MONTE CARLO INTEGRATION IN GENERAL DYMANIC MODELS

by

Peter Müller
Purdue University

Technical Report #90-17

Department of Statistics
Purdue University

April 1990

◇

Monte Carlo Integration in General Dynamic Models

Peter Müller *

Department of Statistics, Purdue University

March 28, 1990

Abstract

This paper suggests the use of Monte Carlo integration to estimate the required integrals in Bayesian analysis of dynamic models. The necessary prior and posterior samples are obtained by simulation.

*This research was supported by NSF grants DMS-8717799 and DMS-8702620 at Purdue University.

1 Introduction

The framework of time series modeling calls for an analysis which treats the model parameters as evolving, non constant quantities. As the dynamic model evolves through time, inferences about the parameters are updated in response to the incoming data and possibly other external information. This can be accomplished most naturally via a Bayesian approach.

Following the notation of Pole and West (1988) the general dynamic model can be defined as:

$$\text{Observation equation: } y_t = f_t(\theta_t) + \nu_t \tag{1}$$

$$\text{Evolution equation: } \theta_t = g_t(\theta_{t-1} + w_t). \tag{2}$$

Here f_t and g_t are known, non-linear functions and ν_t and w_t are random variables with arbitrary specified distribution.

This paper proposes an algorithm for the numerical evaluation of the integrals required in the analysis of the general dynamic model. In Section 2.1 these integrals are stated and it is argued why it is in general impossible to exactly evaluate them and why even numerical integration is difficult. Sections 2.2 and 2.3 review some restricted versions of the dynamic model

for which an exact analysis is feasible. Section 3 states the proposed Monte Carlo integration algorithm. The sample points required for the Monte Carlo integrals are obtained by simulating the dynamic model starting with a sample generated from the initial prior distribution. Section 3.2 gives the algorithm used for the simulation of the dynamic model, leaving the implementation details for the simulation of the observation step for Section 3.3. Section 4 contains an application of the algorithm to a model of advertising awareness taken from West and Harrison (1989).

2 The Analysis of the Dynamic Model

2.1 The General Dynamic Model

The following notation will help to describe the analysis of the general dynamic model stated in (1) and (2): Let D_t denote the information set at time t ($t = 0, 1, \ldots$), i.e. all information relevant to forming beliefs on the parameter vector θ_t. Suppose that the model is closed to inputs of external information, so that $D_{t+1} = \{y_t\} \cup D_t$. It will be assumed that all continuous probability distributions have a density defined with respect to Lebesgue measure, and $p(X)$ will be used generically to denote the density for a random variable X, and $p(X|D)$ the density of X given D. Expressions of the form $(X|D) \sim N(\mu, \sigma^2)$ refer to the conditional distribution of X given D.

Much of the Bayesian analysis in the dynamic model can be done in terms of integrals with respect to the distributions $p(\theta_t|D_{t-1})$ and $p(\theta_t|D_t)$ at time $t = 1, 2, \ldots T$. The distribution $p(\theta_t|D_{t-1})$ reflects the beliefs about θ_t prior to observing y_t. The density $p(\theta_t|D_t)$ summarizes the updated beliefs after observing the data y_t. Since in terms of the isolated experiment at time t $p(\theta_t|D_{t-1})$ represents the prior distribution on θ_t and $p(\theta_t|D_t)$ the posterior distribution, they will in the following be referred to as the prior at time t and the posterior at time t.

Given the initial prior $p(\theta_1|D_0)$, the unknown densities $p(\theta_t|D_t)$ and $p(\theta_t|D_{t-1})$ are recursively determined by

$$p(\theta_t|D_t) \propto p(\theta_t|D_{t-1})p(y_t|\theta_t) \tag{3}$$

$$p(\theta_t|D_{t-1}) = \int p(\theta_t|\theta_{t-1})p(\theta_{t-1}|D_{t-1})d\theta_{t-1}, \tag{4}$$

where $p(y_t|\theta_t)$ is the known likelihood function implied by the sampling model (1) and $p(\theta_t|\theta_{t-1})$ is determined by $g_t()$ and the distribution of ω_t in the system equation (2).

Inference typically requires integration of the prior and/or posterior densities. Examples include point estimation of the parameter vector based on the information available at time t, namely

$$\hat{\theta}_t = E(\theta_t|D_t) = \int \theta_t p(\theta_t|D_t)d\theta_t,$$

and one-step ahead forecast distributions,

$$p(y_{t+1}|D_t) = \int p(y_{t+1}|\theta_{t+1})p(\theta_{t+1}|D_t)d\theta_{t+1}.$$

K-step ahead forecast densities $p(Y_{t+k}|D_t)$ are also be easily obtainable in the framework of the suggested algorithm.

In general, we are interested in estimating integrals of the form

$$E(f(\theta_t)|D_t) = \int f(\theta_t)p(\theta_t|D_t)d\theta_t \quad (5)$$

$$E(f(\theta_t)|D_{t-1}) = \int f(\theta_t)p(\theta_t|D_{t-1})d\theta_t. \quad (6)$$

Here the integrands involve the unknown posterior density (3) or the prior density (4). The absence of conjugate prior/likelihood pairs and the possibly non-linear form of g_t and f_t in (1) and (2) make it in general impossible to analytically derive expressions for the prior distribution (4) and the posterior distribution (3) at time t. Therefore an exact evaluation of the integrals (5) and (6) is impossible, and even numerical integration is very difficult, since most schemes require evaluation of the integrand at specific points

Another feature which distinguishes the integration problem in the general dynamic model from other applications is the sequential nature of the problem, requiring estimation of a sequence of integrals over very closely related densities.

2.2 The Normal Dynamic Linear Model

An exact analysis of the dynamic model is only possible if some restrictions are imposed on the general dynamic model. The Bayesian analysis of dy-

namic models becomes most simple in the framework of the Kalman filter (Kalman (1960)) with the linear evolution and observation equation, which together with normal distributed noise terms and conjugate initial prior lead to a sequence of conjugate prior/likelihood pairs, making the computational burden of evaluating integrals of the form (5) or (6) trivial. This linear dynamic Bayesian model was described and further developed by e.g. Ho and Lee (1964) and Harrison and Stevens (1976).

