Nonlinear and non-Gaussian state-space modeling with Monte Carlo simulations

Tanizaki, Hisashi

Roberto, S. Mariano

(Citation)
Journal of Econometrics,83(1-2):263-290

(Issue Date)
1998

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90000575

Nonlinear and Nonnormal State-Space Modeling with Monte-Carlo Stochastic Simulations

Hisashi TANIZAKI
Kobe University
Kobe 657, JAPAN

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

May, 1995

Abstract: Two nonlinear and nonnormal filters using Monte-Carlo simulation techniques are proposed in this paper. Both of the proposed filters have less computational burden and easier programming than the other nonlinear and nonnormal filters such as the numerical integration procedure and the Monte-Carlo integration approaches with importance sampling and Gibbs sampling. Furthermore, the proposed filters are extended to prediction and smoothing algorithms. Finally, by Monte-Carlo experiments, we compare all the nonlinear and nonnormal procedures.

Key Words: Prediction, Filtering, Smoothing, Numerical Integration, Monte-Carlo Integration with Importance Sampling, Monte-Carlo Integration with Gibbs Sampling, and Monte-Carlo Stochastic Simulations.

1 Introduction

The filtering theory developed by Kalman (1960) and Kalman and Bucy (1961) has been extended in two directions. One extension of the Kalman filter is nonlinearity of measurement and transition equations. Another is nonnormality of error terms in the state-space model. The Kalman filter represented by the conventional linear recursive algorithm is optimal only if the system is linear and the error terms in the system are normal. Therefore, nonlinear and nonnormal filters have been developed by these two extensions.

The nonlinear measurement and transition equations are linearized by a Taylor series expansion and the linearized system is directly applied to the linear recursive algorithm. Thus, the extended Kalman filter, the second-order nonlinear filter and so on are derived (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 nonlinear filters based on the Taylor series expansions give us the biased estimators (see Tanizaki (1993), Tanizaki and Mariano (1994) and Mariano and Tanizaki (1995)).

Furthermore, Meinhold and Singpurwalla (1989) pointed out that the Kalman filter model is not robust under the normality assumption. The state-space model with the nonnormal errors has been developed based on the underlying density functions, where a recursive algorithm of density functions is derived. Kitagawa (1987) and Kramer and Sorenson (1988) evaluate each density by Numerical Integration (NI). Recently, an attempt has been made to apply simulation techniques in order to evaluate a nonlinear and nonnormal equation (for example, Brown and Mariano (1984, 1989), McFadden (1989) and Mariano and Brown (1983, 1989)). Tanizaki (1993), Tanizaki and Mariano (1994) and Mariano and Tanizaki (1995) applied a simulation technique to the filtering theory, utilizing Monte-Carlo Integration with Importance sampling (MII)[1] to obtain each integration included in the recursive algorithm of densities. Moreover,

[1] For MII, see Geweke (1988, 1989a, 1989b) and Shao (1989).

Carlin, Polson and Stoffer (1992) proposed a nonlinear and nonnormal state-space modeling by Monte-Carlo Integration with Gibbs sampling (MIG)[2] , which does not yield a recursive algorithm and includes the nuisance parameters in the underlying density functions. The disadvantage of these three nonlinear and nonnormal procedures based on the densities is to require a great amount of computational burden and a complicated computer programming. In addition, each conditional density function has to be calculated by hand based on the measurement and transition equations in order to make a program of each filter. Improving these problems, by Monte-Carlo approaches, we propose two nonlinear and nonnormal filters; (i) one utilizes the random draws based on the transition equation and the conditional densities obtained from the measurement equation (Monte-Carlo Approach Type 1, i.e., MA1) and (ii) another uses the random draws only (Monte-Carlo Approach Type 2, i.e., MA2). The proposed filters have less computational burden and can be easily implemented. Also, the nonlinear and nonnormal procedures are extended to prediction and smoothing algorithms in the exactly same fashion. Finally, by Monte-Carlo experiments, we compare MA1 prediction (MA1P), MA1 filter (MA1F), MA1 smoothing (MA1S), MA2 prediction (MA2P), MA2 filter (MA2F) and MA2 smoothing (MA2S) with NI prediction (NIP), NI filter (NIF), NI smoothing (NIS), MII prediction (MIIF), MII filter (MIIF), MII smoothing (MIIS), MIG prediction (MIGF), MIG filter (MIGF) and MIG smoothing (MIGS).

2 Overview of Nonlinear and Nonnormal State-Space Modeling

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

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

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

where $t = 1, \cdots, T$. A $p \times 1$ vector, y_t, is observable while a $k \times 1$ 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 have to be assumed but not necessarily normal. 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).

We have two density-based algorithms: one is a recursive algorithm and another is not. The former includes the NI procedure (Kitagawa (1987) and Kramer and Sorenson (1988)) and the MII approach (Tanizaki (1993), Tanizaki and Mariano (1994) and Mariano and Tanizaki (1995)), while the latter gives us the MIG procedure (Carlin, Polson and Stoffer (1992)). The MA1 approach belongs to the latter (i.e., nonrecursive algorithm) but the MA2 approach belongs to the former (i.e., recursive algorithm).

