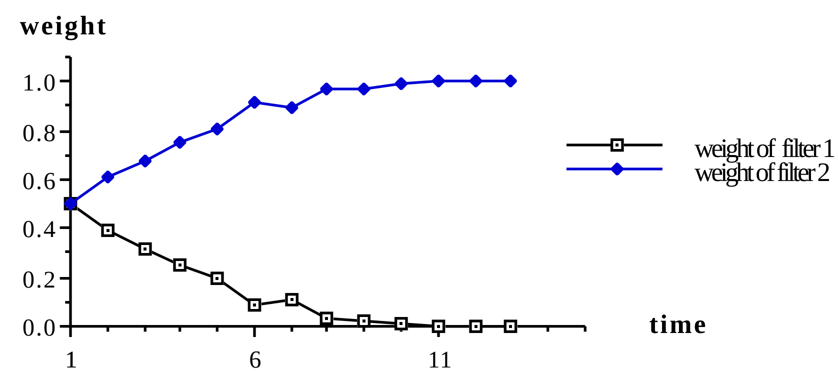


Figure 2

Parallel Filters Procedure



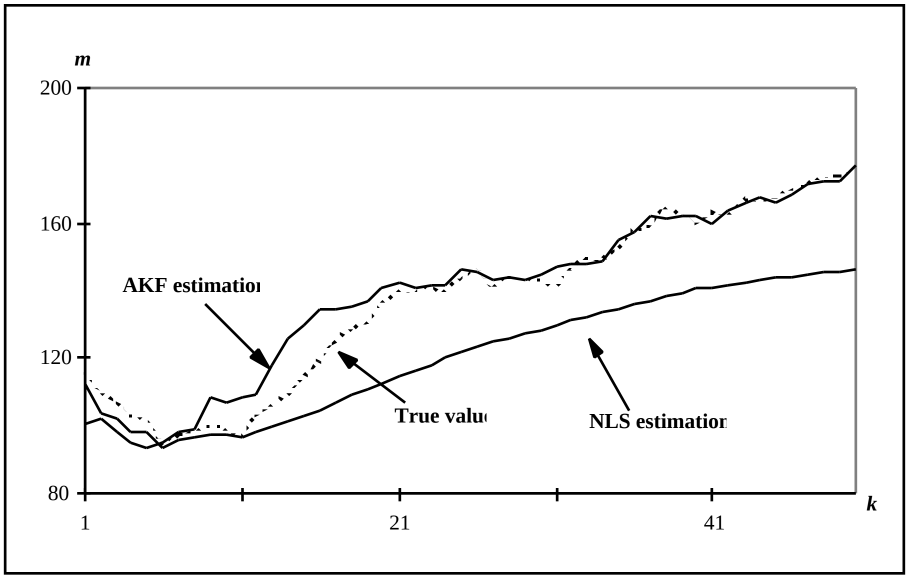
3-a. Observation, Predictions by Individual Filters and Combined Prediction Results