Using the notation of Pole, West and Harrison (1988) the basic normal dynamic linear model (NDLM) can be stated as:

$$\begin{aligned}
\text{Observation Equation: } y_t &= F_t'\theta_t + \nu_t \\
\nu_t &\sim N(0, V_t) \\
\text{Evolution Equation: } \theta_t &= G_t'\theta_{t-1} + \omega_t \\
\omega_t &\sim N(0, \Omega_t) \\
t &= 1, 2, \ldots
\end{aligned}$$

Here y_t is the observation vector, F_t a vector of regression variables, θ_t the parameter vector, G_t the state transition matrix, V_t the observation variance and Ω_t the system variance. The quadruple $M = \{F_t, G_t, V_t, \Omega_t\}$

fully specifies the model.

With conjugate initial prior $(\theta_1|D_0) \sim N(a_1, R_1)$ the prior $p(\theta_t|D_{t-1})$ and posterior $p(\theta_t|D_t)$ are normal with means and variances, which are derived as follows by induction. Assume the prior at time t is already given by $(\theta_t|D_{t-1}) \sim N(a_t, R_t)$. Then the posterior becomes:

$$(\theta_t|D_t) \sim N(m_t, C_t),$$

with $m_t = a_t + A_t e_t$, $C_t = R_t - A_t Q_t A_t'$, $Q_t = F_t' R_t' Ft + V_t$, $A_t = R_t F_t Q_t^{-1}$ and $e_t = y_t - F_t' a_t$. The prior for the subsequent period is then given by

$$(\theta_{t+1}|D_t) \sim N(a_{t+1}, R_{t+1}),$$

where $a_t = G_t m_{t-1}$ and $R_t = G_t C_{t-1} G_t' + \Omega_t$. See West and Harrison (1989) for a proof and more discussion. The NDLM covers a wide range of models, allowing incorporation of trend and seasonal components in G_t, a regression term in F_t and possible intervention, just to name a few of the modeling tools developed in West and Harrison (1989).

2.3 Extensions of the NDLM

Pole, West and Harrison (1988) extended the basic NDLM to allow for non-linearities in the observation equation by introducing a "guide relationship" and assuming a normal prior at each step. The guide relationship is used as an aid in determining the first two moments of the posterior, which together with the normality assumption and the linear evolution suffices to specify the next prior. Another extension of the NDLM is achieved by replacing the normality assumption by a wider class of conjugate priors. Sorenson and Alspach (1971) propose a discrete mixture of normals to represent observation and system noise term.

West, Harrison and Migon (1985) developed an algorithm to analyze dynamic generalized linear models (DGLM). The DGLM is an extension of the NDLM, allowing any density from an exponential family as sampling distribution for the observation y_t and specifying the distribution of the evolution noise term only by its first and second order moments. Let θ_t denote the state vector at time t and η_t the natural parameter of the exponential family sampling distribution $p(y_t|\eta_t)$. Then the observation equation is given by:

$$\text{Observation model: } p(y_t|\eta_t) \quad \propto \quad exp\{V_t^{-1}[y_t\eta_t - a(\eta_t)]\} \qquad (7)$$

$$g(\eta_t) \quad = \quad \lambda_t = F_t'\theta_t, \tag{8}$$

where V_t is a known scale parameter, $a(\eta)$ is a known function, F_t is a vector of regression variables, and $g(.)$ is a known mapping between the parameter η_t and the linear regression λ_t. The evolution equation specifies the distribution of the evolution noise term only by its first and second moments:

$$\begin{aligned} \text{Evolution equation: } \theta_{t+1} &= G_t\theta_t + \omega_t \\ E(\omega_t) &= 0 \\ \text{Var}(\omega_t) &= W_t. \end{aligned}$$

The analysis proceeds by determining, at any time, only the first and second moments of the prior and posterior for the state vector. The distribution of $(\eta_t|D_{t-1})$ is assumed in a form conjugate to the likelihood $p(y_t|\eta_t)$, namely

$$p(\eta_t|D_{t-1}) \propto exp\{s_t[x_t\eta_t - a(\eta_t)]\},$$

for some parameters s_t and x_t. This leads then to a sequence of simple updating equations for the moments of the prior $p(\theta_t|D_{t-1})$ and posterior

$p(\theta_t|D_t)$, which would correspond to an analysis with fully specified distributions, if the prior and posterior distributions were actually normal and the forecast distribution $p(\eta_t|D_{t-1})$ was actually conjugate to the sampling distribution. See West and Harrison (1989), Section 14.3 for a complete description of this approach.

The accessible range of models can be even further widened by abandoning the restriction to analytically tractable models and using numerical integration methods. Pole and West (1988) propose an algorithm combining the computational convenience of conjugate models and the flexibility of using numerical integration. They deal with models which are NDLM conditional on a subset of parameters β:

$$M(\beta) = \{F_t(\beta), G_t(\beta), V_t(\beta), \Omega_t(\beta)\}.$$

The model is analyzed by Gaussian quadrature with respect to β, using the analytical solution for $\{F_t(\beta), G_t(\beta), V_t(\beta), \Omega_t(\beta)\}$ for fixed β's. Gaussian quadrature involves at each stage maintaining an optimal grid $K_t = \{\beta_{t,i}\}$ which is used to estimate the posterior integrals and forecast densities. When updating the grid K_{t-1} to K_t, the $p(\beta_t|D_t)$ and the moments of the NDLM parameters conditional on the β's need to be determined on the

new grid points. Pole and West (1988) suggest using spline interpolation. Linear interpolation is used to maintain required positive definiteness of $C_t(\beta) = Var(\theta_t|\beta_t, D_t)$.

3 The Algorithm

3.1 Monte Carlo Integration in the General Dynamic Model

In Section 1.1. the general dynamic model was stated as

$$\text{Observation equation: } y_t = f_t(\theta_t) + \nu_t \tag{9}$$

$$\text{Evolution equation: } \theta_t = g_t(\theta_{t-1} + w_t), \tag{10}$$

with f_t and g_t known, non-linear functions and ν_t and w_t random variables with arbitrary specified distribution. As argued in Section 2.1, under this model it is in general impossible to exactly evaluate the posterior integrals (5) and prior integrals (6). Even numerical integration is difficult since the posterior (3) and the prior (4) cannot be evaluated. The algorithms mentioned in Sections 2.2 and 2.3 avoid this problem by either making distributional assumptions or by estimating density values, e.g. by spline interpolation or linear interpolation.

One numerical integration approach which fits naturally into the dynamic model framework is Monte Carlo integration (with importance sampling) as described, e.g., in Rubinstein(1981), van Dijk(1984) or Geweke(1989), with the sample points for the Monte Carlo integration coming from a sim-

ulation of the dynamic model. A sample from the original prior $p(\theta_1|D_0)$ is propagated step by step through the time series, thereby always maintaining a sample from the prior and posterior at each time step; these can be used for Monte Carlo estimates of prior and posterior integrals as desired. The task of evaluating the posterior and prior is replaced by the problem of simulating the observation and evolution steps by an appropriate process applied to the sample points.

