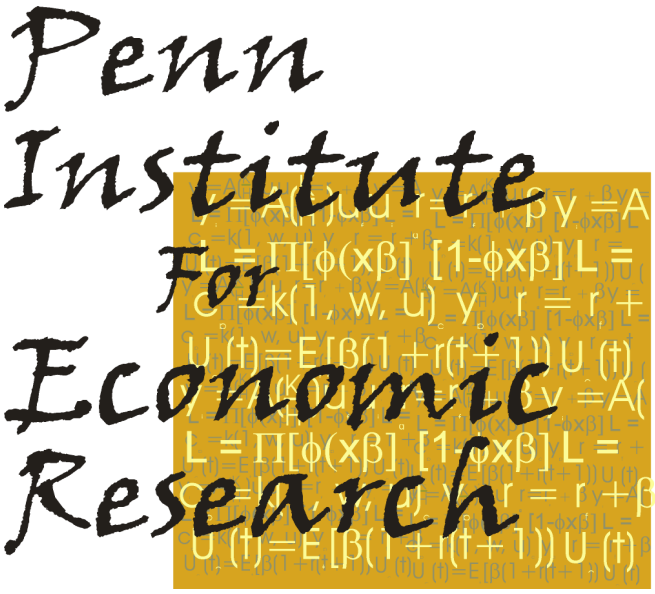

Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19103-6297

pier@ssc.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 97-035

"Nonlinear and Non-Gaussian State-Space Modeling with Monte-Carlo Simulations"

by

Robert S. Mariano and Hisashi Tanizaki

NONLINEAR AND NON-GAUSSIAN STATE-SPACE MODELING WITH MONTE-CARLO SIMULATIONS

HISASHI TANIZAKI
Kobe University
Kobe 657, JAPAN

ROBERTO S. MARIANO
University of Pennsylvania
Philadelphia, PA 19104, USA

December 1995

Abstract:

Two nonlinear and nonnormal filters using Monte-Carlo simulation techniques are proposed in this paper. In terms of programming and computational requirements, both proposed filters are more tractable than other nonlinear filters that have been developed earlier, such as those that use numerical integration or Monte-carlo integration with importance sampling or Gibbs sampling. The proposed filters are extended to prediction and smoothing algorithms. Monte-Carlo experiments are carried out also to assess the statistical merits of these proposed filters.

Key Words: Nonlinear State Space Models, Prediction, Filtering, Smoothing, Numerical Integration, Stochastic Simulation, Monte-Carlo Integration, Importance Sampling, Gibbs Sampling

1 Introduction

In one of Carl Christ's key research interests – macroeconomic forecasting – the linear filtering theory developed in Kalman (1960) and Kalman and Bucy (1961) provides a powerful tool. Various authors – e.g., Mariano and Schleicher (1972), Harvey (1989) and Mariano and Tanizaki (1995) – have approached forecasting with a macroeconometric model subject to measurement errors as a signal extraction problem solved through Kalman filtering.

This filtering theory has been extended in the literature in two directions. One extension comes by way of introducing nonlinearities in the measurement and transition equations. The second extension considers the likely cases of non-Gaussian error distributions in the state-space model. These extensions are non-trivial since the Kalman filter represented by

the conventional linear recursive algorithm is optimal only if the system is linear and error terms in the system are Gaussian.

One way of dealing with nonlinearities in the system is to linearize the measurement and transition equations and apply Kalman's linear recursive algorithm directly to this linearized system. Thus, the extended Kalman filter, the second-order nolinear filter, and so on, are developed. For example, see Wishner, Tabaczynski and Athans (1969), Sorenson and Alspach (1971), Alspach and Sorenson (1972), Gelb (1974), Anderson and Moore (1979), and Tanizaki and Mariano (1992).

However, these filters based on Taylor series expansions may produce large biases (see Tanizaki (1993), Tanizaki and Mariano (1994) and Mariano and Tanizaki (1995)). Furthermore, Meinhold and Singpurwalla (1989) pointed out that the Kalman filter solution is not robust against departures from normality assumptions.

For nonlinear state-space models with non-Gaussian errors, filtering methods are based on recursive algorithms for the evaluation of conditional probability densities. Kitagawa (1987) and Kramer and Sorenson (1988) evaluate each density by numerical integration (NI). A natural alternative is to utilize stochastic simulations that have been applied recently in estimation of and prediction with nonlinear econometric models – Brown and Mariano (1984, 1989), McFadden (1989), and Mariano and Brown (1983, 1989, 1993) – as well as with Bayesian applications, discussed, for example, in Geweke (1988, 1989a,b) and Shao (1989), For applications of these simulation techniques in nonlinear filtering, Tanizaki (1993), Tanizaki and Mariano (1994) and Mariano and Tanizaki (1995) utilized Monte-Carlo integration with importance sampling (MII).

Carlin, Polson and Stoffer (1992) proposed Monte-Carlo Integration with Gibbs sampling (MIG) for nonlinear filters. For the Gibbs sampler, see Gelfand, Hills, Racine-Poon and Smith (1990), Gelfand and Smith (1990), Geman and Geman (1984) and Zeger and Karim (1991). The algorithm porposed by Carlin, Polson and Stoffer is non-recursive and includes nuisance parameters in the underlying functions. Related procedures have also been suggested, in the treatment of nonlinear dynamic models, by Smith (1981), West and Harrison (1990), Pole and West (1990), and Jacquier, Polson and Rossi (1994).

The disadvantage of the density-based nonlinear and nonnormal procedures described above is the programming and computational burden involved in their implementation. In addition, each algorithm requires the calculation of conditional density functions based on the measurement and transition equations.

To rectify these complications, but still using a Monte-Carlo approach, we propose two nonlinear filters in this paper.

1. Monte-Carlo Approach Type 1 (MA1), which uses random draws directly from the appropriate conditional distributions,

2. Monte-Carlo Approach Type 2 (MA2), which utilizes rejection sampling to get random draws based on the transition equation and the conditional densities obtained from the measurement equation.

These new filters are less burdensome computationally and can be easily implemented. We further show in this paper that these filters can be extended directly to prediction and smoothing algorithms. Finally, Monte-Carlo experiments with linear and nonlinear state-space models are carried out to assess the statistical merits of these new nonlinear filters. Also note that density-based algorithms can be either recursive or not. Among the algorithms discussed in this section, NI, MII and MA1 are recursive while MIG and MA2 are non-recursive.

2 Nonlinear and Non-Gaussian State-Space Modeling: An Overview

We consider the following general nonlinear and non-Gaussian state-space model:

$$\text{Measurement Equation:} \quad y_t = h_t(\alpha_t, \epsilon_t), \tag{1}$$

$$\text{Transition Equation:} \quad \alpha_t = f_t(\alpha_{t-1}, \eta_t), \tag{2}$$

where $t = 1, \cdots, T$. The $p \times 1$ vector, y_t, is observable while the $k \times 1$ state vector, α_t, is unobserved. $h_t(\cdot)$ and $f_t(\cdot)$ are $p \times 1$ and $k \times 1$ vector functions, which are assumed to be known. ϵ_t and η_t are mutually independently distributed. Also, the distribution functions of the error terms ϵ_t and η_t are assumed known but not necessarily Gaussian.

Let $P(\cdot|\cdot)$ and Y_s be the conditional density function and the information set up to time s, i.e., $Y_s = \{y_s, y_{s-1}, \cdots, y_1\}$. Define $P_y(y_t|\alpha_t)$ as the density function based on the measurement equation (1) and $P_\alpha(\alpha_t|\alpha_{t-1})$ as the density function obtained from the transition equation (2).

2.1 Recursive Algorithms

In (1) and (2), the main problem is estimating the state vector α_t given observations on y summarized in the information set Y_s. It is a convention in the literature to call this

estimation problem smoothing when $s > t$, filtering when $s = t$, and prediction when $s < t$.

The density algorithm on L-step ahead prediction can be written as:

$$P(\alpha_{t+L}|Y_t) = \int P_\alpha(\alpha_{t+L}|\alpha_{t+L-1})P(\alpha_{t+L-1}|Y_t)\mathrm{d}\alpha_{t+L-1}, \tag{3}$$

where $P_\alpha(\alpha_{t+L}|\alpha_{t+L-1})$ can be derived from the transition equation (2). Given the density $P_\alpha(\alpha_{t+L}|\alpha_{t+L-1})$, the prediction density, i.e., $P(\alpha_{t+L}|Y_t)$ for $L = 1, 2, \cdots$, is recursively obtained.

For filtering, the recursive density algorithm can be formulated as follows (for example, see Kitagawa (1987) and Harvey (1989)):

$$P(\alpha_t|Y_{t-1}) = \int P_\alpha(\alpha_t|\alpha_{t-1})P(\alpha_{t-1}|Y_{t-1})\mathrm{d}\alpha_{t-1}, \tag{4}$$

$$P(\alpha_t|Y_t) = \frac{P_y(y_t|\alpha_t)P(\alpha_t|Y_{t-1})}{\int P_y(y_t|\alpha_t)P(\alpha_t|Y_{t-1})\mathrm{d}\alpha_t}, \tag{5}$$

where the densities $P_y(y_t|\alpha_t)$ and $P_\alpha(\alpha_t|\alpha_{t-1})$ are computed from the measurement equation (1) and the transition equation (2), respectively. Based on the two densities, equation (4) yields $P(\alpha_t|Y_{t-1})$ given $P(\alpha_{t-1}|Y_{t-1})$ and equation (5) yields $P(\alpha_t|Y_t)$ given $P(\alpha_t|Y_{t-1})$. Thus, recursively predicting and updating from one-point to the next, the filtering densities $P(\alpha_t|Y_t)$ are obtained sequentially from $t = 1$ to $t = T$.

Also, the smoothing algorithm is:

$$P(\alpha_t|Y_T) = P(\alpha_t|Y_t)\int \frac{P(\alpha_{t+1}|Y_T)P_\alpha(\alpha_{t+1}|\alpha_t)}{P(\alpha_{t+1}|Y_t)}\mathrm{d}\alpha_{t+1}, \tag{6}$$

where $P(\alpha_t|Y_t)$ and $P(\alpha_{t+1}|Y_t)$ are computed from the filtering algorithm in (4) and (5), while $P_\alpha(\alpha_{t+1}|\alpha_t)$ is obtained from the transition equation (2). The smoothing algorithm is represented by the backward recursion (i.e., $t = T, T-1, \cdots, 1$) shown in equation (6).

Given these smoothing, filtering or prediction densities $P(\alpha_r|Y_s)$ we would be interested in evaluating conditional expectations of certain functions of α_r:

$$g_{r|s} \equiv E(g(\alpha_r)|Y_s) = \int g(\alpha_r)P(\alpha_r|Y_s)\mathrm{d}\alpha_r, \tag{7}$$

Typically we would be interested in the conditional mean and the conditional variance of α_r given Y_s: $\alpha_{r|s} = E(\alpha_r|Y_s)$ and $\sum_{r|s} = E((\alpha_r - \alpha_{r|s})(\alpha_r - \alpha_{r|s})'|Y_s)$.