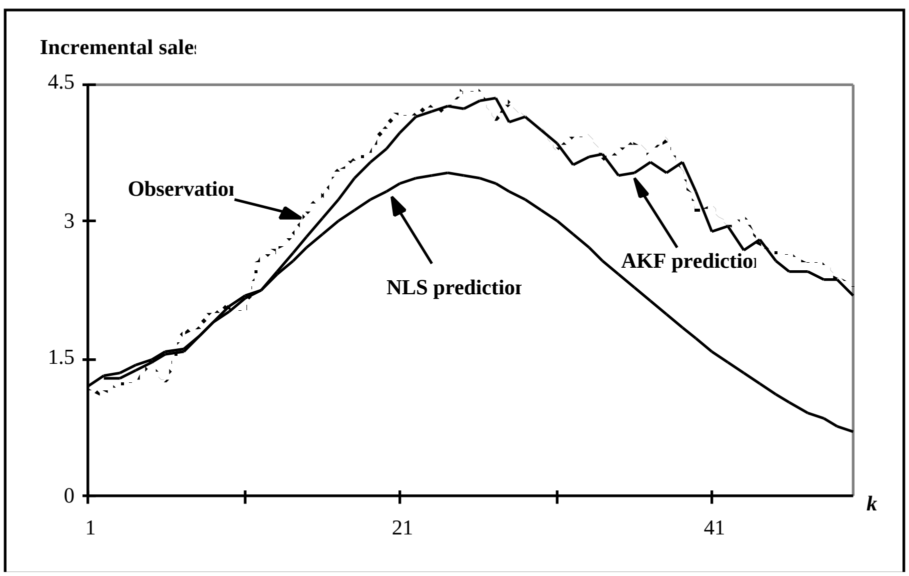
3-b. Weights of the Two Models

Figure 3

Using Parallel Filters Procedure to Estimate Diffusion of Room Air Conditioner



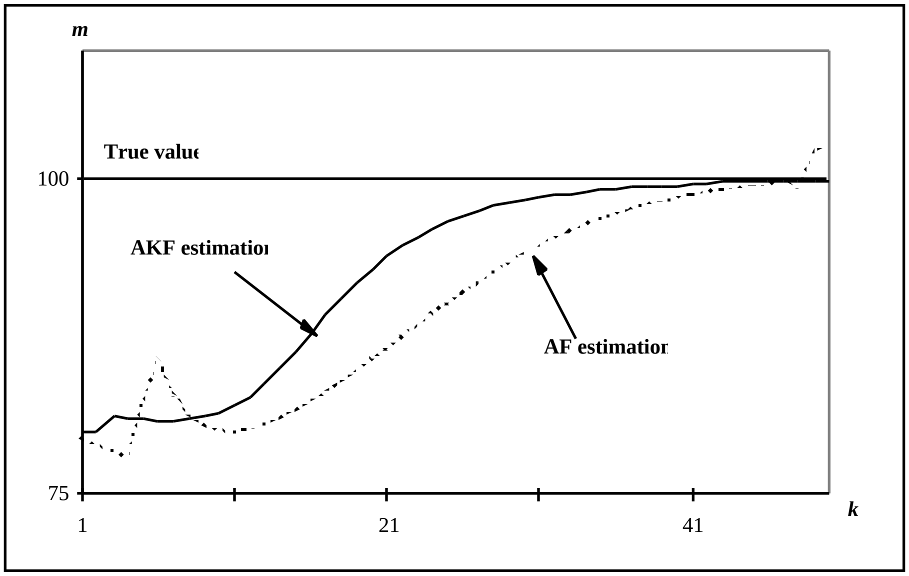
B1-a. Estimation of a Time-varying Parameter *m* (AKF vs. NLS)



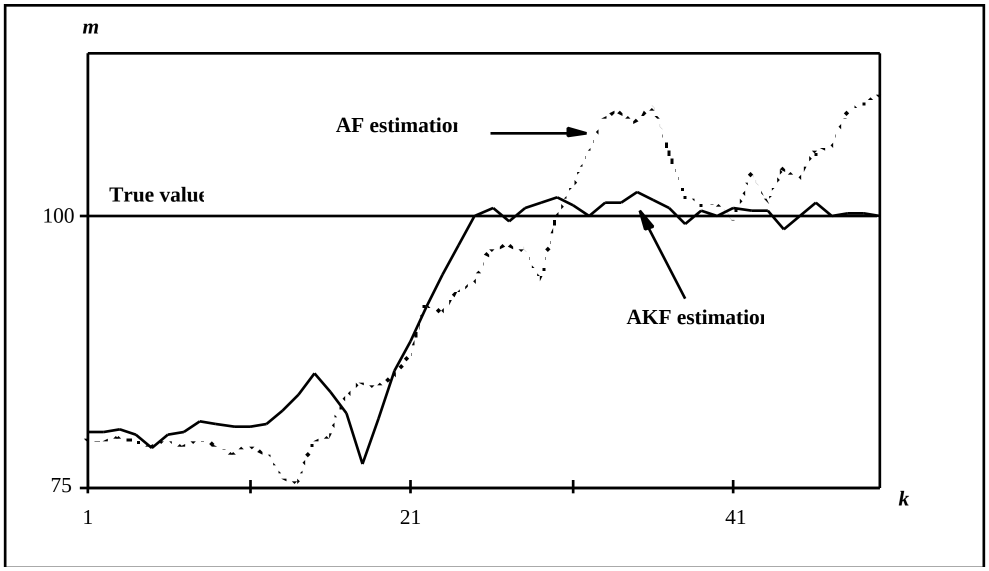
B1-b. Forecasting in the Presence of Parameter Changing Over Time (AKF vs. NLS)