2.1 Recursive Algorithms

Given the above conventional nonlinear and nonnormal state-space model (1) and (2), for evaluation of the state vector α_t we have three kinds of estimates, i.e., prediction, filtering and smoothing.

The density algorithm on L-step ahead prediction is represented 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.

[2] 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 recursive density algorithm on filtering is known 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 $P_\alpha(\alpha_t|\alpha_{t-1})$ and $P_y(y_t|\alpha_t)$ 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, repeating predicting and updating for all t, the filtering densities $P(\alpha_t|Y_t)$, $t = 1, \cdots, T$, can be obtained.

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 above filtering algorithm represented by equations (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).

After computing each density, the expectation of a function $g(\alpha_r)$ is given by:

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

where $(r, s) = (t+L, t), (t, t), (t, T)$. The vector function $g(\cdot)$ is typically specified as $g(\alpha_r) = \alpha_r$ or $g(\alpha_r) = (\alpha_r - \alpha_{r|s})(\alpha_r - \alpha_{r|s})'$, where $\alpha_{r|s} = E(\alpha_r|Y_s)$.

When the unknown parameters are included in the system (1) and (2), the following likelihood function is maximized:

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

because $P(y_t|Y_{t-1}) = \int P_y(y_t|\alpha_t)P(\alpha_t|Y_{t-1})\mathrm{d}\alpha_t$, which corresponds to the denominator of equation (5).

The above integration in equations (3) – (8) is evaluated in two ways, i.e., NI and MII.

2.1.1 Numerical Integration (NI)

Kitagawa (1987) and Kramer and Sorenson (1988) proposed a nonlinear and nonnormal filter using NI. Consider a scalar case of the state vector. NI requires the nodes which are denoted by $\alpha_{i,t}$, $i = 0, \cdots, n$. For all t, $\alpha_{i,t}$ 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 some methods to evaluate integration numerically; a rectangle rule, a sum of trapezoids, Simpson's formula and so on. Here, for simplicity of discussion, we take the NI method by a sum of rectangles.

By NI, equation (3) is evaluated as:

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

which is called NI prediction (NIP).

For an algorithm of NI filter (NIF), equations (4) and (5) are evaluated as:

$$P(\alpha_{i,t}|Y_{t-1}) = \sum_{j=1}^{n} P_\alpha(\alpha_{i,t}|\alpha_{j,t-1})P(\alpha_{j,t-1}|Y_{t-1})(\alpha_{j,t-1} - \alpha_{j-1,t-1}), \tag{10}$$

$$P(\alpha_{i,t}|Y_t) = \frac{P_y(y_t|\alpha_{i,t})P(\alpha_{i,t}|Y_{t-1})}{\sum_{j=1}^{n} P_y(y_t|\alpha_{j,t})P(\alpha_{j,t}|Y_{t-1})(\alpha_{j,t} - \alpha_{j-1,t})}, \tag{11}$$

where $i = 1, \cdots, n$ and $t = 1, \cdots, T$.

Equation (6) is rewritten as:

$$P(\alpha_{i,t}|Y_T) = P(\alpha_{i,t}|Y_t) \sum_{j=1}^{n} \frac{P(\alpha_{j,t+1}|Y_T)P_\alpha(\alpha_{j,t+1}|\alpha_{i,t})}{P(\alpha_{j,t+1}|Y_t)}(\alpha_{j,t+1} - \alpha_{j-1,t+1}), \tag{12}$$

which is called the NI smoothing (NIS) algorithm.

Let us denote $\widetilde{g}_{r|s}$ as the NI estimate of $g_{r|s}$ (which is denoted by equation (7)) for $(r,s) = (t+L,t),(t,t),(t,T)$. Using the above density $P(\alpha_{i,r}|Y_s)$, the expectation of $g_{r|s}$ is evaluated as:

$$\widetilde{g}_{r|s} = \sum_{i=1}^{n} g(\alpha_{i,r})P(\alpha_{i,r}|Y_s)(\alpha_{i,r} - \alpha_{i-1,r}). \tag{13}$$

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 $P_y(y_t|\alpha_{i,t})$). There, the nodes $\alpha_{i,t}$, $i = 1, \cdots, n$ and $t = 1, \cdots, T$, have to be chosen by a researcher.

Finally, the likelihood function (8) is rewritten as:

$$P(Y_T) = \prod_{t=1}^{T} \sum_{j=1}^{n} P_y(y_t|\alpha_{j,t})P(\alpha_{j,t}|Y_{t-1})(\alpha_{j,t} - \alpha_{j-1,t}), \tag{14}$$

which is maximized with respect to the unknown parameter if it is included in the model.