No restrictions on the functional form of the observation equation (1) or on the system equation (2) are required, i.e., f_t and g_t can be any, not necessarily linear, functions. Although no global assumptions about the involved distributions will be necessary, the simulation of the observation step will require that the prior $p(\theta_t|D_{t-1})$ be approximable by a normal density over certain regions.

3.2 Simulating the Dynamic Model

In the following, let π_t denote the prior $p(\theta_t|D_{t-1})$, p_t the posterior $p(\theta_t|D_t)$ and l_t the likelihood function $p(y_t|\theta_t)$ as a function of θ_t. The simulation of the evolution step is simple: Assume a sample $\{\eta_i, i = 1, \ldots n\}$ from the posterior p_{t-1} is available. Then these sample points are easily mapped into a sample $\{\theta_i, i = 1, \ldots n\}$ from π_t by $\theta_i := g_t(\eta_i + \omega_i)$, where the ω_i are

generated from the distribution given by the system equation (10).

The simulation of the observation step, i.e. the transformation of the prior sample from π_t to a posterior sample from p_t, would be conceptually straightforward by an accept/reject procedure: Bayes' theorem, $p_t(\theta_i) \propto \pi_t(\theta_i) l_t(\theta_i)$, implies that the prior sample could be transformed to a posterior sample by deleting and retaining sample points in an accept/reject like way with probabilities $P(\theta_i \text{ is accepted}) \propto l_t(\theta_i)$. This could be implemented by generating $u_i \sim \text{Uniform}(0,1)$ and comparing u_i with $P(\theta_i \text{ is accepted})$: if $u_i > P(\theta_i \text{ is accepted})$ then delete θ_i from the sample, otherwise keep θ_i in the sample. This only requires the evaluation of the likelihood function l_t, which is analytically available; the difficult evaluation of the densities π_t and p_t is avoided. Unfortunately, each accept/reject step will typically lead to a decrease in the sample size, making an unmodified implementation of this approach impractical for simulation of a many-step dynamic model.

The proposed algorithm solves this problem by replacing the accept/reject step by a two step procedure. First the prior sample is transformed into a sample from an envelope density I, which is then in a second step used for the accept/reject procedure. The envelope density as specified in Algorithms 1 and 2 is chosen such that

- a sample from I can be generated by expanding the available prior sample by stratified additional sampling from π_t, restricted to certain regions L_m,
- the accept/reject weights are functions of l_t only and
- the expected sample size after the accept/reject step is equal to the original prior sample size.

The sets L_m are chosen as regions of high likelihood in a way formalized in Algorithm 2.

The following specific algorithm is proposed:

Algorithm 1: Simulation of the General Dynamic Model

Initial Prior: Draw a sample from the initial prior: $\theta_i \sim \pi_1$, $i = 1, \ldots n_0$.

Set Time: $t := 1$.

Observation Step: The details of the implementation of this step will be explained in Algorithm 2.

Extension of Sample Size: By stratified additional sampling from π_t the sample $\{\theta_i, i = 1, \ldots n_0\}$ is extended to a sample from a

density of the form:

$$I \in \mathcal{G} = \{f : f = \sum_{m=1}^{M} \gamma_m \pi_t|_{Lm}, \sum \gamma_m = 1\},$$

where $\pi_t(x)|_{Lm} := \begin{cases} \pi_t(x)/\pi_t(L_m) & \text{if } x \in L_m \\ 0 & \text{otherwise} \end{cases}$,

denotes π_t restricted to L_m. The regions L_m will be specified in equation (11) in Algorithm 2.

Accept/Reject: The density I is now used as envelope density for an accept/reject step which, results in a sample from the posterior p_t:

$$P(\theta_i \text{ is deleted }) \propto 1 - \frac{p_t(\theta_i)}{I(\theta_i)}.$$

Posterior Sample: The posterior sample $\{\theta_i, i = 1, \ldots n_t\}$ can now be used for Monte Carlo estimates of posterior integrals. By choosing $I \in \mathcal{G}$ appropriately some minimum expected sample size for the posterior sample can be guaranteed.

Evolution Step: Generate $\omega_i, i = 1, \ldots n_t$ and set $\eta_i := g_t(\theta_i) + \omega_i$. The set $\{\eta_i, i = 1, \ldots n_t\}$ now forms a sample from π_{t+1}, which can be used to estimate prior integrals.

Iteration: Increment $t := t + 1$ and simulate the next observation step.

The estimates of the posterior integrals can be improved by using Monte Carlo with importance sampling with the sample from the envelope density I. The accept/reject weights p_t/I would become the importance sampling weights. This point will not be further explored here, since this is a refinement of the algorithm which is not essential to the main idea.

3.3 Simulating Bayes' Theorem

To simulate the observation step by Algorithm 1, it is still necessary to specify the choice of $I \in \mathcal{G}$. In the following, π will denote the prior $p(\theta_t|D_{t-1})$ at time t, l the likelihood function $p(y_t|\theta_t)$ and p the posterior $p(\theta_t|D_t)$.

The envelope density of Algorithm 1, $I = \sum \gamma_m \pi|_{Lm}$, can alternatively be written recursively as

$$I_0 = \pi, \ldots, I_m = \alpha_m I_{m-1} + (1 - \alpha_m)\pi|_{Lm}, \ldots, I = I_M.$$

This suggests that the available sample from π can be transformed into a sample from I_M by iterative expansion of the sample size by drawing additional sample points from $\pi|_{Lm}$.

Simulation of the observation step, i.e. Bayes' theorem, can then be done

by the following algorithm:

Algorithm 2: Simulation of Bayes' Theorem

Assume that a sample $\{\theta_i, i = 1, \ldots n\}$ from the current prior π is available.

Initialization: Estimate $\bar{l} = E_\pi l$ by $\frac{1}{n}\sum_1^n l(\theta_i)$ and set $m := 1$. Set the initial accept/reject weights $w_0(\theta_i) = p/I_0 = p/\pi$ to:

$$w_0(\theta_i) = l(\theta_i).$$

Iteration: The following steps iteratively expand the original prior sample to a sample from the envelope density I_M.