Figure B1

AKF(C-D)'s Advantage in Estimating Time-varying Parameters



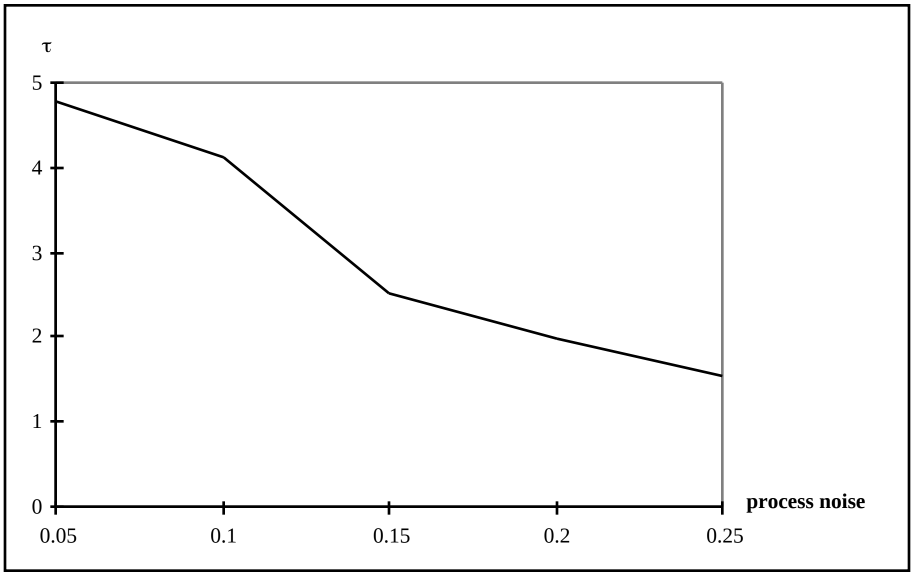
B2-a. Estimation of *m* in the absence of observation noise (AKF vs. AF)



B2-b. Estimation of *m* in the Presence of Observation Noise (AKF vs. AF)

Figure B2

AKF(C-D)'s Advantage in Explicitly Modeling Observation Error



$$\tau = \frac{\text{MAPD(MLS)}}{\text{MAPD(AKF)}}$$

Figure B3

The Impact of Process Noise on the Superiority of AKF(C-D)'s Prediction Power

References

Athans, Michael (1974), "The Importance of Kalman Filtering Methods For Economic Systems," *Annals of Economic and Social Measurement*, 3 (1), 49-64.

Bass, Frank M. (1969), "A New Product Growth Model for Consumer Durable," *Management Science*, 15 (January), 215-227.

Bretschneider, Stuart I. and Vijay Mahajan (1980), "Adaptive Technological Substitution Models," *Technological Forecasting and Social Change*, 18 (October) 129-139.

Carbone, Robert, and Richard L. Longini (1977), "A Feedback Model for Automated Real Estate Assessment, *management Science*, *24*, *241*-248.

Chen, Guangrong and Charles K. Chui (1991), "A Modified Adaptive Kalman Filter for Real-Time Applications," *IEEE Transactions on Aerospace and Electronic System* , 27 (January), 149-153.

Dhebar, Anirudh and Shmuel S. Oren (1985), "Optimal Dynamic Pricing for Expanding Networks," *Marketing Science*, 4 (Fall), 36-351.

Duncan, George, Wilpen L. Gorr, and Janusz Szczypula (1993), "Baysian Forecasting for Seemingly Unrelated Time Series: Application to Local Government Revenue Forecasting," *Management Science,* 39 (March), 275-293.

Easingwood, Christopher J., Vijay Mahajan, and Eitan Muller (1983), "A Nonuniform Influence Innovation Diffusion Model of New Product Acceptance," *Marketing Science*, 2 (Summer), 273-295.

Gorr, Wilpen L. and Cheng Hsu (1985), "An Adaptive Filtering Procedure for Estimating Regression Quantiles," *Management Science*, 31 (August), 1019-1029.

Heeler, Roger M. and Thomas P. Hustad (1980), "Problems in Predicting New Product Growth For Consumer Durables," *Management Science*, 26 (October), 1007-1020.

Horsky, Dan and Leonard S. Simon (1983), "Advertising and the Diffusion of New Products," *Marketing Science*, 2(Winter), 1-17.