For NIP, NIF and NIS, since computational errors drastically grow up as t increases, we need to impose $1 = \sum_{i=1}^{n} P(\alpha_{i,r}|Y_s)(\alpha_{i,r} - \alpha_{i-1,r})$ for $(r,s) = (t+L,t),(t,t),(t,T)$ and $t = 1, \cdots, T$ in order to avoid accumulation of the errors.[3]

The NI procedure can evaluate each expectation quite precisely. When the state vector α_t is extended to multidimensional cases, however,we should keep in mind that a computational burden increases more than proportionally and therefore that NIP, NIF and NIS are intractable in such a case.

2.1.2 Monte-Carlo Integration with Importance Sampling (MII)

Tanizaki (1993), Tanizaki and Mariano (1994) and Mariano and Tanizaki (1995) developed a nonlinear and nonnormal filter with MII, where a recursive algorithm of the density functions is converted to that of the weight functions. Define the weight function $\omega_{r|s} = \dfrac{P(\alpha_r|Y_s)}{P_I(\alpha_r)}$ for $(r,s) = (t+L,t),(t,t),(t,T)$, which is a ratio of the two densities. The density function $P_I(\alpha_t)$ has to be appropriately specified by a researcher,

[3] $1 = \sum_{i=1}^{n} P(\alpha_{i,r}|Y_s)(\alpha_{i,r} - \alpha_{i-1,r})$ comes from $1 = \int P(\alpha_r|Y_s)\mathrm{d}\alpha_r$.

which is called the importance density. Moreover, define $\omega_{i,r|s} = \dfrac{P(\alpha_{i,r}|Y_s)}{P_I(\alpha_{i,r})}$ as the weight function evaluated at $\alpha_{i,r}$, where $\alpha_{i,r}$, $i = 1, \cdots, n$, are the random draws generated from the importance density $P_I(\alpha_r)$.

By MII, equation (3) is given by[4] :

$$\omega_{i,t+L|t} = \frac{1}{n}\sum_{j=1}^{n} \frac{P_\alpha(\alpha_{i,t+L}|\alpha_{j,t+L-1})}{P_I(\alpha_{i,t+L})}\omega_{j,t+L-1|t}, \tag{15}$$

which is called MII prediction (MIIP).

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

$$\omega_{i,t|t-1} = \frac{1}{n}\sum_{j=1}^{n} \frac{P_\alpha(\alpha_{i,t}|\alpha_{j,t-1})}{P_I(\alpha_{i,t})}\omega_{j,t-1|t-1}, \tag{16}$$

$$\omega_{i,t|t} = \frac{P_y(y_t|\alpha_{i,t})\omega_{i,t|t-1}}{\dfrac{1}{n}\displaystyle\sum_{j=1}^{n} P_y(y_t|\alpha_{j,t})\omega_{j,t|t-1}}, \tag{17}$$

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

In order to obtain MII smoothing (MIIS), equation (6) is rewritten as:

$$\omega_{i,t|T} = \omega_{i,t|t}\frac{1}{n}\sum_{j=1}^{n} \frac{\omega_{j,t+1|T}P_\alpha(\alpha_{j,t+1}|\alpha_{i,t})}{\omega_{j,t+1|t}P_I(\alpha_{j,t+1})}, \tag{18}$$

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

$$\widehat{g}_{r|s} = \frac{1}{n}\sum_{i=1}^{n} g(\alpha_{i,r})\omega_{i,r|s}, \tag{19}$$

where $\widehat{g}_{r|s}$ denotes the MII estimate of $g_{r|s}$ based on the MII for $(r, s) = (t + L, t), (t, t), (t, T)$. Thus, MIIP, MIIF and MIIS depend on the number of random numbers (i.e., n) and choice of the importance density (i.e., $P_I(\alpha_t)$).

The likelihood function (8) is rewritten as:

$$P(Y_T) = \prod_{t=1}^{T}\frac{1}{n}\sum_{j=1}^{n} P_y(y_t|\alpha_{j,t})\omega_{j,t|t-1}. \tag{20}$$

For MIIP, MIIF and MIIS, since simulation errors drastically grow up as t increases, we need to impose $1 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\omega_{i,r|s}$ for $(r, s) = (t+L, t), (t, t), (t, T)$ and $t = 1, \cdots, T$ in order to avoid accumulation of simulation errors.[5]

[4] For derivation of equation (15), using the weight functions $\omega_{t+L|t}$ and $\omega_{t+L-1|t}$, the prediction density (3) is rewritten 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 (15).

In the exactly same fashion, equations (4) – (6) reduce to equations (16) – (18).

[5] $1 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}\omega_{i,r|s}$ comes from $1 = \displaystyle\int P(\alpha_r|Y_s)\mathrm{d}\alpha_r$. See Koop (1994) for a survey of integration methods.