Also, when there are unknown parameters in the model (see (1) and (2)), we can write out the likelihood of the sample as

$$P(Y_T) = \prod_{t=1}^{T} P(y_t|Y_{t-1})$$
$$= \prod_{t=1}^{T} \left(\int P_y(y_t|\alpha_t) P(\alpha_t|Y_{t-1}) \mathrm{d}\alpha_t \right). \quad (8)$$

Note that $P(y_t|Y_{t-1})$ is given by the denominator of equation (5).

The integals in (3) - (8) are evaluated through numerical integration and Monte-Carlo simulations.

2.1.1 Numerical Integration (NI)

Kitagawa (1987) and Kramer and Sorenson (1988) proposed a nonlinear filter using NI. Consider the case of a scalar state vector. NI requires the nodes which are denoted by $\alpha_{i,t}$, $i = 0, \cdots, n$. These nodes, specified by the researcher, are sorted by size with respect to i, i.e., $\alpha_{0,t}$ is the smallest value and $\alpha_{n,t}$ the largest one. There are various conventional methods for numerical approximations to integrals: use of rectangular areas, trapezoids, Simpson's formula and so on. Here, for simplicity of discussion, we take the NI method using sums of rectangles.

By NI, equation (3) is evaluated as follows for NI prediction (NIP):

$$P(\alpha_{i,t+L}|Y_t) \approx \sum_{j=1}^{n} P_\alpha(\alpha_{i,t+L}|\alpha_{j,t+L-1}) P(\alpha_{j,t+L-1}|Y_t)(\alpha_{j,t+L-1} - \alpha_{j-1,t+L-1}).$$

For the NI filter (NIF) algorithm, we approximate (4) and (5) as follows:

$$P(\alpha_{i,t}|Y_{t-1}) \approx \sum_{j=1}^{n} P_\alpha(\alpha_{i,t}|\alpha_{j,t-1}) P(\alpha_{j,t-1}|Y_{t-1})(\alpha_{j,t-1} - \alpha_{j-1,t-1}),$$

$$P(\alpha_{i,t}|Y_t) \approx \frac{P_y(y_t|\alpha_{i,t}) P(\alpha_{i,t}|Y_{t-1})}{\sum_{j=1}^{n} P_y(y_t|\alpha_{j,t}) P(\alpha_{j,t}|Y_{t-1})(\alpha_{j,t} - \alpha_{j-1,t})},$$

where $i = 1, \cdots, n$ and $t = 1, \cdots, T$. For the NI smoothing (NIS) algorithm, equation (6) is rewritten as:

$$P(\alpha_{i,t}|Y_T) \approx P(\alpha_{i,t}|Y_t) \sum_{j=1}^{n} \frac{P(\alpha_{j,t+1}|Y_T) P_\alpha(\alpha_{j,t+1}|\alpha_{i,t})}{P(\alpha_{j,t+1}|Y_t)} (\alpha_{j,t+1} - \alpha_{j-1,t+1}).$$

The NI estimate of $g_{r|s}$, (as defined in (7)), is then written as

$$\tilde{g}_{r|s} = \sum_{i=1}^{n} g(\alpha_{i,r}) P(\alpha_{i,r}|Y_s)(\alpha_{i,r} - \alpha_{i-1,r}).$$

According to the density approximation based on NI, each density is specified by the number of segments (i.e., n), location of nodes (i.e., $\alpha_{i,t}$), and the value at each node (i.e., $P_\alpha(\alpha_{i,t}|\alpha_{i,t-1})$ and $P_y(y_t|\alpha_{i,t})$).

Finally, for purposes of estimating any unknown parameters in the model, numerical maximum likelihood estimation can be implemented with the likelihood function (8) approximated as:

$$P(Y_T) \approx \prod_{t=1}^{T} \left(\sum_{j=1}^{n} P_y(y_t|\alpha_{j,t}) P(\alpha_{j,t}|Y_{t-1})(\alpha_{j,t} - \alpha_{j-1,t}) \right),$$

For NIP, NIF and NIS, computational errors drastically grow as time t increases. To avoid accumulation of the errors, we need to impose

$$\sum_{i=1}^{n} P(\alpha_{i,r}|Y_s)(\alpha_{i,r} - \alpha_{i-1,r}) = 1,$$

as a consequence of the fact that

$$\int P(\alpha_r|Y_s) \mathrm{d}\alpha_r = 1.$$

The NI procedures can evaluate expectations quite precisely. However, when the state vector α_t is extended to multidimensional cases, the computational burden increases more than proportionately and all these numerical integration procedures (NIP, NIF and NIS) can quikly become intractable as the dimension of α_t increases.

2.1.2 Monte-Carlo Integration with Importance Sampling (MII)

Tanizaki (1993), Tanizaki and Mariano (1994) and Mariano and Tanizaki (1995) developed a nonlinear filter with MII, where a recursive algorithm for density functions is developed through a sequential calculation of weight functions. Define the weight function $\omega_{r|s} = P(\alpha_r|Y_s)/P_I(\alpha_r)$, where the importance density function $P_I(\alpha_t)$ is specified by the researcher. Further, define $\omega_{i,r|s} = P(\alpha_{i,r}|Y_s)/P_I(\alpha_{i,r})$ as the weight function evaluated at $\alpha_{i,r}$, where $\alpha_{i,r}$, $i = 1, \cdots, n$, are random draws from the importance density $P_I(.)$.

By MII, equation (3) is given by:

$$\omega_{i,t+L|t} \approx \frac{1}{n} \sum_{j=1}^{n} \frac{P_\alpha(\alpha_{i,t+L}|\alpha_{j,t+L-1})}{P_I(\alpha_{i,t+L})} \omega_{j,t+L-1|t}, \tag{9}$$

the formula for MII prediction (MIIP). This equation comes from rewriting the prediction density as

$$\omega_{t+L|t} = \int \frac{P_\alpha(\alpha_{t+L}|\alpha_{t+L-1})}{P_I(\alpha_{t+L})}\omega_{t+L-1|t}P_I(\alpha_{t+L-1})\mathrm{d}\alpha_{t+L-1}.$$

Generating n random numbers of α_{t+L-1} from the importance density $P_I(\alpha_{t+L-1})$ and evaluating the weight function $\omega_{t+L|t}$ at the random numbers $\alpha_{i,t+L}$ generated from $P_I(\alpha_{t+L})$, we obtain equation (9).

An algorithm of MII filter (MIIF) is represented by the following two equations:

$$\omega_{i,t|t-1} \approx \frac{1}{n}\sum_{j=1}^{n}\frac{P_\alpha(\alpha_{i,t}|\alpha_{j,t-1})}{P_I(\alpha_{i,t})}\omega_{j,t-1|t-1},$$

$$\omega_{i,t|t} \approx \frac{P_y(y_t|\alpha_{i,t})\omega_{i,t|t-1}}{\frac{1}{n}\sum_{j=1}^{n}P_y(y_t|\alpha_{j,t})\omega_{j,t|t-1}},$$

where $i = 1, \cdots, n$ and $t = 1, \cdots, T$. Thus, a recursive algorithm for the weight functions, rather than for densities, is derived.

To obtain MII smoothing (MIIS), we rewrite equation (6) as:

$$\omega_{i,t|T} \approx \omega_{i,t|t}\frac{1}{n}\sum_{j=1}^{n}\frac{\omega_{j,t+1|T}P_\alpha(\alpha_{j,t+1}|\alpha_{i,t})}{\omega_{j,t+1|t}P_I(\alpha_{j,t+1})},$$

Similarly, the expectation of $g(\alpha_t)$ is evaluated as:

$$\widehat{g}_{r|s} = \frac{1}{n}\sum_{i=1}^{n}g(\alpha_{i,r})\omega_{i,r|s},$$

where $\widehat{g}_{r|s}$ denotes the MII estimate of $g_{r|s}$ based on MII. Thus, MIIP, MIIF and MIIS depend on the number of random numbers (i.e., n) and choice of the importance density. Further more, the likelihood function (8) is rewritten as:

$$P(Y_T) \approx \prod_{t=1}^{T}\left(\frac{1}{n}\sum_{j=1}^{n}P_y(y_t|\alpha_{j,t})\omega_{j,t|t-1}\right).$$

For MIIP, MIIF and MIIS, we impose the following condition to avoid accumulation of simulation errors[1]:

$$\frac{1}{n}\sum_{i=1}^{n}\omega_{i,r|s} = 1,$$

[1]Unless $(1/n)\sum\omega_{i,t|t-1} = 1$ is satisfied, $\omega_{i,t|t}$ becomes unrealistic because it depends on $\omega_{i,t|t-1}$. Since $\omega_{i,t+1|t}$ depends on $\omega_{i,t|t}$, $\omega_{i,t+1|t}$ is also biased, i.e., $(1/n)\sum\omega_{i,t+1|t} = 1$ does not hold. Accordingly, we need the restriction $(1/n)\sum\omega_{i,r|s} = 1$ for $(r,s) = (t,t-1),(t,t),(t,T)$ and all t in order to avoid accumulation of simulation errors..

This comes from $\int P(\alpha_r|Y_s)d\alpha_r = 1$. See Koop (1994) for a survey of integration methods.

The MII procedure can be easily implemented even when the state vector is multidimensional - an attractive feature over the NI approach. It is easily shown that $\widehat{g}_{r|s}$ is consistent, but with a convergence rate of $\sqrt{n}$.[2]

2.2 Monte-Carlo Integration with Gibbs Sampling (MIG)

Carlin, Polson and Stoffer (1992) proposed a solution to multivariate state-space modeling, where they allowed for the possibilities of non-Gaussian errors and nonlinear functions in the state-space model. Their crucial assumptions are that distribution functions of ϵ_t and η_t depend on nuisance parameters γ_t and λ_t, which are assumed to have prior densities, $P_\gamma(\gamma_t)$ and $P_\lambda(\lambda_t)$, in a Bayesian framework. From equation (1) and the density function of ϵ_t, the density of y_t given α_t and γ_t is represented by $P_y(y_t|\alpha_t, \gamma_t)$ and, from equation (2) and η_t, the density of α_t given α_{t-1} and λ_t is $P_\alpha(\alpha_t|\alpha_{t-1}, \lambda_t)$. Note that the initial density of α_0 given λ_0 is assumed to be $P_\alpha(\alpha_0|\lambda_0)$.

Under the assumption that $P_y(y_t|\alpha_t, \gamma_t)$, $P_\gamma(\gamma_t)$, $P_\alpha(\alpha_t|\alpha_{t-1}, \lambda_t)$ and $P_\lambda(\lambda_t)$ are known, the following three posterior density functions for $t = 1, \cdots, T$ (i.e., $P_{\gamma_*}(\gamma_t|\cdot)$, $P_{\lambda_*}(\lambda_t|\cdot)$ and $P_{\alpha_*}(\alpha_t|\cdot)$) are obtained using Bayes theorem. $P_{\gamma_*}(\gamma_t|y_t, \alpha_t)$ is computed from $P_y(y_t|\alpha_t, \gamma_t)$ and $P_\gamma(\gamma_t)$, $P_{\lambda_*}(\lambda_t|\alpha_t, \alpha_{t-1})$ is obtained from $P_\alpha(\alpha_t|\alpha_{t-1}, \lambda_t)$ and $P_\lambda(\lambda_t)$, and $P_{\alpha_*}(\alpha_t|y_t, \alpha_{t-1}, \alpha_{t+1}, \gamma_t, \lambda$ is derived from $P_y(y_t|\alpha_t, \gamma_t)$, $P_\alpha(\alpha_t|\alpha_{t-1}, \lambda_t)$ and $P_\alpha(\alpha_{t+1}|\alpha_t, \lambda_{t+1})$. Given the three posterior densities, Gibbs sampling is performed to obtain random draws from the joint density function of $(Y_T, A_T, \Gamma_T, \Lambda_T)$, where $A_T = \{\alpha_T, \alpha_{T-1}, \cdots, \alpha_0\}$, $\Gamma_T = \{\gamma_T, \gamma_{T-1}, \cdots, \gamma_1\}$ and $\Lambda_T = \{\lambda_T, \lambda_{T-1}, \cdots, \lambda_0\}$.