Jain, Dipak C. and Ram C. Rao (1989), "Effect of Price on the Demand for Durables: Modeling, Estimation and Findings," *Journal of Business and Economic Statistics*, 8 (April), 163-170.

Johnston, John (1984), *Econometric Methods*, New York, McGraw-Hill Book Company.

Kahl, Douglas R. and Johannes Ledolter (1983), "A Recursive Kalman Filter Forecasting Approach," *Management Science*, 29 (November), 1325-1333.

Kalish, Shlomo (1983), "Monopolist Pricing With Dynamic Demand and Production Cost," *Marketing Science*, 2 (Spring), 135-160.

----------- (1985), "A New Product Adoption Model with Pricing, Advertising and Uncertainty," *Management Science*, 31 (December), 1569-1585.

Ledolter, Johannes and Douglas R. Kahl (1984), "Adaptive Filtering: An Empirical Evaluation," *Journal of the Operational Research Society*, 35 (4), 337-345.

Lenk, Peter J. and Ambar G. Rao (1990), "New Models From Old: Forecasting Product Adoption by Hierarchical Bayes Procedures," *Marketing Science*, 9 (Winter), 42-53.

Lewis, L. Frank (1986), *Optimal Estimation*, John Wiley & Sons, New York.

Lilien, Gary L., Ambar G. Rao, and Shlomo Kalish (1981), "Bayesian Estimation and Control of Detailing Effort in a Repeat Purchase Diffusion Environment," *Management Science*, 27 (May), 493-506.

Mahajan, Vijay, Charlotte H. Mason, and V. Srinivasan (1986), "An Evaluation of Estimation Procedures for New Product Diffusion Models," in Mahajan and Wind (eds.) *Innovation Diffusion Models of New Product Acceptance*, Cambridge, MA: Ballinger Publishing Co., 203-232.

Mahajan, Vijay, Eitan Muller, and Frank M. Bass (1990), "New Product Diffusion Models in Marketing: a Review and Directions for Research," *Journal of Marketing*, 54 (January), 1-26.

Mahajan, Vijay and Subhash Sharma (1986), "A Simple Algebraic Estimation Procedure for Innovation Diffusion Models of New Product Acceptance," *Technological Forecasting and Social Change* , 30 (December), 331-346.

Meade, Nigel (1985), "Forecasting Using Growth Curves: An Adaptive Approach," *Journal of the Operational Research Society*, 36, 1103-1115.

Morrison, G. Wayne and David H. Pike (1977), "Kalman Filtering Applied to Statistical Forecasting," *Management Science*, 23 (March), 768-778.

Norton, John A. and Frank M. Bass (1987), "A Diffusion Theory Model of Adoption and Substitution for Successive Generations of High-Technology Products," *Management Science* 33 (September), 1069-1086.

Schmittlein, David C. and Vijay Mahajan (1982), "Maximum Likelihood Estimation for an Innovation Diffusion Model of New Product Acceptance," *Marketing Science*, 1 (Winter): 57-78.

Simon, Herman and Karl-Heinz Sebastian (1987), " Diffusion and Advertising: The German Telephone Campaign," *Management Science*, 33(April), 451-466.

Slade, Margaret E. (1989), "Modeling Stochastic and Cyclical Components of Technical Change," *Journal of Econometrics*, 41, 363-383.

Srinivasan, V. and Charlotte H. Mason (1986), "Nonlinear Least Squares Estimation of New Product Diffusion Models," *Marketing Science*, 5 (Spring), 169-178.

Stengel, Robert F. (1986), *Stochastic Optimal Control: Theory and Application*, John Wiley & Sons, New York.

Sultan, Fareena, John U. Farley, and Donald R. Lehmann (1990), "A Meta-Analysis of Application of Diffusion Models," *Journal of Marketing Research*, 27 (February), 70-77.

Tegene, Abebayehu (1990), "The Kalman Filter Approach for Testing Structural Change in the Demand for Alcoholic Beverages," *Applied Economics* 22, 1407-1416.

----------- (1991), "Kalman Filter and the Demand for Cigarettes," *Applied Economics* 23 , 1175-1182.

Wildt, Albert R. and Russell S. Winer (1983), "Modeling and Estimation in Changing Market Environments," *The Journal of Business*, 56 (July), 365-388.

Xie, Jinhong and Marvin A. Sirbu (1995), "Price Competition and Compatibility in the Presence of Network Externalities," *Management Science*, 41. (May) 909-926.