Additional Sampling: Increase the sample size by a factor $\frac{1}{\alpha_m}$ by additional sampling from $\pi|_{Lm}$, where

$$L_m = \{\theta : w_{m-1}(\theta) > \bar{l}\}. \tag{11}$$

This expands the current sample to a sample from I_m.

Weights: The new weights p/I_m are given by:

$$\frac{p}{I_m} \propto w_m := \begin{cases} w_{m-1} & \text{if } \theta \notin L_m \\ q_m w_{m-1} & \text{if } \theta \in L_m \end{cases} \tag{12}$$

where

$$q_m := (1 + \frac{1-\alpha_m}{\alpha_m} \frac{1}{I_{m-1}(L_m)})^{-1}.$$

Stopping Rule: Set $m := m + 1$ and continue iterating until $m = M$, where

$$M = min\{m : \prod_{k=1}^{m} q_k \leq \frac{\bar{l}}{l^*}\}. \tag{13}$$

Here $l^* = \sup(l(\theta))$.

Accept/Reject: Reject sample points with probability

$$P(\theta_i \text{ is deleted }) \propto 1 - \frac{p_t(\theta_i)}{I(\theta_i)} = 1 - w_M.$$

Posterior Sample: The posterior sample $\{\theta_i, i = 1, \ldots n\}$ can now be used for Monte Carlo estimates of posterior integrals.

To derive the expression (12) for w_m, observe the following: The regions L_m are such that $L_M \subset \ldots \subset L_1$, implying $\pi|_{Lm} = I_{m-1}|_{L_m}$. With this the

expression for I_n can be written as:

$$\begin{aligned} I_m &= \alpha_m I_{m-1} + (1-\alpha_m) I_{m-1}|_{L_m} \\ &\propto \begin{cases} I_{m-1} & \text{if } \theta \notin L_m \\ I_{m-1}(1 + \frac{1-\alpha_m}{\alpha_m} \frac{1}{I_{m-1}(L_m)}) \quad = I_{m-1}/q_m & \text{if } \theta \in L_m. \end{cases} \end{aligned}$$

From this, (12) follows.

The number of iterations (13) is chosen such that the expected sample size of the posterior sample after the accept/reject step is greater than or equal to the original prior sample size. This leads to the stopping rule (13) (see appendix). If the α_m are chosen such that $q_m \leq \frac{1}{2}$ then obviously

$$M \approx \log(2) log \frac{l^*}{\bar{l}}.$$

There remains the problem of generating from $\pi|_{Lm}$. This is the point at which the simulation algorithm requires some approximation; in particular it will be assumed that π can, on the regions L_m, be approximated by a multivariate normal distribution restricted to L_m. Compared with the global distributional assumptions and linearity restrictions on the functional form of system and evolution equation that are required for available analytical and numerical algorithms, this local approximation of the prior density $\pi|_{Lm}$

seems quite weak. The bias which this approximation introduces into the Monte Carlo integral estimates does not accumulate over time, since adding the noise term ω_t in each evolution step amounts to "discounting" the past history at a certain rate, so that typically only the most recent observations have a strong influence on the prior and posterior at any given point in time.

The following simple example illustrates the suggested algorithm:

Example:

Assume $X \sim N(\theta, 1)$, where θ is unknown. Let the prior $\pi(\theta)$ be $N(2, 1)$. Then, if $x = -1$ is observed, the likelihood is $l(\theta) = \phi(\theta - (-1))$, where $\phi(.)$ denotes the standard normal density, giving the posterior $p(\theta)$ as $(\theta|x) \sim N(0.5, 0.5)$.

Following Algorithm 2, we first drew a sample from the prior: $\theta_i \sim N(2,1), i = 1, \ldots 100$. In the first step of the iteration $L_1 = \{l(\theta_i) > \bar{l}\}$ was the interval $(-\infty, 1.29]$, where 1.29 was obtained as $\max\{\theta_i : l(\theta_i) > \bar{l}, i = 1, \ldots 100\}$. Since the likelihood function is analytically available, it would obviously be possible to obtain L_1 exactly. But the only penalty for not accurately identifying L_m is that the expected final posterior sample size will not exactly match the initial prior sample size. The parameter α_1 was (rather arbitrarily) chosen such that q_1 evaluates to 1/2. With all parameters set, the additional sample was taken from $\pi|L_1$, increasing the

total sample size to 122. The new points were added to the sample, and the weights updated to $w_1(\theta) := w_0(\theta)/2 = l(\theta)/2$, if $\theta \in L_1$, respectively left unchanged $w_1(\theta) := w_0(\theta)$ otherwise.

The stopping rule was determined by the ration $\bar{l}/l^*$, which was estimated to be .076, based on the likelihood values $l(\theta_i)$ of the sampled points. Although $l^* = \sup(l(\theta))$ could have been obtained analytically here, this was not done, since underestimation of l^* by taking the sample maximum will only affect the posterior sample size.

The stopping criterion was not yet met. So the iteration continued with $L_2 = \{w_1 > \bar{l}\} = (-\infty, 0.97]$. In the second iteration the sample size was increased to 151. After the forth cycle the stopping criterion was met, and a sample of 191 points resulted from the last envelope density I_4. The final accept/reject step reduced the sample size back to $n = 106$.

The graph in Figure 1 shows the prior density, the four envelope densities and the posterior density, plotted proportional to the sample size; i.e. , if one curve is above another, then a sample corresponding to the first curve can always be transformed by accept/reject to a sample corresponding to the second. The algorithm continues to expand the sample size until the curve corresponding to the final envelope density I_4 is everywhere above the posterior density; i.e., the envelope density sample can by accept/reject be

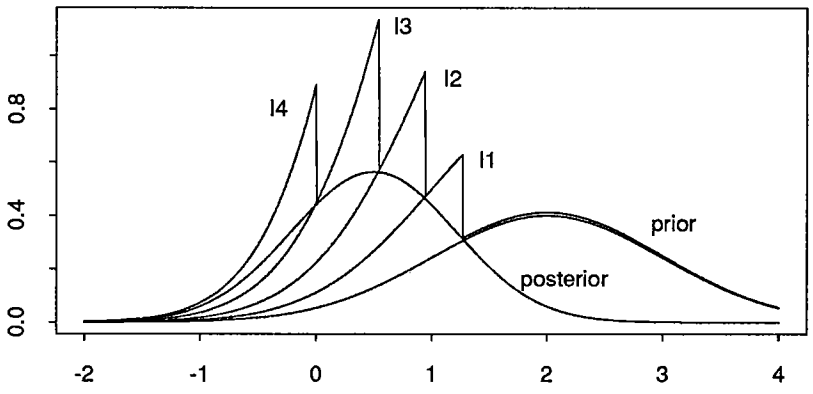

Figure 1: Prior π, envelope densities I_1, I_2, I_3, I_4 and posterior p.

transformed to a posterior sample with expected sample size 100.