Let us define $\overline{g}_{r|s}$ as the MIG estimate of $g_{r|s}$. Since $\alpha_{i,r}$ is the random draw from $P(\alpha_r|Y_s)$, this approximation is simply

$$\overline{g}_{r|s} = \frac{1}{n}\sum_{i=1}^{n} g(\alpha_{i,r}).$$

[2] As shown in Geweke (1989a,1989b), MII gives us a consistent estimate as the number of random draws increases. Note that $g(\alpha_{i,r})\omega_{i,r|s}$ in $\widehat{g}_{r|s}$ is the random number generated from $P_I(\alpha_r)$. Define

$$E\big(g(\alpha_{i,r})\omega_{i,r|s}\big) = E\big(g(\alpha_r)|Y_s\big) \equiv g_{r|s},$$

$$E\big(g(\alpha_{i,r})g(\alpha_{i,r})'\omega^2_{i,r|s}\big) = E\big(g(\alpha_r)g(\alpha_r)'\omega_{r|s}|Y_s\big) \equiv \Sigma_{r|s}.$$

The standard central limit theorem then gives

$$\sqrt{n}(\widehat{g}_{r|s} - g_{r|s}) \longrightarrow N(0, \Sigma_{r|s}).$$

Thus, for the MIG estimate of $g_{r|s}$, the procedure developed by Carlin, Polson and Stoffer (1992) utilizes random draws only. In comparing estimates of $g_{r|s} \equiv E(g(\alpha_r)|Y_s)$, i.e., equations $\widetilde{g}_{r|s}$, $\widehat{g}_{r|s}$ and $\overline{g}_{r|s}$, we see that we need the density functions or the weight functions for $\widetilde{g}_{r|s}$ and $\widehat{g}_{r|s}$ but not for $\overline{g}_{r|s}$. For $\overline{g}_{r|s}$, the important point is to generate random draws from the three posterior densities.

The likelihood function (8) is evaluated as

$$\begin{aligned} P(Y_T) &= \iiint P(Y_T, A_T, \Gamma_T, \Lambda_T) \mathrm{d}A_T \mathrm{d}\Gamma_T \mathrm{d}\Lambda_T \\ &= \iiint \left(P_\alpha(\alpha_0|\lambda_0) \prod_{t=1}^{T} P_y(y_t|\alpha_t, \gamma_t) P_\alpha(\alpha_t|\alpha_{t-1}, \lambda_t) P_\gamma(\gamma_t) P_\lambda(\lambda_t) \mathrm{d}A_T \mathrm{d}\Gamma_T \mathrm{d}\Lambda_T \right) \\ &\approx \prod_{t=1}^{T} \left(\frac{1}{n} \sum_{i=1}^{n} P_y(y_t|\alpha_{i,t}, \gamma_{i,t}) \right), \end{aligned}$$

where the joint density of $\{Y_T, A_T, \Gamma_T, \Lambda_T\}$ is given by $P(Y_T, A_T, \Gamma_T, \Lambda_T)$.

The MIG approach requires random draws only, using the Gibbs sampler, but the Gibbs sampling procedure has *ad hoc* assumption such as choice of the prior distribution for nuisance parameters, i.e., γ_t and λ_t. The MIG procedure has the computational disadvantage of requiring a great amount of computation and data storage: $\alpha_{i,t}$, $\gamma_{i,t}$ and $\lambda_{i,t}$ for $t = 1, \cdots, T$ and $i = 1, \cdots, n$ have to be stored to obtain the unconditional random draws. (see Carlin and Polson (1991) and Carlin, Polson and Stoffer (1992)).

3 New Nonlinear Filters Based on Monte-Carlo Simulations

For the NI and the MII approaches, choice of $\alpha_{i,t}$ is one of the critical problems, because precision of the estimates depends on location of the nodes for NI and choice of the importance density for MII. For the MIG procedure, we need to specify the prior densities of the nuisance parameters. The nonlinear procedures proposed in this section do not require such assumptions

3.1 Monte-Carlo Approach: Type 1 (MA1)

In this section, we use random draws from appropriate conditional distributions to obtain prediction, filtering and smoothing estimates. Let $\alpha_{i,r|s}$ be the i-th random draw from the density function of α_r given Y_s.

MA1 Prediction (MA1P): The MA1 prediction (MA1P) estimate is obtained simply as follows. Suppose that $\alpha_{i,t+L-1|t}$, $i = 1, \cdots, n$, are available. The L-step ahead prediction algorithm (3) is represented as

$$P(\alpha_{t+L}|Y_t) \approx \frac{1}{n}\sum_{i=1}^{n} P_\alpha(\alpha_{t+L}|\alpha_{i,t+L-1|t}),$$

where $P(\alpha_{t+L}|Y_t)$ is approximated as a mixture distribution. We can generate random draws of α_{t+L} given Y_t as follows.

(i) Pick $\alpha_{i,t+L-1|t}$ randomly (i.e., pick i with probability $1/n$).

(ii) Given $\alpha_{i,t+L-1|t}$, generate random draws of $\alpha_{t+L|t}$ (i.e., $\alpha_{j,t+L|t}$) from the transition equation $\alpha_{j,t+L|t} = f_{t+L}(\alpha_{i,t+L-1|t}, \eta_{j,t+L})$ after generating a random number for η_{t+L}, i.e., $\eta_{j,t+L}$ for $j = 1, \cdots, n$.

Thus, given $\alpha_{i,t|t}$, $\alpha_{i,t+L|t}$ is recursively obtained for $L = 1, 2, \cdots$. $\alpha_{i,t|t}$ is generated in the following filtering algorithm.

MA1 Filtering (MA1F). The MA1 filtering (MA1F) estimate is derived as follows. Suppose the random draws $\alpha_{i,t-1|t-1}$ are available. Given $\alpha_{i,t-1|t-1}$ for $i = 1, \cdots, n$, we consider generating $\alpha_{j,t|t}$ for $j = 1, \cdots, n$, which implies the recursive algorithm based on random draws. Thus, we consider generating random numbers of α_t directly from the filtering density $P(\alpha_t|Y_t)$. The filtering density $P(\alpha_t|Y_t)$ given by equation (5) is represented as:

$$\begin{aligned} P(\alpha_t|Y_t) &= \frac{P_y(y_t|\alpha_t)P(\alpha_t|Y_{t-1})}{\int P_y(y_t|\alpha_t)P_\alpha(\alpha_t|Y_{t-1})\mathrm{d}\alpha_t} \\ &\propto \int P_y(y_t|\alpha_t)P_\alpha(\alpha_t|\alpha_{t-1})P(\alpha_{t-1}|Y_{t-1})\mathrm{d}\alpha_{t-1} \\ &\approx \frac{1}{n}\sum_{i=1}^{n} P_y(y_t|\alpha_t)P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}) \\ &\propto \frac{1}{n}\sum_{i=1}^{n} w_1(\alpha_t; y_t)P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}), \end{aligned} \tag{10}$$

where $w_1(\alpha_t; y_t)$ satisfies:

$$\begin{aligned} &w_1(\alpha_t; y_t) \propto P_y(y_t|\alpha_t), \\ &0 \leq w_1(\alpha_t; y_t) \leq 1. \end{aligned}$$

Thus, the filtering density $P(\alpha_t|Y_t)$ is approximately proportional to

$$\frac{1}{n}\sum_{i=1}^{n} w_1(\alpha_t; y_t)P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}).$$

When we cannot obtain the explicit functional form of the density $P(\alpha_t|Y_t)$, the following random number generation is helpful. We may use rejection sampling (described further in Appendix 1) to obtain a random observation from $P(\alpha_t|Y_t)$. The distribution which we want to sample is dominated by the $P_\alpha(\alpha_t|\alpha_{j,t-1|t-1})$ density. Therefore, for $t = 1, \cdots, T$, we choose $\alpha_{j,t-1|t-1}$ with probability $1/n$ (i.e., we choose j), sample the $P_\alpha(\alpha_t|\alpha_{j,t-1|t-1})$ distribution and accept it with probability $w_1(\alpha_t; y_t)$.

In general, summarizing the above, the MA1F estimate is given by the following procedure.

(i) Pick j with probalility $1/n$, i.e., choose $\alpha_{j,t-1|t-1}$ for $j = 1, \cdots, n$.

(ii) Given j, generate a random draw of α_t (i.e., $\alpha_{i,t}$) from the transition equation $\alpha_{i,t} = f_t(\alpha_{j,t-1|t-1}, \eta_{i,t})$, where $\eta_{i,t}$ is the i-th random number of η_t.

(iii) (a) If $\alpha_{i,t}$ is accepted with probability $w_1(\alpha_{i,t}; y_t)$, it is taken as a random draw from the filtering density at time t, i.e., $\alpha_{i,t|t}$.

(b) If $\alpha_{i,t}$ is not accepted with probability $w_1(\alpha_{i,t}, y_t)$, go to (ii) again.

(iv) Repeat (i) – (iii) for $i = 1, \cdots, n$.

In this procedure, if the acceptance probability is small, the steps (ii) and (iii-b) are repeated indefinitely. In order to avoid this situation, if the acceptance probability is too small, we may go to (i), pick up another j and repeat the above procedure.

Thus, the random numbers generated from equation (10) are obtained.

MA1 Smoothing (MA1S): Finally, the MA1 smoothing (MA1S) estimate is represented as follows. Let $\alpha_{i,t-1|t-1}$ and $\alpha_{j,t+1|T}$ be the random draws of α_{t-1} given Y_{t-1} and α_{t+1} given Y_T. Suppose that $\alpha_{i,t|t}$ for $t = 1, \cdots, T$ and $i = 1, \cdots, n$ are available, which are obtained from the MA1F procedure shown above.