The MII procedure can be easily implemented in multidimensional cases of the state vector, which is an attractive feature over the NI approach. It is easily shown that $\widehat{g}_{r|s}$ is consistent but convergence is quite slow as $\sqrt{n}$.[6]

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 nonnormal errors and nonlinear functions in the state-space model. They introduced nuisance parameters γ and λ into the model. Their crucial assumptions are that distribution functions of ϵ_t and η_t depend on nuisance parameters γ_t and λ_t and that the distribution functions of γ_t and λ_t are given by $P_\gamma(\gamma_t)$ and $P_\lambda(\lambda_t)$, which are called the prior densities in the 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 above setup, the MIG procedure (i.e., Monte-Carlo integration with Gibbs sampler) is performed as follows:

1. 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's theorem.
   (a) $P_{\gamma_*}(\gamma_t|y_t, \alpha_t)$ comes from $P_y(y_t|\alpha_t, \gamma_t)$ and $P_\gamma(\gamma_t)$.
   (b) $P_{\lambda_*}(\lambda_t|\alpha_t, \alpha_{t-1})$ is obtained from $P_\alpha(\alpha_t|\alpha_{t-1}, \lambda_t)$ and $P_\lambda(\lambda_t)$.
   (c) $P_{\alpha_*}(\alpha_t|y_t, \alpha_{t-1}, \alpha_{t+1}, \gamma_t, \lambda_t, \lambda_{t+1})$ is derived from $P_y(y_t|\alpha_t, \gamma_t)$, $P_\alpha(\alpha_t|\alpha_{t-1}, \lambda_t)$ and $P_\alpha(\alpha_{t+1}|\alpha_t, \lambda_{t+1})$.[7]
2. Given the values with the superscript (j), generate the random numbers with $(j+1)$ for filtering, smoothing and prediction as follows.
   (a) MIG Filtering (MIGF):
      i. Generate $\gamma_s^{(j+1)}$ from $P_{\gamma_*}(\gamma_s|y_s, \alpha_s^{(j)})$ for $s = 1, \cdots, t$.
      ii. Generate $\lambda_s^{(j+1)}$ from $P_{\lambda_*}(\lambda_s|\alpha_s^{(j)}, \alpha_{s-1}^{(j)})$ for $s = 1, \cdots, t$.
      iii. Generate

$$\begin{aligned}
&\alpha_s^{(j+1)} \quad \text{from} \quad P_{\alpha_*}(\alpha_s|\alpha_{s+1}^{(j)}, \lambda_s^{(j)}, \lambda_{s+1}^{(j)}), && \text{for} \quad s = 0,\\
&\alpha_s^{(j+1)} \quad \text{from} \quad P_{\alpha_*}(\alpha_s|y_s, \alpha_{s-1}^{(j)}, \alpha_{s+1}^{(j)}, \gamma_s^{(j)}, \lambda_s^{(j)}, \lambda_{s+1}^{(j)}), && \text{for} \quad s = 1, \cdots, t-1,\\
&\alpha_s^{(j+1)} \quad \text{from} \quad P_{\alpha_*}(\alpha_s|y_s, \alpha_{s-1}^{(j)}, \gamma_s^{(j)}, \lambda_s^{(j)}), && \text{for} \quad s = t,
\end{aligned}$$

[6] As shown in Geweke (1989a,1989b), MII gives us a consistent estimate as the number of random draws increases. Note that $\widehat{g}_{r|s} = \frac{1}{n}\sum_{i=1}^{n} g(\alpha_{i,r})\omega_{i,r|s}$, where $g(\alpha_{i,r})\omega_{i,r|s}$ is the random number generated from $P_I(\alpha_r)$. Define $E\big(g(\alpha_{i,r})\omega_{i,r|s}\big) = E\big(g(\alpha_r)|Y_s\big) \equiv g_{r|s}$ and $E\big(g(\alpha_{i,r})g(\alpha_{i,r})'\omega_{i,r|s}^2\big) = E\big(g(\alpha_r)g(\alpha_r)'\omega_{r|s}|Y_s\big) \equiv \Sigma_{r|s}$. By the central limit theorem, it is shown that

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

[7] Taking into account the endpoint cases, we have

$$\begin{aligned}
&P_{\alpha_*}(\alpha_t|\alpha_{t+1}, \lambda_t, \lambda_{t+1}), && \text{for} \quad t = 0,\\
&P_{\alpha_*}(\alpha_t|y_t, \alpha_{t-1}, \alpha_{t+1}, \gamma_t, \lambda_t, \lambda_{t+1}), && \text{for} \quad t = 1, \cdots, T-1.\\
&P_{\alpha_*}(\alpha_t|y_t, \alpha_{t-1}, \gamma_t, \lambda_t), && \text{for} \quad t = T,
\end{aligned}$$

for $j = 1, \cdots, J$ and $t = 1, \cdots, T$.