4 A Model of Advertising Awareness

4.1 The Model

This non-linear, non-normal example is taken from West and Harrison (1989). The number of individuals in a survey who are "aware" of certain TV commercials is modeled. Awareness is measured by the proportion of the TV viewers who have seen these TV commercials during a recent, fixed time interval. The data was collected by questioning in weekly intervals sampled members of the TV viewing population as to whether or not they have seen the TV commercials in question. The survey was taken over a period of 75 weeks with a constant sample size of $n = 66$.

The variables involved are:

Y_t: number of positive responses out of the sample size n

X_t: extent of advertising (measured in units called "TVR").

The time series for Y_t and X_t are plotted in Figure 2 and 3.

The following parameters are used to model the total population awareness at time t:

α_t : lower threshold of awareness, i.e. the minimum level of awareness expected at time t.

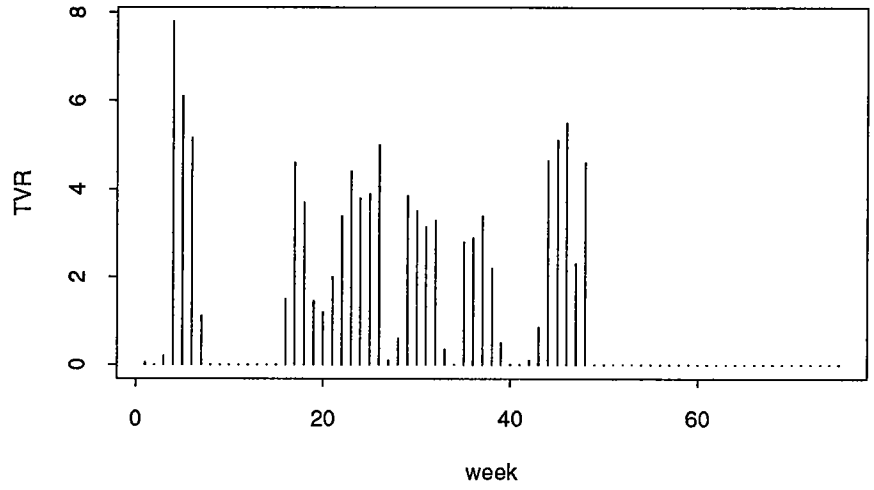

Figure 2: Extent of advertising in TVR.

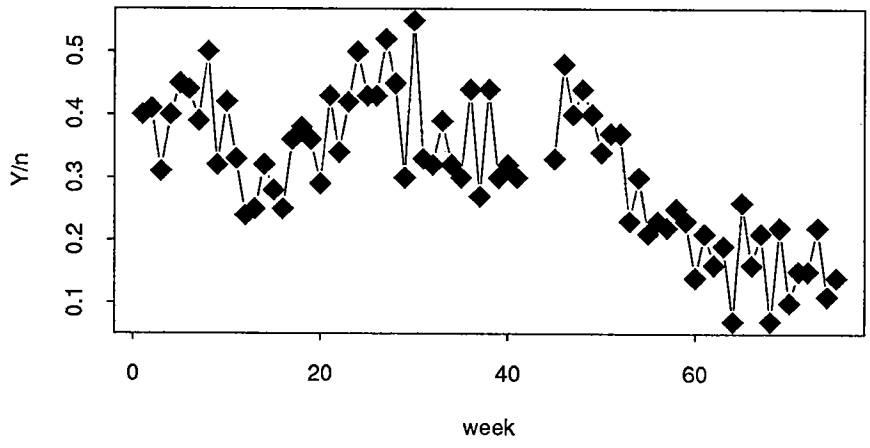

Figure 3: Awareness response as proportion: Y_t/n_t (weeks 42, 43 and 44 have missing observations).

β_t : upper threshold of awareness, i.e. the maximum level of awareness expected at time t.

E_t : effect of current and past weekly advertising, adding up with α_t to the total population awareness $\mu_t = E_t + \alpha_t$ at time t .

ρ_t : decay of awareness: if $X_t = 0$ then the effect of past advertising is expected to decay exponentially by: $E_t = \rho_t E_{t-1}$.

κ_t : penetration parameter: the meaning of this parameter becomes clear from the evolution equation.

The deterministic part of the evolution equation is given by:

$$E_t = \rho_t E_{t-1} + [1 - \exp(-\kappa_t X_t)][\beta_t - (\alpha_t + \rho_t E_{t-1})]. \tag{14}$$

The effect of past advertising is decaying by the factor ρ_t. The effect of current advertising is proportional to $[1 - \exp(-\kappa_t X_t)]$ and bounded above by $[\beta_t - (\alpha_t + \rho_t E_{t-1})]$.

Let $\theta_t = (\alpha_t, \beta_t, E_t, \rho_t, \kappa_t)'$ denote the parameter vector at time t. Assume a multivariate normal evolution noise with a covariance matrix corre-

sponding to a discount factor of 0.97:

$$\omega_t \sim N(0, U_t) \text{ with } U_t = 0.03C_t,$$

where C_t is the variance of the current posterior $p(\theta_t|D_t)$. The concept of using a discount factor to specify W_t corresponds to specifying the relative increase of the current "uncertainty" C_t. See West and Harrison (1989), section 2.4.2 for a more complete justification.

The complete model takes the form:

$$\begin{aligned}
\text{Observation equation: } p(Y_t|\mu_t) &\propto \mu_t^{Y_t}(1-\mu_t)^{n-Y_t} \\
\mu_t &= \alpha_t + E_t \\
\text{Evolution equation: } \theta_{t+1} &= g(\theta_t + \omega_t) \\
\omega_t &\sim N(0, U_t) \\
n &= 66 \text{ and } t = 1, \ldots 75,
\end{aligned}$$

where $g()$ is the identity in $\alpha_t, \beta_t, \rho_t$ and κ_t and determined by (14) for E_t. This is a non-linear, non-normal dynamic model with a system equation in the form of (10) and a sampling model slightly more general than (9).