To obtain the MA1S estimate, note that each component in equation (6) is transformed as follows:

$$\begin{aligned} P(\alpha_t|Y_t) &= \frac{1}{P(y_t|Y_{t-1})} \int P_y(y_t|\alpha_t) P_\alpha(\alpha_t|\alpha_{t-1}) P(\alpha_{t-1}|Y_{t-1}) \mathrm{d}\alpha_{t-1} \\ &\approx \frac{1}{P(y_t|Y_{t-1})} \frac{1}{n} \sum_{i=1}^{n} P_y(y_t|\alpha_t) P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}), \end{aligned}$$

$$\int \frac{P(\alpha_{t+1}|Y_T) P_\alpha(\alpha_{t+1}|\alpha_t)}{P(\alpha_{t+1}|Y_t)} \mathrm{d}\alpha_{t+1} \approx \frac{1}{n} \sum_{j=1}^{n} \frac{P_\alpha(\alpha_{j,t+1|T}|\alpha_t)}{P(\alpha_{j,t+1|T}|Y_t)}.$$

$P(\alpha_{t+1}|Y_t)$ in the denominator of equation (6) is rewritten as:

$$\begin{aligned} P(\alpha_{t+1}|Y_t) &= \int P_\alpha(\alpha_{t+1}|\alpha_t)P(\alpha_t|Y_t)\mathrm{d}\alpha_t \\ &\approx \frac{1}{n}\sum_{m=1}^{n} P_\alpha(\alpha_{t+1}|\alpha_{m,t|t}). \end{aligned}$$

Accordingly, the smoothing density (6) is approximated as:

$$\begin{aligned} P(\alpha_t|Y_T) &\approx \frac{1}{P(y_t|Y_{t-1})}\left(\frac{1}{n}\sum_{i=1}^{n} P_y(y_t|\alpha_t)P_\alpha(\alpha_t|\alpha_{i,t-1|t-1})\right)\left(\frac{1}{n}\sum_{j=1}^{n}\frac{P_\alpha(\alpha_{j,t+1|T}|\alpha_t)}{P(\alpha_{j,t+1|T}|Y_t)}\right) \\ &\propto \left(\frac{1}{n}\sum_{i=1}^{n} P_y(y_t|\alpha_t)P_\alpha(\alpha_t|\alpha_{i,t-1|t-1})\right)\left(\frac{1}{n}\sum_{j=1}^{n}\frac{P_\alpha(\alpha_{j,t+1|T}|\alpha_t)}{P(\alpha_{j,t+1|T}|Y_t)}\right) \\ &\propto \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{q_{j,t}}P_y(y_t|\alpha_t)P_\alpha(\alpha_{j,t+1|T}|\alpha_t)P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}) \\ &\propto \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{q_{j,t}}w_1(\alpha_t;y_t)w_2(\alpha_t;\alpha_{j,t+1|T})P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}), \end{aligned} \tag{11}$$

for $t = T-1, T-2, \cdots, 1$, where $q_{j,t}$ satisfies the following two conditions:

$$\begin{aligned} \text{(i)}\quad q_{j,t} &\propto \frac{1}{n}\sum_{m=1}^{n} P_\alpha(\alpha_{j,t+1|T}|\alpha_{m,t|t}) \\ &\approx \int P_\alpha(\alpha_{j,t+1|T}|\alpha_t)P(\alpha_t|Y_t)\mathrm{d}\alpha_t \\ &= P(\alpha_{j,t+1|T}|Y_t), \end{aligned}$$

$$\text{(ii)}\quad \sum_{j=1}^{n}\frac{1}{q_{j,t}} = 1,$$

and $w_1(\alpha_t;y_t)$ and $w_2(\alpha_t;\alpha_{j,t+1|T})$ satisfy:

$$\begin{aligned} &w_1(\alpha_t;y_t) \propto P_y(y_t|\alpha_t), \\ &w_2(\alpha_t;\alpha_{j,t+1|T}) \propto P_\alpha(\alpha_{j,t+1|T}|\alpha_t), \\ &0 \le w_1(\alpha_t;y_t)w_2(\alpha_t;\alpha_{j,t+1|T}) \le 1. \end{aligned}$$

Thus, given $\alpha_{i,t-1|t-1}$ and $\alpha_{j,t+1|T}$, the smoothing density $P(\alpha_t|Y_T)$ is proportional to the righthand side of equation (11).

When we cannot explicitly obtain the functional form of the posterior (or smoothing) density $P(\alpha_t|Y_T)$, we may utilize rejection sampling to obtain a random observation from the smoothing density $P(\alpha_t|Y_T)$ for $t = T-1, T-2, \cdots, 1$. This is a backward recursive algorithm - that is, for $t = T-1, T-2, \cdots, 1$, we choose $\alpha_{i,t-1|t-1}$ with probability $1/n$ and

$\alpha_{j,t+1|T}$ with probability $1/q_{j,t}$ (i.e, we choose i and j randomly), sample the $P_\alpha(\alpha_t|\alpha_{i,t-1|t-1})$ random variable and accept it with probability $w_1(\alpha_t;y_t)w_2(\alpha_t;\alpha_{j,t+1|T})$, where $\alpha_{i,T|T}$ is the random draw from the filtering density. Thus, the random draw of α_t given information Y_T can be generated. Thus, based on $w_1(\alpha_t;y_t)w_2(\alpha_t;\alpha_{j,t+1|T})P_\alpha(\alpha_t|\alpha_{i,t-1|t-1})$, we can generate the random draws for α_t from $P(\alpha_t|Y_T)$ for $t=T-1,T-2,\cdots,1$.

In equation (11), we require an extra computational burden to compute $q_{j,t}$. Therefore, we approximately use $q_{j,t}=n$ for all j. The posterior density $P(\alpha_t|Y_T)$ is approximated by:

$$P(\alpha_t|Y_T) \propto \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_1(\alpha_t;y_t)w_2(\alpha_t;\alpha_{j,t+1|T})P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}).$$

Thus, we generate the random numbers of α_t, given $\alpha_{i,t-1|t-1}$ and $\alpha_{j,t+1|T}$, where i and j are chosen with probability $1/n$ and probability $1/n$, respectively.

Each density function used for the computational procedure is summarized as follows.

(i) For the prediction density,

$$P(\alpha_{t+L}|Y_t) \approx \frac{1}{n}\sum_{i=1}^{n} P_\alpha(\alpha_{t+L}|\alpha_{i,t+L-1|t}), L=1,2,\cdots,$$

(ii) For the filtering density,

$$P(\alpha_t|Y_t) \propto \frac{1}{n}\sum_{i=1}^{n} w_1(\alpha_t;y_t)P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}), t=1,2,\cdots,T,$$

(iii) For the smoothing density,

$$P(\alpha_t|Y_T) \propto \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{1}{q_{j,t}} w_1(\alpha_t;y_t)w_2(\alpha_t;\alpha_{j,t+1|T})P(\alpha_t|\alpha_{i,t-1|t-1}), t=T,T-1,\cdots,1,$$

where $\alpha_{i,t-1|t-1}$ denote the random draws generated from $P(\alpha_{t-1}|Y_{t-1})$ and $\alpha_{j,t+1|T}$ represent the random draws from $P(\alpha_{t+1}|Y_T)$. The distribution which we want to sample is dominated by the $P_\alpha(\alpha_t|\alpha_{i,t-1|t-1})$ density. Therefore, we may use rejection sampling to obtain a random observation from $P(\alpha_t|Y_t)$ or $P(\alpha_t|Y_T)$. That is, we choose i or (i,j) randomly, sample a $P_\alpha(\alpha_t|\alpha_{i,t-1|t-1})$ random variable and accept it with probability $w_1(\alpha_t;y_t)$ or $w_1(\alpha_t;y_t)w_2(\alpha_t;\alpha_{j,t+1|T})$. Thus, we recursively generate the random numbers for prediction, filtering and smoothing.

Using the random draws $\alpha_{i,r|s}$ for $(r,s)=(t+L,t),(t,t),(t,T)$ that are generated, we can then calculate $g^*_{r|s}$, the MA1 estimate of $g_{r|s}$.

$$g^*_{r|s} = \frac{1}{n}\sum_{i=1}^{n} g(\alpha_{i,r|s}).$$

We call this the Monte-Carlo predictor (MA1P) when $(r,s)=(t+L,t)$, the Monte-Carlo filter (MA1F) when $(r,s)=(t,t)$ and the Monte-Carlo smoother (MA1S) when $(r,s)=(t,T)$.

The likelihood function is given by

$$\begin{aligned} P(Y_T) &= \prod_{t=1}^{T}\left(\int P_y(y_t|\alpha_t)P(\alpha_t|Y_{t-1})\mathrm{d}\alpha_t\right) \\ &\approx \prod_{t=1}^{T}\left(\frac{1}{n}\sum_{i=1}^{n}P_y(y_t|\alpha_{i,t|t-1})\right). \end{aligned}$$

The features of the above procedure are:

(i) We do not need the exact functional form of $P_y(y_t|\alpha_t)$, $P_\alpha(\alpha_t|\alpha_{t-1})$ and $P(\alpha_r|Y_s)$ for $(r,s)=(t+L,t),(t,t),(t,T)$.

(ii) The random draws of α_t have to be directly generated from the posterior density $P(\alpha_r|Y_s)$.

(iii) Programming is very easy, compared with the other nonlinear filters (NI, MII, and MIG) summarized in the preceding section.

The following examples illustrate the implementation of the above procedure. In all four examples, we assume that the functional specification of the measurement and transition equations is known. Further more,

$$\begin{pmatrix}\epsilon_t\\ \eta_t\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}H_t & 0\\ 0 & Q_t\end{pmatrix}\right)$$

where H_t and Q_t are known.

Example 1 (Linear Model): Consider the linear and Gaussian state-space model.

Measurement Equation: $y_t = Z_t\alpha_t + \epsilon_t$,

Transition Equation: $\alpha_t = T_t\alpha_{t-1} + \eta_t$,

For prediction, we pick $\alpha_{i,t|t}$ randomly, generate normal random draws of η_{t+L} for $L=1,2,\cdots$, and have the random numbers of $\alpha_{t+L|t}$ given Y_t from $N(T_{t+L}\alpha_{i,t+L-1|t}, Q_{t+L})$ for $L=1,2,\cdots$.

For filtering, from $P_y(y_t|\alpha_t)$ and $P_\alpha(\alpha_t|\alpha_{t-1})$, the posterior density $P(\alpha_t|Y_t)$ is obtained. Picking $\alpha_{i,t-1|t-1}$ randomly, $P(\alpha_t|Y_t)$ is represented as average of the normal densities with mean $\Sigma_t\mu_{i,t}$ and variance Σ_t, i.e.,

$$P(\alpha_t|Y_t) \approx \frac{1}{n}\sum_{i=1}^{n}N(\Sigma_t\mu_{i,t},\Sigma_t),$$

where

$$\mu_{i,t} = Q_t^{-1} T_t \alpha_{i,t-1|t-1} + Z_t' H_t^{-1} y_t,$$

$$\Sigma_t^{-1} = Z_t' H_t^{-1} Z_t + Q_t^{-1}.$$

Note that $\alpha_{i,t-1|t-1}$ is chosen at random. Thus, the random draw for α_t given Y_t is obtained from the mixture distribution.