As J goes to infinity, $\{\gamma_s^{(J+1)}, \lambda_s^{(J+1)}, \alpha_s^{(J+1)}\}_{s=1}^t$ approaches a set of random draws from the joint distribution of $\{\gamma_s, \lambda_s, \alpha_s\}_{s=1}^t$ given $\{y_s\}_{s=1}^t$. Let $\{\gamma_{i,s}, \lambda_{i,s}, \alpha_{i,s}\}_{s=1}^t$ be the i-th set of random draws (i.e., $\{\gamma_{i,s}^{(J+1)}, \lambda_{i,s}^{(J+1)}, \alpha_{i,s}^{(J+1)}\}_{s=1}^t$ for enough large J). Suppose that n sets of random draws are generated, i.e., $\{\gamma_{i,s}, \lambda_{i,s}, \alpha_{i,s}\}_{s=1}^t$ for $i = 1, \cdots, n$. Then, the filtering density is approximated as

$$P(\alpha_t|Y_t) = \frac{1}{n}\sum_{i=1}^{n} P_{\alpha_*}(\alpha_t|y_t, \alpha_{i,t-1}, \gamma_{i,t}, \lambda_{i,t}).$$

Thus, the filtering estimate based on MIG is obtained.

(b) MIG Smoothing (MIGS):

i. Generate $\gamma_s^{(j+1)}$ from $P_{\gamma_*}(\gamma_s|y_s, \alpha_s^{(j)})$ for $s = 1, \cdots, T$.

ii. Generate $\lambda_s^{(j+1)}$ from $P_{\lambda_*}(\lambda_s|\alpha_s^{(j)}, \alpha_{s-1}^{(j)})$ for $s = 1, \cdots, T$.

iii. Generate

$$\begin{aligned}
&\alpha_s^{(j+1)} \quad \text{from} \quad P_{\alpha_*}(\alpha_s|\alpha_{s+1}^{(j)}, \lambda_s^{(j)}, \lambda_{s+1}^{(j)}), && \text{for} \quad s = 0,\\
&\alpha_s^{(j+1)} \quad \text{from} \quad P_{\alpha_*}(\alpha_s|y_s, \alpha_{s-1}^{(j)}, \alpha_{s+1}^{(j)}, \gamma_s^{(j)}, \lambda_s^{(j)}, \lambda_{s+1}^{(j)}), && \text{for} \quad s = 1, \cdots, T-1,\\
&\alpha_s^{(j+1)} \quad \text{from} \quad P_{\alpha_*}(\alpha_s|y_s, \alpha_{s-1}^{(j)}, \gamma_s^{(j)}, \lambda_s^{(j)}), && \text{for} \quad s = T.
\end{aligned}$$

Similarly, the smoothing density is approximated as

$$P(\alpha_t|Y_T) = \begin{cases} \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} P_{\alpha_*}(\alpha_t|y_t, \alpha_{i,t-1}, \alpha_{i,t+1}, \gamma_{i,t}, \lambda_{i,t}, \lambda_{i,t+1}), & \text{for } t = 1, \cdots, T-1, \\ \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} P_{\alpha_*}(\alpha_t|y_t, \alpha_{i,t-1}, \gamma_{i,t}, \lambda_{i,t}), & \text{for } t = T, \end{cases}$$

Note that in the case of $t = T$ the smoothing density is equivalent to the filtering density.

(c) MIG Prediction (MIGP):

i. Generate $\gamma_s^{(j+1)}$ from

$$\begin{aligned}
&P_{\gamma_*}(\gamma_s|y_s, \alpha_s^{(j)}), && \text{for } s = 1, \cdots, t,\\
&P_{\gamma_*}(\gamma_s|y_s^{(j)}, \alpha_s^{(j)}), && \text{for } s = t+1, \cdots, t+L.
\end{aligned}$$

ii. Generate $\lambda_s^{(j+1)}$ from $P_{\lambda_*}(\lambda_s|\alpha_s^{(j)}, \alpha_{s-1}^{(j)})$ for $s = 1, \cdots, t+L$.

iii. Generate

$$\begin{aligned}
&\alpha_s^{(j+1)} \quad \text{from} \quad P_{\alpha_*}(\alpha_s|\alpha_{s+1}^{(j)}, \lambda_s^{(j)}, \lambda_{s+1}^{(j)}), && \text{for} \quad s = 0,\\
&\alpha_s^{(j+1)} \quad \text{from} \quad P_{\alpha_*}(\alpha_s|y_s, \alpha_{s-1}^{(j)}, \alpha_{s+1}^{(j)}, \gamma_s^{(j)}, \lambda_s^{(j)}, \lambda_{s+1}^{(j)}), && \text{for} \quad s = 1, \cdots, t,\\
&\alpha_s^{(j+1)} \quad \text{from} \quad P_{\alpha_*}(\alpha_s|y_s^{(j)}, \alpha_{s-1}^{(j)}, \alpha_{s+1}^{(j)}, \gamma_s^{(j)}, \lambda_s^{(j)}, \lambda_{s+1}^{(j)}), && \text{for} \quad s = t+1, \cdots, t+L-1,\\
&\alpha_s^{(j+1)} \quad \text{from} \quad P_{\alpha_*}(\alpha_s|y_s^{(j)}, \alpha_{s-1}^{(j)}, \gamma_s^{(j)}, \lambda_s^{(j)}), && \text{for} \quad s = t+L.
\end{aligned}$$

iv. Generate $y_s^{(j+1)}$ from $P_y(y_s|\alpha_s^{(j)}, \gamma_s^{(j)})$ for $s = t+1, \cdots, t+L$.[8]

[8] Note that $P_y(y_s|\alpha_s, \gamma_s)$ comes from the measurement equation (1).