Following West and Harrison the initial prior is chosen as

$$\begin{aligned}
\theta_0|D_0 &\sim N(m_0, C_0) \\
m_0 &= (0.1, 0.85, 0.9, .02, 0.3)' \\
C_0 &= 0.0001 \begin{bmatrix} 6.25 & 6.25 & 0 & 0 & 0 \\ 6.25 & 406.25 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 2.25 & 0 \\ 0 & 0 & 0 & 0 & 100 \end{bmatrix}.
\end{aligned}$$

The initial distribution is denoted here by $p(\theta_0|D_0)$ rather than $p(\theta_1|D_0)$ as in the description of Algorithm 1, because it models the beliefs on θ before the first evolution step.

4.2 Analysis as a Dynamic Generalized Linear Model

In their analysis of the advertising awareness model, West and Harrison first linearized the evolution equation:

$$\theta_t \approx g_t(m_{t-1}) + G_t * (\theta_{t-1} + \omega_t - m_{t-1}),$$

where

$$\begin{aligned} m_{t-1} &= E(\theta_{t-1}|D_{t-1}) \\ G_t &= \frac{\partial g_t(z)}{\partial z}|_{m_{t-1}}. \end{aligned}$$

The linearized model was then analyzed as a dynamic generalized linear model, following the DGLM algorithm described in Section 2.3 The sampling model (7) of the DGLM specializes to (The analysis is easier in terms of the parameter μ_t rather than the natural parameter $\eta_t = log(\mu_t/(1-\mu_t).)$:

$$\begin{aligned} p(Y_t|\mu_t) &\sim \text{Binomial}(\mu_t, n) \\ \mu_t &= (1,0,0,0,1)'\theta_t = \alpha_t + E_t. \end{aligned}$$

Using the updating equations of the DGLM, as given e.g. in West and Harrison (1989) in Section 14.3.3., it is then straightforward to derive for each week $E(\theta_t|D_{t-1})$, $E(\theta_t|D_t)$ and other quantities of interest in the analysis of the model. In the next section the resulting data analysis is compared with the analysis obtained from the simulation algorithm.

4.3 Analysis by Simulation

The same model was analyzed with the simulation algorithm suggested in Section 3. The simulation was started with a sample of size $n = 100$ from the initial distribution $p(\theta_0|D_0)$. (The prior specified by West and Harrison models the beliefs on θ before the first evolution step. Therefore we have to start with $p(\theta_0|D_0)$, rather than $p(\theta_1|D_0)$, as used in Algorithm 1.) By simulating the first evolution step, this sample was then transformed into a sample from the prior distribution $p(\theta_1|D_0)$. Then the information updating in the first observation step is simulated by Algorithm 2, leading to a sample from the posterior $p(\theta_1|D_1)$. From there the process iterates through all time steps, making at any time t a prior and a posterior sample available, which can be used to estimate prior and posterior integrals of the form (6) and (5).

The one step ahead forecast function gives the expected response in week $t+1$, based on the currently available information D_t:

$$f_t = E(\mu_{t+1}|D_t) = \int (\alpha_{t+1} + E_{t+1}) p(\theta_t|D_{t-1}) d\theta_{t+1}.$$

This is an integral of the form (6) and can be estimated by a Monte Carlo integral using the sample points from $p(\theta_t|D_{t-1})$. The forecasts are shown in Figure 4. The solid line plots the estimates of f_t; the dotted trajectory is

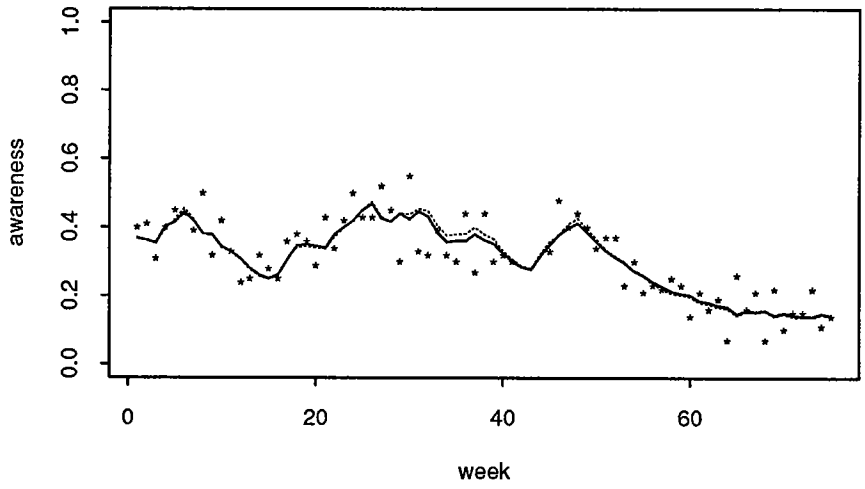

Figure 4: One step ahead forecasts.

the estimate of f_t using the dynamic generalized linear model. The actual responses are plotted as points. Figure 5 shows the estimated trajectory of the lower threshold parameter α_t. The solid line plots estimates of the posterior mean $E(\alpha_t|D_t)$, with the dashed lines representing one posterior standard deviation margins, i.e. $E(\alpha_t|D_t) + \sigma$ and $E(\alpha_t|D_t) - \sigma$, where $\sigma^2 = \text{var}(\alpha_t|D_t)$. (The posterior variance should be distinguished from the numerical variance, i.e. mean squared error, of the Monte Carlo estimator of $E(\alpha_t|D_t)$, which is σ^2/n, where $n = 100$ is the number of sample points.)

The dotted graph is the posterior estimate from the analysis as a DGLM.

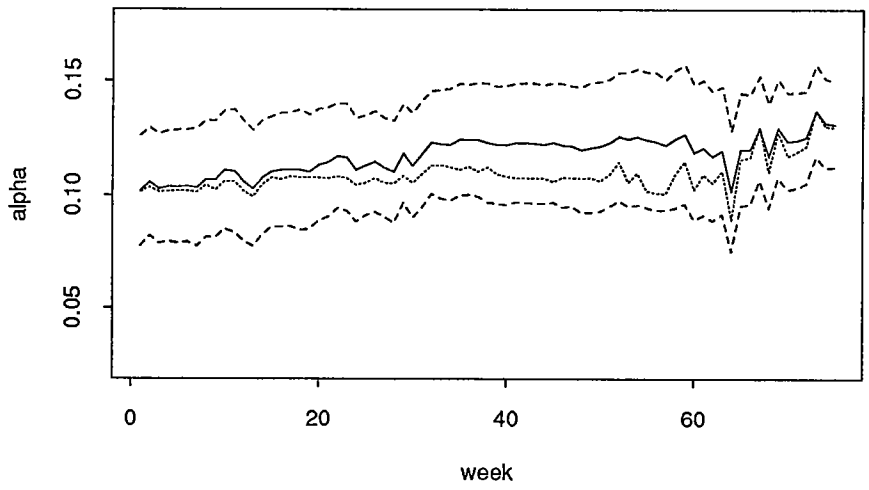

Figure 5: Lower threshold α_t.