For smoothing, from $P_y(y_t|\alpha_t)$, $P_\alpha(\alpha_t|\alpha_{t-1})$ and $P_\alpha(\alpha_{t+1}|\alpha_t)$, the posterior density $P(\alpha_t|Y_T)$ is represented as an average of the normal densities with mean $\Sigma_t^* \mu_{ij,t}^*$ and variance Σ_t^*, i.e.,

$$P(\alpha_t|Y_T) \approx \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{q_{j,t}} N(\Sigma_t^* \mu_{ij,t}^*, \Sigma_t^*),$$

where

$$\mu_{ij,t}^* = Q_t^{-1} T_t \alpha_{i,t-1|t-1} + Z_t' H_t^{-1} y_t + \alpha_{j,t+1|T} Q_{t+1}^{-1} T_{t+1}$$

$$\Sigma_t^{*-1} = Z_t' H_t^{-1} Z_t + Q_t^{-1} + T_{t+1}' Q_{t+1}^{-1} T_{t+1}.$$

$\alpha_{i,t-1|t-1}$ and $\alpha_{j,t+1|T}$ are picked with probability $1/n$ and probability $1/q_{j,t}$, respectively, and the random draw for α_t given Y_T is obtained. The endpoint case (i.e., $\alpha_{j,T|T}$) denotes the filtering estimate.

Thus, in the linear and normal case, we do not need to use rejection sampling for random number generation.

Example 2 Now, consider the case where the transition equation is nonlinear and the measurement equation is linear:

Measurement Equation: $y_t = Z_t \alpha_t + \epsilon_t,$

Transition Equation: $\alpha_t = f_{1,t}(\alpha_{t-1}) + \eta_t.$

For prediction, given $\alpha_{i,t|t}$ and normal random draws of η_{t+L} for $L = 1, 2, \cdots$, we obtain the random draw of α_{t+L} given Y_t recursively, using $N(f_{1,t+L}(\alpha_{i,t+L-1|t}), Q_{t+L})$ for $L = 1, 2, \cdots$. i in $\alpha_{i,t|t}$ is chosen with equal probability, i.e., $1/n$.

For filtering, from $P_y(y_t|\alpha_t)$ and $P_\alpha(\alpha_t|\alpha_{t-1})$, the posterior density $P(\alpha_t|Y_t)$ is represented as an average of the normal densities with mean $\Sigma_t \mu_{i,t}$ and variance Σ_t, i.e.,

$$P_y(y_t|\alpha_t) \approx \frac{1}{n} \sum_{i=1}^{n} N(\Sigma_t \mu_{i,t}, \Sigma_t),$$

where

$$\mu_{i,t} = Q_t^{-1} f_{1,t}(\alpha_{i,t-1|t-1}) + Z_t' H_t^{-1} y_t,$$

$$\Sigma_t^{-1} = Z_t' H_t^{-1} Z_t + Q_t^{-1}.$$

Picking $\alpha_{i,t-1|t-1}$ at random, the random draw for α_t given Y_t is obtained from $N(\Sigma_t \mu_{i,t}, \Sigma_t)$.

For smoothing, from $P_y(y_t|\alpha_t)$, $P_\alpha(\alpha_t|\alpha_{t-1})$ and $P_\alpha(\alpha_{t+1}|\alpha_t)$, the posterior density $P(\alpha_t|Y_T)$ shown in equation (11) cannot be obtained explicitly. Therefore, the random numbers from the posterior density are generated utilizing rejection sampling. We pick $\alpha_{i,t-1|t-1}$ and $\alpha_{j,t+1|T}$ randomly. Then, we generate a random draw from $N(\Sigma_t \mu_{i,t}, \Sigma_t)$ and accept it with probability $w_2(\alpha_t; \alpha_{j,t+1|T})$, where the acceptance probability is taken as the exponential part of the normal density, i.e.,

$$w_2(\alpha_t; \alpha_{j,t+1|T}) = \exp\left(-\frac{1}{2}(\alpha_{j,t+1|T} - f_{1,t+1}(\alpha_t))' Q_{t+1}^{-1} (\alpha_{j,t+1|T} - f_{1,t+1}(\alpha_t))\right).$$

i and j are chosen with probability $1/n$ and probability $1/q_{j,t}$.

Thus, rejection sampling is used for smoothing but not for filtering.

Example 3. Here, the transition equation is linear and the measurement equation is non-linear.

$$\begin{aligned} &\text{Measurement Equation:} \quad y_t = h_{1,t}(\alpha_t) + \epsilon_t, \\ &\text{Transition Equation:} \quad \alpha_t = T_t \alpha_{t-1} + \eta_t, \end{aligned} \tag{12}$$

For prediction, given $\alpha_{i,t|t}$, generating normal random draws of η_{t+L} for $L = 1, 2, \cdots$, we obtain the random number of α_{t+L} given Y_t recursively, using $N(T_{t+L}\alpha_{i,t+L-1|t}, Q_{t+L})$ for $L = 1, 2, \cdots$.

For filtering, from $P_y(y_t|\alpha_t)$ and $P_\alpha(\alpha_t|\alpha_{t-1})$, the posterior density $P_\alpha(\alpha_t|Y_t)$ cannot be obtained explicitly. Therefore, we may apply rejection sampling. We pick $\alpha_{i,t-1|t-1}$ randomly, then generate a normal random draw $N(T_t\alpha_{i,t-1|t-1}, Q_t)$ and accept it with probability $w_1(\alpha_t; y_t)$, where

$$w_1(\alpha_t; y_t) = \exp\left(-\frac{1}{2}(y_t - h_{1,t}(\alpha_t))' H_t^{-1} (y_t - h_{1,t}(\alpha_t))\right).$$

For smoothing, from $P_y(y_t|\alpha_t)$, $P_\alpha(\alpha_t|\alpha_{t-1})$ and $P_\alpha(\alpha_{t+1}|\alpha_t)$, the posterior density $P_\alpha(\alpha_t|Y_T)$ cannot be obtained explicitly. Therefore, the random numbers from the posterior density

(11) are generated utilizing rejection sampling. We draw randomly from the normal density with mean $\Sigma_t^* \mu_t^*$ and variance Σ_t^*, i.e., $N(\Sigma_t^* \mu_t^*, \Sigma_t^*)$, where

$$\mu_t^* = Q_t^{-1} T_t \alpha_{t-1} + \alpha_{t+1} Q_{t+1}^{-1} T_{t+1},$$

$$\Sigma_t^{*-1} = Q_t^{-1} + T_{t+1}' Q_{t+1}^{-1} T_{t+1}.$$

Therefore, picking $\alpha_{i,t-1|t-1}$ and $\alpha_{j,t+1|T}$ randomly, we can get the random draw for α_t given Y_T from $N(\Sigma_t^* \mu_{ij,t}^*, \Sigma_t^*)$, where

$$\mu_{ij,t}^* = Q_t^{-1} T_t \alpha_{i,t-1|t-1} + Z_t' H_t^{-1} y_t + \alpha_{j,t+1|T} Q_{t+1}^{-1} T_{t+1}.$$

Then, we accept the normal random draw with probability $w_1(\alpha_t; y_t)$.

Example 4. Finally, we consider the case where both the transition equation and the measurement equation are nonlinear:

$$\begin{aligned} &\text{Measurement Equation:} \quad y_t = h_{1,t}(\alpha_t) + \epsilon_t, \\ &\text{Transition Equation:} \quad \alpha_t = f_{1,t}(\alpha_{t-1}) + \eta_t, \end{aligned} \tag{13}$$

where $h_{1,t}$, $f_{1,t}(\cdot)$, H_t and Q_t are assumed to be known.

For prediction, given $\alpha_{i,t|t}$ and normal random draws of η_{t+L} for $L = 1, 2, \cdots$, we obtain the random draw for α_{t+L} given Y_t recursively, using $N(f_{1,t+L}(\alpha_{i,t+L-1|t}), Q_{t+L})$ for $L = 1, 2, \cdots$.

For filtering, from $P_y(y_t|\alpha_t)$ and $P_\alpha(\alpha_t|\alpha_{t-1})$, the posterior density $P_\alpha(\alpha_t|Y_t)$ is not available explicitly. Therefore, we apply rejection sampling. We pick $\alpha_{i,t-1|t-1}$ randomly, then, generate a normal random draw $N(f_{1,t}(\alpha_{t-1}), Q_t)$ and accept it with probability $w_1(\alpha_t; y_t)$.

For smoothing, from $P_y(y_t|\alpha_t)$, $P_\alpha(\alpha_t|\alpha_{t-1})$ and $P_\alpha(\alpha_{t+1}|\alpha_t)$, the posterior density $P_\alpha(\alpha_t|Y_T)$ cannot be obtained explicitly. Random draws from the posterior density are generated utilizing rejection sampling. We pick $\alpha_{i,t-1|t-1}$ and $\alpha_{j,t+1|T}$ randomly, generate a random draw from $N\left(f_{1,t}(\alpha_{i,t-1|t-1}), Q_t\right)$, and accept it with probability $w_1(\alpha_t; y_t) w_2(\alpha_t; \alpha_{j,t+1|T})$.

3.2 Monte-Carlo Type 2 (MA2)

In this section, we propose another nonlinear filter using Monte-Carlo simulations as an alternative procedure for dealing with the nonlinear state-space model. Prediction, filtering and smoothing algorithms are developed under this approach.

First, note that the joint density function of (Y_s, A_T) can be written as

$$P(Y_s, A_T) = P(Y_s|A_s) P(A_T),$$

where

$$P(Y_s|A_s) = \prod_{t=1}^{s} P_y(y_t|\alpha_t),$$

$$P(A_T) = P_\alpha(\alpha_0) \prod_{t=1}^{T} P_\alpha(\alpha_t|\alpha_{t-1}).$$

$P_y(y_t|\alpha_t)$ and $P_\alpha(\alpha_t|\alpha_{t-1})$ are obtained directly from the measurement and transition equations. Therefore, the conditional density of α_r given information Y_s is given by:

$$P(\alpha_r|Y_s) = \frac{\int P(Y_s|A_s)P(A_T)\mathrm{d}A_{T,-r}}{\int P(Y_s|A_s)P(A_T)\mathrm{d}A_T},$$

where $A_{T,-r} = \{\alpha_T, \cdots, \alpha_{r+1}, \alpha_{r-1}, \cdots, \alpha_0\}$, the subset of A_t excluding α_r. The expectation of $g(\alpha_r)$ given Y_s may then be calculated as

$$g_{r|s} \equiv E(g(\alpha_r)|Y_s) = \frac{\int g(\alpha_r)P(Y_s|A_s)P(A_T)\mathrm{d}A_T}{\int P(Y_s|A_s)P(A_T)\mathrm{d}A_T},$$

which is equivalent to equation (7).

Let us define $\bar{\bar{g}}_{r|s}$ as the MA2 estimate of $g_{r|s}$ and $A_{i,T} = \{\alpha_{i,T}, \alpha_{i,T-1}, \cdots, \alpha_{i,0}\}$, $i = 1, \cdots, n$, as the random draws of A_T generated from $P(A_T)$. That is, $A_{i,s}$ is a set of random draws from the transition equation $\alpha_{i,t} = f_t(\alpha_{i,t-1}, \eta_{i,t})$ for $t = 1, \cdots, s$, where $\eta_{i,t}$ is a random number generated from the underlying density.

Given n set of random draws of A_T, (i.e., $A_{i,T}$ for $i = 1, \cdots, n$), $\bar{\bar{g}}_{r|s}$ is evaluated as:

$$\bar{\bar{g}}_{r|s} = \frac{\frac{1}{n}\sum_{i=1}^{n} g(\alpha_{i,r})P(Y_s|A_{i,s})}{\frac{1}{n}\sum_{j=1}^{n} P(Y_s|A_{j,s})}. \tag{14}$$

We refer to $\bar{\bar{g}}_{r|s}$ as MA2 prediction (MA2P) when $(r, s) = (t + L, t)$, MA2 filtering (MA2F) when $(r, s) = (t, t)$, and MA2 smoothing (MA2S) when $(r, s) = (t, T)$.