The prediction density is approximated as

$$P(\alpha_{t+L}|Y_t) = \frac{1}{n}\sum_{i=1}^{n} P_{\alpha_*}(\alpha_{t+L}|y_{i,t+L}, \alpha_{i,t+L-1}, \alpha_{i,t+L+1}, \gamma_{i,t+L}, \lambda_{i,t+L}, \lambda_{i,t+L+1}),$$

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

3. Let us define $\overline{g}_{r|s}$ as the MIG estimate of $g_{r|s}$. Since $\alpha_{i,r}$ is the random draw generated from $P(\alpha_r|Y_s)$ for $(r,s) = (t+L,t), (t,t), (t,T)$, we can obtain the expectation as follows:

$$\overline{g}_{r|s} = \frac{1}{n}\sum_{i=1}^{n} g(\alpha_{i,r}). \tag{21}$$

Thus, for the MIG estimate of $g_{r|s}$, the nonlinear and nonnormal state-space models developed by Carlin, Polson and Stoffer (1992) utilize the random draws only.[9]

The likelihood function (8) is evaluated as

$$P(Y_T) = \prod_{t=1}^{T} \frac{1}{n}\sum_{i=1}^{n} P_y(y_t|\alpha_{i,t}, \gamma_{i,t}), \tag{22}$$

because the joint density of $\{Y_T, A_T, \Gamma_T, \Lambda_T\}$ is given by

$$P(Y_T, A_T, \Gamma_T, \Lambda_T) = P_\alpha(\alpha_0|\lambda_0)\prod_{t=1}^{T} P_y(y_t|\alpha_t, \gamma_t)P_\alpha(\alpha_t|\alpha_{t-1}, \lambda_t)P_\gamma(\gamma_t)P_\lambda(\lambda_t),$$

where $A_T = \{\alpha_T, \alpha_{T-1}, \cdots, \alpha_0\}$, $\Gamma_T = \{\gamma_T, \gamma_{T-1}, \cdots, \gamma_1\}$ and $\Lambda_T = \{\lambda_T, \lambda_{T-1}, \cdots, \lambda_0\}$.

We take two examples to show the above procedure; Example 1 (Linear Model) and Example 2 (Nonlinear Model).

Example 1 (Linear Model): Consider the linear model: $h_t(\alpha_t, \epsilon_t) = Z_t\alpha_t + \epsilon_t$ and $f_t(\alpha_{t-1}, \eta_t) = T_t\alpha_{t-1} + \eta_t$, where $\epsilon_t \sim N(0, \gamma_t\Sigma_\epsilon)$ and $\eta_t \sim N(0, \lambda_t\Sigma_\eta)$. Z_t, T_t, Σ_η and Σ_ϵ are exogenous.[10] Moreover, assume that the prior densities are $\gamma_t \sim \text{IG}\left(\frac{1}{2}, \frac{p}{2}\right)$ and $\lambda_t \sim \text{IG}\left(\frac{1}{2}, \frac{q}{2}\right)$, where $\text{IG}(\cdot,\cdot)$ denotes the inverse (reciprocal) gamma distribution.[11] Then, the following posterior densities can be derived from the Bayesian rule.

$$P_{\gamma_*}(\gamma_t|y_t, \alpha_t) = \text{IG}\left(\frac{(y_t - Z_t\alpha_t)'\Sigma_\epsilon^{-1}(y_t - Z_t\alpha_t) + 1}{2}, \frac{1+p}{2}\right), \tag{23}$$

[9] Compare each estimate of $g_{r|s} \equiv E\big(g(\alpha_r)|Y_s\big)$, i.e., equations (13), (19) and (21). We need the density functions or the weight functions for (13) and (19) but not for (21). The important point is to generate random draws from the three posterior densities.

[10] In Carlin, Polson and Stoffer (1992), the prior densities of Σ_η and Σ_ϵ are also assumed and the posterior densities are derived by the Bayesian rule. In this paper, however, we assume for simplicity of discussion that Σ_η and Σ_ϵ are known and fixed.

[11] Let x be a random variable from a gamma distribution with parameters (a,b). That is, the density function of x is given by: $f(x) = \frac{a^b}{\Gamma(b)}e^{-ax}x^{b-1}$. Define $z = 1/x$. Then, z follows an inverse gamma distribution with parameters (a,b), i.e., $z \sim \text{IG}(a,b)$. In general, when $z \sim \text{IG}(a,b)$, $2a/z$ follows a Chi-square distribution with $2b$ degrees of freedom. Accordingly, $\gamma_t \sim \text{IG}\left(\frac{1}{2}, \frac{p}{2}\right)$ implies that $1/\gamma_t$ is a Chi-square random variable with p degrees of freedom.