The DGLM estimates are well within the one posterior standard deviation margins of the simulation estimate. Around week 30 the posterior on α_t shifts to a slightly higher level. The DGLM analysis adjusts to this new level only towards the end of the campaign, when the exponentially decreasing response values give increasingly sharper information on the lower threshold.

Figures 6 through 8 give the same plots for the parameters β_t, ρ_t and κ_t. Again, the estimates from the DGLM model and the simulation estimates stay always within one posterior standard deviation. The only noticeable

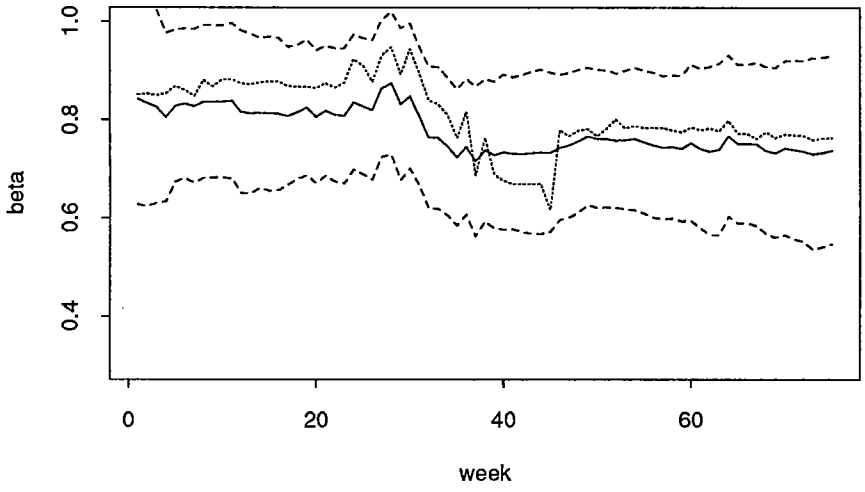

Figure 6: Upper threshold β_t.

difference occurs in the estimate for β_t around week 40. This downward jump in the trajectory of β_t corresponds to an overestimation of κ_t in the same time period. (The model takes a long time to readjust the posterior estimates, because the weeks 42, 43 and 44 have missing response values.)

The observed minor discrepancies between the DGLM analysis and the simulation analysis are due to approximations made in both algorithms.

In Figure 9 the beta density implicitly assumed for $p(\mu_{41}|D_{40})$ in the DGLM model is compared with a density estimate obtained from the sample points, which have been generated by the simulation algorithm. Figure

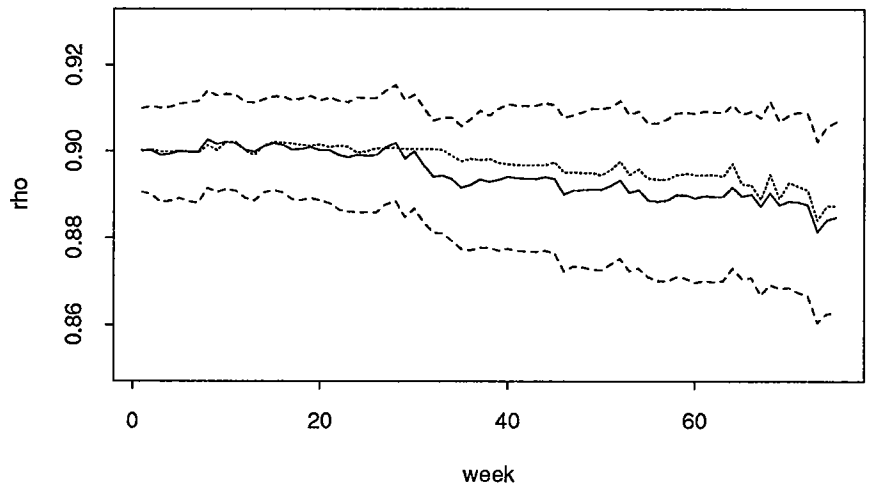

Figure 7: Decay parameter ρ_t.

10 plots a density estimate for the posterior $p(\kappa_{40}|D_{40})$ together with the corresponding normal density. In both cases the unimodal beta, respectively normal distribution does not adequately reproduce the multimodal density. This does not contradict the DGLM, since this algorithm is designed for problems where the prior distributions and the evolution noise are not specified with densities, but only specified by their first and second moments.

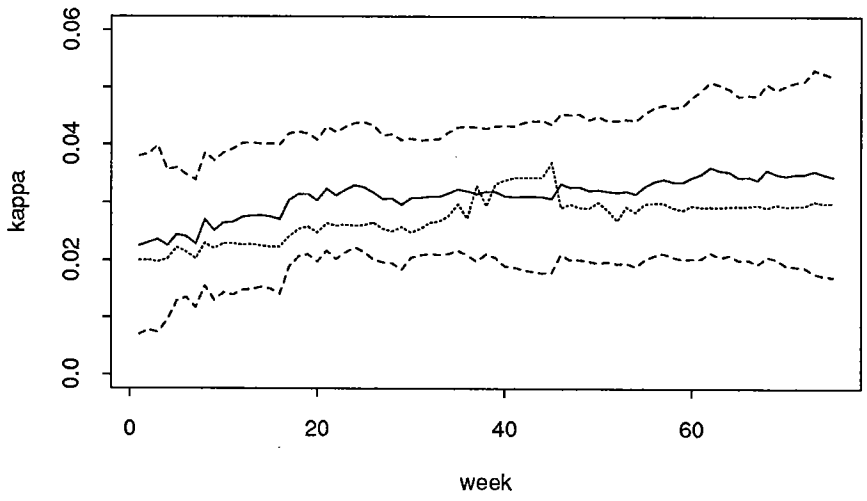

Figure 8: Penetration parameter ρ_t.