The following formula is easy and convenient for computing $\bar{\bar{g}}_{r|s}$. Since $P(Y_s|A_{i,s}) = \prod_{t=1}^{s} P_y(y_t|\alpha_{i,t})$, we can rewrite (14) as

$$\bar{\bar{g}}_{r|s} = \sum_{i=1}^{n} g(\alpha_{i,r})\omega_{i,s}, \tag{15}$$

where the weight function $\omega_{i,s}$ is recursively obtained as follows:

$$\omega_{i,s} = \frac{\prod_{t=1}^{s} P_y(y_t|\alpha_{i,t})}{\sum_{j=1}^{n}\prod_{t=1}^{s} P_y(y_t|\alpha_{j,t})} = \frac{P_y(y_s|\alpha_{i,s})\omega_{i,s-1}}{\sum_{j=1}^{n} P_y(y_s|\alpha_{j,s})\omega_{j,s-1}}. \tag{16}$$

Moreover, the initial values of the weight function are given by

$$\omega_{i,0} = \frac{1}{n}, \tag{17}$$

for $i = 1, \cdots, n$. Thus, (14) or (15) provides the MA2 estimate of $g(\alpha_r)$ given information Y_s, where Monte-Carlo simulation is utilized based on random draws $A_{i,s} = \{\alpha_{i,s}, \alpha_{i,s-1}, \cdots, \alpha_{i,0}\}$ generated from the transition equation. Note that $\omega_{i,s}$ represents the weight function such that $1 = \sum_{i=1}^{n} \omega_{i,s}$. Take L-step ahead prediction as an example. The L-step ahead prediction $g_{t+L|t}$ is represented as:

$$g_{t+L|t} = \frac{\int g(\alpha_{t+L})P(Y_t|A_t)P(A_{t+L})\mathrm{d}A_{t+L}}{\int P(Y_t|A_t)P(A_{t+L})\mathrm{d}A_{t+L}}.$$

Thus, the prediction estimate obtained by the Monte-Carlo simulation is given by:

$$\overline{\overline{g}}_{t+L|t} = \sum_{i=1}^{n} g(\alpha_{i,t+L})\omega_{i,t},$$

where $L = 1, 2, \cdots$.

The features of the above prediction, filtering and smoothing are that the random numbers of α_t (i.e., $\alpha_{i,t}$, $i = 1, \cdots, n$) are generated from the transition equation (2) for all t, and that the algorithm requires the functional form of the density function of y_t given α_t (i.e., $P_y(y_t|\alpha_t)$) for all t. We do not need *ad hoc* assumptions such as choice of the nodes for NI, that of the importance density for MII and that of the prior density for MIG. Note that, similar to MIG, but different from NI, MII and MA1, the MA2 procedure does not yield a recursive algorithm.

The computation of MA2 proceeds as follows:

(i) The random draws of the initial state variable α_0 are generated from the initial density $P_\alpha(\alpha_0)$, which are denoted by $\alpha_{i,0}$, $i = 1, \cdots, n$.

(ii) Given $\alpha_{i,0}$, the random numbers $\alpha_{i,t}$ are obtained from the transition equation $\alpha_{i,t} = f_t(\alpha_{i,t-1}, \eta_{i,t})$ using the random draws $\eta_{i,t}$ for $i = 1, \cdots, n$ and $t = 1, \cdots, T$.

(iii) Given the initial weight (17), use equation (16) to obtain the weight functions for $i = 1, \cdots, n$ and $t = 1, \cdots, T$.

(iv) $\overline{\overline{g}}_{r|s}$ can be computed by equation (15).

Thus, the transition equation (2) is utilized in order to generate the random numbers of A_T. The density function $P_y(y_t|\alpha_t)$ comes from the measurement equation (1). The proposed estimator needs to derive the density $P_y(y_t|\alpha_t)$ only, while for both the NI and the MII procedures, prior computation of $P_\alpha(\alpha_t|\alpha_{t-1})$ and $P_y(y_t|\alpha_t)$ is required. This implies much easier programming for MA2 in comparison with NI, MII and MIG. Moreover, the proposed procedure is more attractive with respect to computational time than the other procedures.

Unknown parameters can be estimated by numerical maximum likelihood estimation, using the following expression for the density function of Y_T:

$$P(Y_T) = \int P(Y_T|A_T)P(A_T)\mathrm{d}A_T$$

and the following approximation based on random draws for A_T:

$$P(Y_T) \approx \frac{1}{n}\sum_{i=1}^{n} P(Y_T|A_{i,T}).$$

Under appropriate conditions, the theorems proved by Geweke (1989a, 1989b), which are related to the asymptotic behavior of the Bayes mean using MII, hold without any modification in the case of MA2P, MA2F and MA2S proposed above. With $\Sigma_{r|s}$ defined as:

$$\Sigma_{r|s} = \frac{\int (g(\alpha_r) - g_{r|s})(g(\alpha_r) - g_{r|s})'(P(Y_s|A_s))^2 P(A_T)\mathrm{d}A_T}{(\int P(Y_s|A_s)P(A_T)\mathrm{d}A_T)^2},$$

and the sample variance of $\overline{\overline{\alpha}}_{r|s}$ given by

$$\overline{\overline{\Sigma}}_{r|s} = \frac{\sum_{i=1}^{n}(g(\alpha_{i,r}) - \overline{\overline{g}}_{r|s})(g(\alpha_{i,r}) - \overline{\overline{g}}_{r|s})'(P(Y_s|A_{i,s}))^2}{(\sum_{i=1}^{n} P(Y_s|A_{i,s}))^2},$$

as n goes to infinity, the following results hold:

(i) $\overline{\overline{g}}_{r|s} \xrightarrow{\text{a.s.}} g_{r|s}$,

(ii) $\sqrt{n}(\overline{\overline{g}}_{r|s} - g_{r|s}) \xrightarrow{\text{d}} N(0, \Sigma_{r|s})$,

$n\overline{\overline{\Sigma}}_{r|s} \xrightarrow{\text{a.s.}} \Sigma_{r|s}$,

where " $\xrightarrow{\text{a.s.}}$ " and " $\xrightarrow{\text{d}}$ " denote almost sure convergence and convergence in distribution, respectively. The above asymptotic properties indicate that $\overline{\overline{g}}_{r|s}$ is consistent, with a convergence rate of $\sqrt{n}$.

Finally, note that we would expect from equation (16) that simulation errors would increase as time t gets large. Accordingly, the MA2 procedure is intractable when T is large.In particular, when α_t is generated by a nonstationary process, MA2 would not perform well, since the random draws used in the procedure are generated form the transition equation alone.

4 Monte-Carlo Experiments

In this section, we use Monte-Carlo experiments to compare the nonlinear algorithms covered in section 2 and 3 and summarized in the following table.

Table 1: Algorithms for Nonlinear and Non-Gaussian State-Space Models

Estimate of $g_{r\|s}$	Estimate of $\alpha_{r\|s}$	Prediction	Filtering	Smoothing	Section
$\widetilde{g}_{r\|s}$	$\widetilde{\alpha}_{r\|s}$	NIP	NIF	NIS	Section 2.1.1
$\widehat{g}_{r\|s}$	$\widehat{\alpha}_{r\|s}$	MIIP	MIIF	MIIS	Section 2.1.2
$\overline{g}_{r\|s}$	$\overline{\alpha}_{r\|s}$	MIGP	MIGF	MIGS	Section 2.2
$g^*_{r\|s}$	$\alpha^*_{r\|s}$	MA1P	MA1F	MA1S	Section 3.1
$\overline{\overline{g}}_{r\|s}$	$\overline{\overline{\alpha}}_{r\|s}$	MA2P	MA2F	MA2S	Section 3.2

We deal with a linear and Gaussian model, as a benchmark, in Section 4.1, then consider a nonlinear transition equation in Section 4.2 and go into the general case, where both measurement and transition equations are nonlinear, in Section 4.3.

4.1 Linear and Gaussian Model

Consider the following linear scalar system:

$$\begin{aligned} &\text{Measurement Equation:} && y_t = \alpha_t + \epsilon_t, \\ &\text{Transition Equation:} && \alpha_t = \delta\alpha_{t-1} + \eta_t, \\ & && \begin{pmatrix}\epsilon_t \\ \eta_t\end{pmatrix} \sim N\left(\begin{pmatrix}0 \\ 0\end{pmatrix}, \begin{pmatrix}1 & 0 \\ 0 & 1\end{pmatrix}\right), \\ & && \alpha_0 \sim N(0,1), \end{aligned} \tag{18}$$

where $t = 1, \cdots, T$. We compare the following estimates: the Kalman (K) estimate $a_{r|s}$[3], the NI estimate $\widetilde{\alpha}_{r|s}$[4], the MII estimate $\widehat{\alpha}_{r|s}$[5], the MIG estimate $\overline{\alpha}_{r|s}$[6], the MA1 estimate $\alpha^*_{r|s}$ and the MA2 estimate $\overline{\overline{\alpha}}_{r|s}$. Note that the Kalman filter estimate $a_{t|s}$ for $s = t-1, t, T$ gives us the optimal solution in this example because the system (18) is a linear and Gaussian model.

The simulation procedure is as follows:

(i) Generating standard normal random numbers for ϵ_t and η_t for $t = 1, \cdots, T$, we obtain a set of data y_t and α_t, $t = 1, \cdots, T$, from the system, where $T = 40$.

(ii) We choose $n = 50$ for NI and MII, $n = 250$ for MIG, $n = 1000$ for MA1, and $n = 2500$ for MA2 to equalize computational burden for all the procedures.

(iii) Given Y_s, obtain $a_{r|s}$, $\widetilde{\alpha}_{r|s}$, $\widehat{\alpha}_{r|s}$, $\overline{\alpha}_{r|s}$, and $\overline{\overline{\alpha}}_{r|s}$ for $t = 1, \cdots, T$ and $(r, s) = (t, t-1), (t, t), (t, T)$.

(iv) Repeat (i)–(iii) G times and compute the bias (BIAS) and the root mean square error (RMSE) for each estimate, defined as:

$$\text{BIAS}(\hat{\alpha}_{r|s}) = \frac{1}{T}\sum_{r=1}^{T}\left(\frac{1}{G}\sum_{g=1}^{G}(\alpha_{r|s}^{(g)} - \alpha_r^{(g)})\right),$$

---

[3]We define $a_{r|s}$ as the K prediction estimate (KP) when $(r, s) = (t, t-1)$, the K filtering estimate (Kalman filter estimate, i.e., KF) when $(r, s) = (t, t)$ and the K smoothing estimate (Kalman smoothed estimate, i.e., KS) when $(r, s) = (t, T)$. In the case of linear and normal cases, $a_{r|s}$ is given by the conventional linear recursive algorithms, which are derived from equations (3) – (6). In the case of nonlinear state-space model (Nonliner Models I and II), K implies the extended Kalman filter in this paper, which is the nonlinear filter based on the first-order Taylor series expansion.