$$P_{\lambda_*}(\lambda_t|\alpha_t,\alpha_{t-1}) = \mathrm{IG}\left(\frac{(\alpha_t - T_t\alpha_{t-1})'\Sigma_\eta^{-1}(\alpha_t - T_t\alpha_{t-1}) + 1}{2}, \frac{1+q}{2}\right), \tag{24}$$

$$P_{\alpha_*}(\alpha_t|y_t,\alpha_{t-1},\alpha_{t+1},\gamma_t,\lambda_t,\lambda_{t+1}) = N(B_t b_t, B_t), \tag{25}$$

where

$$B_t^{-1} = \frac{\Sigma_\eta^{-1}}{\lambda_t} + \frac{T_{t+1}'\Sigma_\eta^{-1}T_{t+1}}{\lambda_{t+1}} + \frac{Z_t'\Sigma_\epsilon^{-1}Z_t}{\gamma_t}, \qquad b_t = \frac{\alpha_{t-1}'T_t'\Sigma_\eta^{-1}}{\lambda_t} + \frac{\alpha_{t+1}'\Sigma_\eta^{-1}T_{t+1}}{\lambda_{t+1}} + \frac{y_t'\Sigma_\epsilon^{-1}Z_t}{\gamma_t}.$$

It is shown that, in the case of $\gamma_t = \gamma$ and $\lambda_t = \lambda$, the posterior densities $P_{\gamma_*}(\gamma_t|\cdot)$ and $P_{\lambda_*}(\lambda_t|\cdot)$ depend on Y_T for smoothing and Y_t for filtering and prediction.

Example 2 (Nonlinear Model): Consider the following nonlinear model: $h_t(\alpha_t,\epsilon_t) = h_{1,t}(\alpha_t) + \epsilon_t$ and $f_t(\alpha_{t-1},\eta_t) = f_{1,t}(\alpha_{t-1}) + \eta_t$. For the error terms and the nuisance parameters, the same assumptions as Example 1 are used. Then, Example 1 is modified as follows. We replace $Z_t\alpha_t$ in equation (23) by $h_{1,t}(\alpha_t)$, $T_t\alpha_{t-1}$ in equation (24) by $f_{1,t}(\alpha_{t-1})$ and $P_{\alpha_*}(\alpha_t|\cdot) = N(B_t b_t, B_t)$ in equation (25) by $P_{\alpha_*}(\alpha_t|\cdot) \propto w_1(\alpha_t)w_2(\alpha_t)N\big(f_{1,t}(\alpha_{t-1}),\lambda_t\Sigma_\epsilon\big)$, where

$$\begin{aligned} w_1(\alpha_t) &= \exp\left(-\frac{1}{2\lambda_{t+1}}\big(\alpha_{t+1} - f_{1,t+1}(\alpha_t)\big)'\Sigma_\eta^{-1}\big(\alpha_{t+1} - f_{1,t+1}(\alpha_t)\big)\right), \\ w_2(\alpha_t) &= \exp\left(-\frac{1}{2\gamma_t}\big(y_t - h_{1,t}(\alpha_t)\big)'\Sigma_\epsilon^{-1}\big(y_t - h_{1,t}(\alpha_t)\big)\right). \end{aligned}$$

Note that we have $0 \le w_1(\alpha_t)w_2(\alpha_t) \le 1$. The distribution which we want to sample is dominated by the $N\big(f_{1,t}(\alpha_{t-1}),\lambda_t\Sigma_\epsilon\big)$ density. Therefore, we may use rejection sampling to obtain a random observation from $P_{\alpha_*}(\alpha_t|\cdot)$.[12] That is, we sample a $N\big(f_{1,t}(\alpha_{t-1}),\lambda_t\Sigma_\epsilon\big)$ random variable and accept it with probability

[12] 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 $f(x) \le cg(x)$ 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 $0 \le w(x) \le 1$. 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 $u \le w(v)$, and return to (i) otherwise.

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

$$\mathrm{Prob}\big(X \le x | u \le w(v)\big) = \frac{\mathrm{Prob}\big(X \le x, u \le w(v)\big)}{\mathrm{Prob}\big(u \le w(v)\big)},$$

where the numerator and the denominator are represented as:

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

$$\begin{aligned} \mathrm{Prob}\big(u \le w(v)\big) &= \mathrm{Prob}\big(X \le \infty, u \le w(v)\big) \\ &= F(\infty)/c \\ &= 1/c. \end{aligned}$$

Therefore, we have:

$$\mathrm{Prob}\big(X \le x | u \le w(v)\big) = F(x).$$

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

$w_1(\alpha_t)w_2(\alpha_t)$.

For the NI and the MIIF procedures, we have to compute $P_\alpha(\alpha_t|\alpha_{t-1})$ and $P_y(y_t|\alpha_t)$ by hand and choose location of the nodes for NI and the random draws from the importance density for MII. The MIG approach requires the random draws only, using the Gibbs sampler, where we do not need *ad hoc* assumptions such as choice of nodes or importance density.