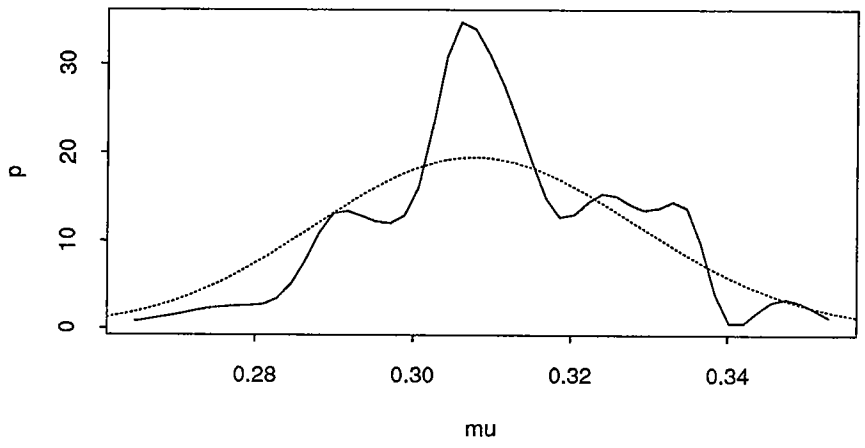

Figure 9: Density estimate for $p(\mu_{41}|D_{40})$.

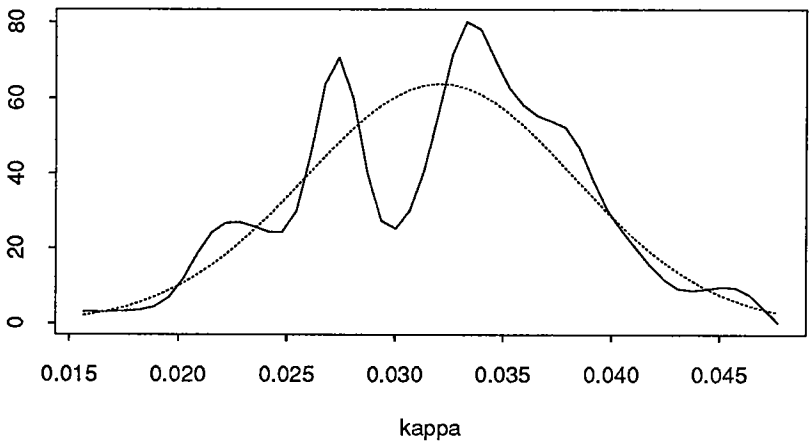

Figure 10: Density estimate for $p(\kappa_{40}|D_{40})$.

5 Conclusion

While the simulation of the evolution step is a straightforward implementation of the system equation, the observation equation, i.e. Bayes' theorem, lacks such a natural interpretation as a mapping from points of a prior sample to points of a posterior sample. In this paper a two step procedure was used, first expanding the prior sample into a sample from an envelope density and then contracting by accept/reject to a sample from the posterior. This approach works well in the context of the dynamic model where the prior is essentially the posterior from the previous step and is typically not significantly changed by updating with the likelihood. For the simulation algorithm this means that only a moderate expansion of the sample size is required. In the application in Section 4 it was only rarely necessary to take more than one iteration step in the construction of the envelope density.

Appendix

Stopping Rule for the Iterative Construction of the Envelope Density

The expected proportion of retained sample points in the accept/reject procedure is $\bar{w}/w^*$, where $\bar{w} = E_{I_M} w_M$ and $w^* = \sup(w_M)$. Let $\bar{X}|_A$ denote $E(X|A)$ and $\mu_m = I_m(L_m)$. Then $\bar{w}$ and w^* are given by:

$$\begin{aligned}
\bar{w_m} &= (1-\mu_m)\bar{w_m}|_{L_m^c} + \mu_m \bar{w_m}|_{L_m} \\
&= (1-\mu_m)\bar{w_{m-1}}|_{L_m^c} + \mu_m q_m \bar{w_{m-1}}|_{L_m} \\
&= \alpha_m \bar{w_{m-1}} \\
w_m^* &= \max(\bar{l}, q_m w_{m-1}^*).
\end{aligned}$$

Then the expected proportion of retained sample points after shrinking takes the form:

$$\begin{aligned}
\frac{\bar{w_m}}{w_m^*} &= \frac{\bar{l}}{w_m^*} \prod_{k=1}^{m} \alpha_k \\
&= \frac{1}{\max(\prod_1^m q_k l^*, \bar{l})} \bar{l} \prod_{k=1}^{m} \alpha_k.
\end{aligned}$$

Since the sample size after the m-th step is increased by a factor $1/\prod \alpha_k$

the following stopping rule will give an expected posterior sample size equal to the original prior sample size:

$$M = \min\{m : \prod_{1}^{m} q_k \leq \frac{\bar{l}}{l^*}\}.$$

References

[1] Geweke, J., "Bayesian inference in econometric models using Monte Carlo integration," *Econometrica*, 1989 forthcoming.

[2] Harrison, P.S. and Stevens, C.F., "Bayesian forecasting," *J.R. Stat. Soc. B* , vol. 38 , pp. 205-247 , 1976 .

[3] Ho, Y.C. and Lee, R.C.K., "A Bayesian approach to problems in stochastic estimation and control," *IEEE Trans. Autom. Control* , vol. 9 , pp. 333-339 , 1964 .

[4] Kalman, R.E., "An approach to linear filtering and prediction theory," *Trans ASME* , vol. 82D , pp. 35-45 , 1960 .

[5] Pole, A. and West, M., *"Efficient numerical integration in dynamic models,"* Technical Report, Department of Statistics - Univ of Warwick, May 1988.

[6] Pole, A., West, M., and Harrison, J., "Nonnormal and nonlinear dynamic Bayesian modelling," in *Bayesian Analysis of Time Series and Dynamic Models*, ed. J.C. Spall , Marcell Dekker, 1988.

[7] Rubinstein, R.Y., *Simulation and the Monte Carlo Method*, J. Wiley, 1981.

[8] Serfling, R., *Approximation Theorems of Mathematical Statistics*, J. Wiley, 1980.

[9] Sorenson, H.W. and Alspach, D.L., "Recursive Bayesian estimation using Gaussian sums," *Automatica* , vol. 7 , pp. 465-479 , 1971

.

[10] van Dijk, H.K., *Posterior analysis of econometric models using Monte Carlo integration*, Reproduktie Wondestein, Erasmus University Rotterdam, Rotterdam, Netherlands, 1984.

[11] West, M. and Harrison, J., *Bayesian Forecasting and Dynamic Models*, Springer Verlag, 1989.

[12] West, M., Harrison, P.J. and Migon, H.S., "Dynamic generalised linear models and Bayesian forecasting (with discussion)," *J. Amer. Statist. Assoc.*, vol. 80, pp. 73-97, 1985 .