[4]For NIF, half of the nodes are chosen from the interval $[a_{t|t-1} - 4\Sigma_{t|t-1}^{1/2}, a_{t|t-1} + 4\Sigma_{t|t-1}^{1/2}]$ and the rest of the nodes are from $[a_{t|t} - 4\Sigma_{t|t}^{1/2}, a_{t|t} + 4\Sigma_{t|t}^{1/2}]$, where $a_{t|t}$ and $\Sigma_{t|t}$ denote the K filtering estimates of α_t and its variance (i.e., KF) and $a_{t|t-1}$ and $\Sigma_{t|t-1}$ denote the one-step ahead K prediction estimates of α_t and its variance (i.e., KP).

[5]For MIIF, we generate the random numbers for $\alpha_{i,t}$ from the normal density $N(a_{t|t-1}, 4\Sigma_{t|t-1})$ with probability 1/2 and from the normal density $N(a_{t|t}, 4\Sigma_{t|t})$ with probability 1/2, which corresponds to the importance density represented by:

$$P_I(\alpha_t) = \frac{1}{2}N(a_{t|t-1}, 4\Sigma_{t|t-1}) + \frac{1}{2}N(a_{t|t}, 4\Sigma_{t|t}).$$

For precision of density approximation, the importance density $P_I(\alpha_t)$ should be larger than $P(\alpha_t|Y_{t-1})$ and $P(\alpha_t|Y_t)$.

See, for example, Tanizaki (1993), Tanizaki and Mariano (1994) and Mariano and Tanizaki (1995) for choice of the importance density.

[6]The error terms are assumed to be $\epsilon_t \sim N(0, \gamma_t\Sigma_\epsilon)$ and $\eta_t \sim N(0, \lambda_t\Sigma_\eta)$, where $\Sigma_\epsilon = \Sigma_\eta = 1$. Also, we assume that the nuisance parameters are distributed as $\gamma_t \sim \text{IG}\left(\frac{1}{2}, \frac{1}{2}\right)$ and $\lambda_t \sim \text{IG}\left(\frac{1}{2}, \frac{1}{2}\right)$, where $x \sim \text{IG}(a, b)$ implies that $2a/x$ is distributed as a Chi-square random variable with $2b$ degrees of freedom.

$$\mathrm{RMSE}(\hat{\alpha}_{r|s}) = \frac{1}{T}\sum_{r=1}^{T}\left(\frac{1}{G}\sum_{g=1}^{G}(\alpha_{r|s}^{(g)} - \alpha_r^{(g)})^2\right)^{1/2},$$

where $\hat{\alpha}_{r|s}$ represents $a_{r|s}$, $\tilde{\alpha}_{r|s}$, $\widehat{\alpha}_{r|s}$, $\overline{\alpha}_{r|s}$ $\alpha^*_{r|s}$ or $\overline{\overline{\alpha}}_{r|s}$. The superscript (g) denotes the g-th simulation run; we take $G = 1000$. Thus, $\alpha_r^{(g)}$ denotes the simulated state variable at time r in the g-th simulation run.

We would expect K to be better than any other estimator because the system is linear and Gaussian.

Table 2a: Linear and Normal Model ($\delta = 1.0$)

		Prediction	Filtering	Smoothing
BIAS	K	0.0069	0.0031	0.0003
	NI	0.0069	0.0031	0.0003
	MII	0.0082	0.0045	0.0011
	MIG	0.1025	0.0440	0.0724
	MA1	0.0086	0.0047	0.0004
	MA2	0.0105	0.0045	0.0012
RMSE	K	1.2786	0.7845	0.6718
	NI	1.2789	0.7845	0.6718
	MII	1.2906	0.7946	0.6827
	MIG	3.6450	1.9469	2.7571
	MA1	1.3004	0.8175	0.6939
	MA2	1.6105	1.1480	1.3319

The results are in Table 2a. NI is very close to K. MIG and MA2 are poor estimators, which diverge from the true value as t increases. K, NI and MII are slightly better than MA1 with respect to BIAS and RMSE but not too different. Since MIG gives us poor results in Table 2a and a computational burden of MIG is too large, hereafter we do not consider MIG. Also, the result of MA2 is not reliable from Table 2a. Therefore, we do not pay too much attention to MA2.

Table 2b: Linear and Normal Model

		AVE	SER	RMSE	10%	25%	50%	75%	90%
$\delta = 0.5$	K	0.444	0.224	0.231	0.140	0.310	0.500	0.610	0.690
	NI	0.444	0.224	0.231	0.140	0.310	0.500	0.610	0.690
	MII	0.439	0.229	0.237	0.120	0.310	0.490	0.610	0.680
	MA1	0.374	0.214	0.248	0.095	0.240	0.410	0.540	0.620
	MA2	0.425	0.250	0.260	0.090	0.290	0.460	0.610	0.710
$\delta = 1.0$	K	0.959	0.079	0.089	0.865	0.930	0.980	1.010	1.030
	NI	0.959	0.079	0.089	0.870	0.930	0.980	1.010	1.030
	MII	0.958	0.080	0.091	0.855	0.930	0.980	1.010	1.030
	MA1	0.963	0.106	0.113	0.830	0.920	0.980	1.030	1.060
	MA2	0.999	0.042	0.042	0.980	0.990	1.000	1.010	1.025

Table 2b shows the estimation results for the unknown parameter δ, where the true parameter value is taken as $\delta = 0.5, 1.0$. MA1 shows larger RMSE than any other estimators. We obtain the result that NI and MII are close to K. In the case of $\delta = 0.5$, AVE of MA1 is small, compared with that of K, NI and MII.

4.2 Nonlinear Model I

Consider the following nonlinear scalar system:

$$\text{Measurement Equation:} \quad y_t = \alpha_t + \epsilon_t,$$

$$\text{Transition Equation:} \quad \alpha_t = (1 - \delta + \delta\alpha_{t-1}^2)^{1/2}\eta_t,$$

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right),$$

$$\alpha_0 \sim N(0, 1),$$

where $0 \leq \delta < 1$ and $t = 1, \cdots, T$. We take $\delta = 0.5, 0.9$ The transition equation follows the ARCH(1) process, while the measurement equation consists of the ARCH(1) term and the error. In this Monte-Carlo experiment, the unconditional variance of α_t is assumed to be one. The measurement equation is linear but the transition equation is nonlinear. In this experiment, therefore, we do not expect K to perform well, since under this procedure, the transition equation is linearized with respect to α_{t-1} and η_t.

Table 3a: Nonlinear Model I

			Prediction	Filtering	Smoothing
$\delta = 0.5$	BIAS	K	0.0018	0.0029	0.0029
		NI	0.0018	0.0027	0.0020
		MII	0.0020	0.0037	0.0029
		MA1	0.0023	0.0028	0.0020
		MA2	0.0036	0.0042	0.0009
	RMSE	K	1.0242	0.7043	0.7043
		NI	1.0232	0.6878	0.6785
		MII	1.0270	0.6946	0.6857
		MA1	1.0243	0.6942	0.6911
		MA2	1.1175	0.8037	0.9296
$\delta = 0.9$	BIAS	K	−0.0041	0.0021	0.0021
		NI	−0.0041	0.0024	0.0020
		MII	−0.0043	0.0037	0.0030
		MA1	−0.0041	0.0019	0.0006
		MA2	−0.0042	−0.0012	−0.0026
	RMSE	K	1.2029	0.6565	0.6565
		NI	1.1983	0.5903	0.5794
		MII	1.2285	0.5821	0.5681
		MA1	1.2051	0.5857	0.5712
		MA2	1.2556	0.9874	1.0614

We compare precision of the estimated state variables for K, NI, MII, MA1 and MA2. The results are in Table 3a. NI, MII and MA1 perform better than the other estimators, while NI, MII and MA1 give us similar results. As expected, NI, MII and MA1 are better than K. MA2 is the worst estimator, showing divergence from the true value as time t increases.

Table 3b: Nonlinear Model I

		AVE	SER	RMSE	10%	25%	50%	75%	90%
$\delta = 0.5$	K	0.326	0.281	0.331	0.010	0.010	0.310	0.530	0.730
	NI	0.437	0.330	0.336	0.010	0.010	0.500	0.740	0.840
	MII	0.452	0.358	0.361	0.010	0.040	0.470	0.800	0.930
	MA1	0.465	0.350	0.351	0.010	0.075	0.500	0.800	0.930
	MA2	0.703	0.299	0.362	0.100	0.595	0.820	0.920	0.970
$\delta = 0.9$	K	0.636	0.291	0.393	0.170	0.460	0.680	0.890	0.990
	NI	0.612	0.364	0.464	0.010	0.205	0.790	0.880	0.940
	MII	0.773	0.306	0.331	0.110	0.730	0.910	0.990	0.990
	MA1	0.805	0.258	0.275	0.440	0.730	0.910	0.990	0.990
	MA2	0.900	0.148	0.148	0.770	0.880	0.950	0.990	0.990

In Table 3b, the estimates of δ are compared for K, NI, MII, MA1 and MA2. The appropriate likelihood function is maximized by a simple grid search, where we take $\delta = 0.01, 0.02, \cdots, 0.99$. The estimates by MA1 is the closest to the true value in the bias criterion, except for MA2 (we do not consider MA2 because it is not reliable from the results in Table 3a). In the case of $\delta = 0.9$, the MA1 estimates give us the smallest bias and smallest RMSE.

4.3 Nonlinear Model II

Consider the following nonlinear scalar system:

$$\begin{aligned} &\text{Measurement Equation:} \quad y_t = \frac{1}{1+\exp(-\alpha_t + \epsilon_t)}, \\ &\text{Transition Equation:} \quad \alpha_t = \frac{1}{1+\exp(-\alpha_{t-1} + \eta_t)}, \\ &\qquad \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right), \\ &\qquad \alpha_0 \sim N(0,1), \end{aligned} \tag{19}$$

where $t = 1, \cdots, T$. In this experiment, both the measurement and the transition equations are nonlinear. The measurement equation can be transformed into a linear function in the state variable:

$$\log\left(\frac{1}{y_t} - 1\right) = -\alpha_t + \epsilon_t.$$

For K, equation (19), rather than the above equation, is approximated by the first-order Taylor series expansion.

Table 4: Nonlinear Model II

		Prediction	Filtering	Smoothing
BIAS	K	0.0256	0.0072	−0.0021
	NI	−0.0004	−0.0004	−0.0005
	MII	−0.0020	−0.0019	−0.0022
	MA1	−0.0016	−0.0016	−0.0012
	MA2	−0.0009	−0.0009	−0.0009
RMSE	K	0.2041	0.2135	0.2151
	NI	0.2017	0.1974	0.1972
	MII	0.2035	0.1992	0.1990
	MA1	0.2017	0.1975	0.1972
	MA2	0.2018	0.1975	0.1974

The results are in Table 4. K is the worst estimator. NI, MA1 and MA2 are better than K and MII, but not too different from MII. In this nonlinear state-space model, K is clearly biased.

5 Summary

In this paper, two Monte-Carlo procedures are proposed for nonlinear filtering. One requires only random draws from appropriate conditonal distributions. Theother utilizes the density function of y_t given α_t and the random draws from the transition equation. Both procedures improve the other nonlinear filters developed in the past in three aspects: computational time, simplicity of computer programming and absence of *ad hoc* assumptions.

The NI procedure proposed in Kitagawa (1987) and Kramer and Sorenson (1988) has the following disadvantages: (i) location of nodes has to be set by the researcher, (ii) we have to derive the densities $P_\alpha(\alpha_t|\alpha_{t-1})$ and $P_y(y_t|\alpha_t)$, and (iii) computational time increases more than proportionately as the dimension of the state variable increases.

The problems of the MII procedure developed by Tanizaki (1993), Tanizaki and Mariano (1994) and Mariano and Tanizaki (1995) are: (i) an appropriate choice of the importance density $P_I(\alpha_t)$ has to be made, (ii) the densities $P_\alpha(\alpha_t|\alpha_{t-1})$ and $P_y(y_t|\alpha_t)$ have to be derived, as in the NI approach.

The MIG approach (Carlin, Polson and Stoffer (1992)) also has some problems: (i) *ad hoc* assumptions have to be made about the distributions of nuisance parameters, (ii) use of the Gibbs sampler requires a great amount of data storage, and (iii) the MIG also takes a

lot of time computationally.

As an alternative simple procedure, we have proposed the MA1 estimator. By random draws only, we evaluate estimates of prediction, filtering and smoothing. For random number generation, rejection sampling may be adopted. MA1 does not need to derive the functional form of $P_y(y_t|\alpha_t)$; however, it is used to obtain the weight $w_1(\alpha_t; y_t)$ in Section 3.1.

The MA2 approach also addressses the problems in NI, MII and MIG. The measurement equation is utilized for deriving the density $P_y(y_t|\alpha_t)$ while the transition equation is used to generate the random numbers of the state variable α_t.

The Monte-Carlo experiments have shown the result that MA1 performs almost as well as NI and MII. In addition to simplicity of computer programming and absence of *ad hoc* assumptions, precision of the estimates is also close to the NI procedure. Accordingly, the MA1 approach proposed in this paper might be recommended for nonlinear and non-Gaussian state-space modeling.

References

Alspach, D.L. and H.W. Sorenson (1972), " Nonlinear Bayesian Estimation Using Gaussian Sum Approximations, " *IEEE Transactions on Automatic Control*, Vol.AC-17, No.4, pp.439–448.

Anderson, B.D.O. and J.B. Moore (1979), *Optimal Filtering*, Prentice-Hall, New York.

Boswell, M.T., S.D. Gore, G.P. Patil and C. Taillie (1993), " The Art of Computer Generation of Random Variables, " in *Handbook of Statistics, Vol.9*, edited by C.R. Rao, pp.661–721, North-Holland.

Brown, B.W. and R.S. Mariano (1984), " Residual-based Procedures for Prediction and Estimation in a Nonlinear Simultaneous System, " *Econometrica*, Vol.52, No.2, pp.321–343.

Brown, B.W. and R.S. Mariano (1989), " Measures of Deterministic Prediction Bias in Nonlinear Models, " *International Economic Review*, Vol.30, No.3, pp.667–684.

Carlin, B.P. and N.G. Polson (1991), " Inference for Nonconjugate Bayesian Models Using the Gibbs Sampler, " *Canadian Journal of Statistics*, Vol.19, pp.399–405.

Carlin, B.P., N.G. Polson and D.S. Stoffer (1992), " A Monte Carlo Approach to Nonnormal and Nonlinear State Space Modeling, " *Journal of the American Statistical Association*, Vol.87, pp.493–500.

Gelb, A. (Ed.) (1974), *Applied Optimal Estimation*, MIT Press.

Gelfand, A.E., S.E. Hills, H.A. Racine-Poon, and A.F.M. Smith (1990), " Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling, " *Journal of the American Statistical Association*, Vol.85, No.412, pp.972–985.

Gelfand, A.E. and A.F.M. Smith (1990), " Sampling-Based Approaches to Calculating Marginal Densities, " *Journal of the American Statistical Association*, Vol.85, No.410, pp.398–409.

Geman, S. and D. Geman (1984), " Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, " *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. Pami-6, No.6, pp.721–741.

Geweke, J. (1988), " Antithentic Acceleration of Monte Carlo Integration in Bayesian Inference, " *Journal of Econometrics*, Vol.38, pp.73–90.

Geweke, J. (1989a), " Modeling with Normal Polynomial Expansions, " in *Economic Complexity: Chaos, Sunspots, Bubbles and Nonlinearity*, edited by Barnett *et al.*, Cambridge University Press.

Geweke, J. (1989b), " Bayesian Inference in Econometric Models Using Monte Carlo Integration, " *Econometrica*, Vol.57, pp.1317–1339.

Harvey, A.C. (1989), *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge University Press.

Jacquier, E., N.G. Polson and P.E. Rossi (1994), " Bayesian Analysis of Stochastic Volatility Models, " *Journal of Business and Economic Statistics*, Vol.12, No.4, pp.371–417 (with discussion).

Kalman, R.E., 1960, " A New Approach to Linear Filtering and Prediction Problems, " *Journal of Basic Engineering, Transactions ASME*, Ser.D, 82, pp.35–45.

Kalman, R.E. and R.S. Bucy, 1961, " New Results in Linear Filtering and Prediction Theory, " *Journal of Basic Engineering, Transactions ASME*, Ser.D, 83, pp.95–108.

Kitagawa, G. (1987), " Non-Gaussian State-Space Modeling of Nonstationary Time Series, " *Journal of the American Statistical Association*, Vol.82, pp.1032–1063 (with discussion).

Knuth, D.E. (1981), *The Art of Computer Programming, Vol.2* (Seminumerical Algorithms, 2nd ed.), Addison-Wesley.

Koop, G. (1994), " Recent Progress in Applied Bayesian Econometrics, " *Journal of Economic Surveys*, Vol.8, No.1, pp.1–34.

Kramer, S.C. and Sorenson H.W. (1988), " Recursive Bayesian Estimation Using Piece-wise Constant Approximations, " *Automatica*, Vol.24, No.6, pp.789–801.

McFadden, D. (1989), " A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration, " *Econometrica*, Vol.57, No.5, pp.995–1026.

Mariano, R.S. and B.W. Brown (1983), " Asymptotic Behavior of Predictors in a Nonlinear Simultaneous System, " *International Economic Review*, Vol.24, No.3, pp.523–536.

Mariano, R.S. and B.W. Brown (1989), " Stochastic Simulation, Prediction and Validation of Nonlinear Models, " in *Economics in Theory and Practice: An Eclectic Approach*, edited by L.R. Klein and J. Marquez, pp.17–36, Kluwer Academic Publishers.

Mariano, R.S. and S. Schleicher (1972), " Kalman Filtering and Forecasting with Preliminary Data in Macroeconometric Models, " *University of Pennsylvania Discussion Paper.*

Mariano, R.S. and H. Tanizaki (1995), " Prediction of Final Data with Use of Preliminary and/or Revised Data, " *Journal of Forecasting*, Vol.14, No.4, pp.351–380.

Meinhold, R.J. and N.D. Singpurwalla (1989), " Robustification of Kalman Filter Models, " *Journal of the American Statistical Association*, Vol.84, pp.479–486.

O'Hagan, A. (1994), *Kendall's Advanced Theory of Statistics, Vol.2B* (Bayesian Inference), Edward Arnold.

Pole, A. and M. West (1990), " Efficient Bayesian Learning in Non-linear Dynamic Models, " *Journal of Forecasting*, Vol.9, No.2, pp.119–136.

Shao, J. (1989), " Monte Carlo Approximations in Bayesian Decision Theory, " *Journal of the American Statistical Association*, Vol.84, No.407, pp.727–732.

Smith, J.Q. (1981), " The Multiparameter Steady Model, " *Journal of the Royal Statistical Society*, Ser.B, Vol.43, No.2, pp.256–260.

Sorenson, H.W. and D.L. Alspach (1971), " Recursive Bayesian Estimation Using Gaussian Sums, " *Automatica*, Vol.7, pp.465–479.

Tanizaki, H. (1993), *Nonlinear Filters: Estimation and Applications*, (Lecture Notes in Economics and Mathematical Systems, No.400), Springer-Verlag.

Tanizaki, H. and R.S. Mariano (1992), " Nonlinear Filters Based on Taylor Series Expansions, " unpublished manuscript.

Tanizaki, H. and R.S. Mariano (1994), " Prediction, Filtering and Smoothing in Nonlinear and Nonnormal Cases Using Monte-Carlo Integration, " *Journal of Applied Econometrics*, Vol.9, No.2, pp.163–179.

West, M. and J. Harrison (1989), *Bayesian Forecasting and Dynamic Models*, Springer-Verlag.

Wishner, R.P., J.A. Tabaczynski and M. Athans (1969), " A Comparison of Three Non-Linear Filters, " *Automatica*, Vol.5, pp.487–496.

Zeger, S.L. and M.R. Karim (1991), " Generalized Linear Models with Random Effects: A Gibbs Sampling Approach, " *Journal of the American Statistical Association*, Vol.86, No.413, pp.79–86.

Appendix: Rejection Sampling

The rejection sampling is as follows: Let x be a random variable from a density function $f(\cdot)$. When we want to generate random draws from $f(\cdot)$, we need to find the density $g(\cdot)$ which satisfies $f(x) \leq cg(x)$ for all x, where c is constant. For $g(\cdot)$ we should choose the distribution function such that we can easily generate random draws. Define $w(x) \equiv f(x)/cg(x)$. Note that $0 \leq w(x) \leq 1$. Let u be a uniform random number between zero and one and v be a random number from $g(\cdot)$. Then, we take the following procedures: (i) generate u from a uniform distribution between zero and one, (ii) generate v from a density $g(\cdot)$, and (iii) take v as x if $u \leq w(v)$, and return to (i) otherwise.

It is known that the random number generator by rejection sampling is inefficient when the acceptance probability $w(v)$ is close to zero.

We can prove the above random number generation procedure as follows: Note that

$$\text{Prob}(X \leq x | u \leq w(v)) = \frac{\text{Prob}(X \leq x, u \leq w(v))}{\text{Prob}(u \leq w(v))},$$

where the numerator and the denominator are represented as:

$$\begin{aligned} \text{Prob}(X \leq x, u \leq w(v)) &= \int_{-\infty}^{x} \text{Prob}(u \leq w(v) | v = t) g(t) \mathrm{d}t \\ &= \int_{-\infty}^{x} w(t) g(t) \mathrm{d}t \\ &= F(x)/c, \end{aligned}$$

$$\begin{aligned} \text{Prob}(u \leq w(v)) &= \text{Prob}(X \leq \infty, u \leq w(v)) \\ &= F(\infty)/c \\ &= 1/c, \end{aligned}$$

where $F(x)$ denotes the cumulative distribution function of x, i.e., $\mathrm{d}F(x)/\mathrm{d}x = f(x)$.

Therefore, we have:

$$\text{Prob}(X \leq x | u \leq w(v)) = F(x).$$

See, for example, Knuth (1981), Boswell, Gore, Patil and Taillie (1993) and O'Hagan (1994) for more details on rejection sampling.