The computational disadvantage of the above procedure is a great amount of 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 in order to obtain the unconditional random draws, i.e., $\{\alpha_{i,t}^{(J+1)}, \gamma_{i,t}^{(J+1)}, \lambda_{i,t}^{(J+1)}\}$ for large J. Also, a computational amount of order $J \times T \times n$ is required, which is quite large. Especially, the random number generation by rejection sampling sometimes takes a lot of time computationally, which is also one of the disadvantages.[13]

3 Method of Monte-Carlo Stochastic Simulations

3.1 Monte-Carlo Approach: Type 1 (MA1)

In this section, we propose a new nonlinear and nonnormal filter using a Monte-Carlo simulation technique. The method of Monte-Carlo stochastic simulations is utilized as an alternative solution to the nonlinear and nonnormal state-space model. Prediction, filtering and smoothing algorithms are introduced. 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 and nonnormal procedure proposed here does not requires such assumptions to be taken by a researcher.

First, note that the joint density function of (Y_s, A_T) is written as $P(Y_s, A_T) = P(Y_s|A_s)P(A_T)$, where $P(Y_s|A_s) = \prod_{t=1}^{s} P_y(y_t|\alpha_t)$ and $P(A_T) = P_\alpha(\alpha_0)\prod_{t=1}^{T} P_\alpha(\alpha_t|\alpha_{t-1})$. $P_y(y_t|\alpha_t)$ and $P_\alpha(\alpha_t|\alpha_{t-1})$ are obtained from the measurement and transition equations. Therefore, the conditional density of α_r given information Y_s is given by:

$$P(\alpha_r|Y_s) = \frac{\int P(Y_s|A_s)P(A_T)\mathrm{d}A_{T,-r}}{\int P(Y_s|A_s)P(A_T)\mathrm{d}A_T}, \tag{26}$$

where $A_{T,-r} = \{\alpha_T, \cdots, \alpha_{r+1}, \alpha_{r-1}, \cdots, \alpha_0\}$, which is the set that α_r is excluded from A_t.[14] The expectation of $g(\alpha_r)$ given Y_s, $g_{r|s} \equiv E\big(g(\alpha_r)|Y_s\big)$, is obtained as follows.

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

which is equivalent to equation (7).

Let us define $\overline{\overline{g}}_{r|s}$ as the MA1 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 $P(A_T) \equiv P_\alpha(\alpha_0)\prod_{t=1}^{T} P_\alpha(\alpha_t|\alpha_{t-1})$. That is, $A_{i,s}$ is a set of random draws from the transition equation $\alpha_{i,t} = f_t(\alpha_{i,t-1}, \eta_{i,t})$, where $\eta_{i,t}$ is a random number generated from the underlying density.

[13] The random number generator by rejection sampling is known to be inefficient if $w_1(\alpha_t)w_2(\alpha_t)$ is close to zero. See, for example, Carlin and Polson (1991) and Carlin, Polson and Stoffer (1992).

[14] Note that $A_{t,-t} \equiv A_{t-1}$.

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

$$\overline{\overline{g}}_{r|s} = \frac{\frac{1}{n}\sum_{i=1}^{n} g(\alpha_{i,r}) P(Y_s|A_{i,s})}{\frac{1}{n}\sum_{j=1}^{n} P(Y_s|A_{j,s})}. \tag{27}$$

In this paper, $\overline{\overline{g}}_{r|s}$ is called the MA1 prediction (MA1P) when $(r, s) = (t + L, t)$ the MA1 filtering (MA1F) when $(r, s) = (t, t)$ and the MA1 smoothing (MA1S) when $(r, s) = (t, T)$, respectively.

The following formula is easy and convenient to compute the estimate of $g_{r|s}$ (i.e., $\overline{\overline{g}}_{r|s}$). Since $P(Y_s|A_{i,s}) = \prod_{t=1}^{s} P_y(y_t|\alpha_{i,t})$, equation (27) is rewritten as:

$$\overline{\overline{g}}_{r|s} = \sum_{i=1}^{n} g(\alpha_{i,r})\omega_{i,s}, \tag{28}$$

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

$$\omega_{i,s} = \frac{\prod_{t=1}^{s} P_y(y_t|\alpha_{i,t})}{\sum_{j=1}^{n}\prod_{t=1}^{s} P_y(y_t|\alpha_{j,t})} = \frac{P_y(y_s|\alpha_{i,s})\omega_{i,s-1}}{\sum_{j=1}^{n} P_y(y_s|\alpha_{j,s})\omega_{j,s-1}}. \tag{29}$$

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

$$\omega_{i,0} = \frac{1}{n}, \tag{30}$$

for $i = 1, \cdots, n$. Thus, $\overline{\overline{g}}_{r|s}$ denotes the MA1 estimate of $g(\alpha_r)$ given information Y_s,[15] where the Monte-Carlo simulation technique is utilized based on the 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 $1 = \sum_{i=1}^{n} \omega_{i,s}$.

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

[15] Take an example of L-step ahead prediction. The L-step ahead prediction $g_{t+L|t}$ is represented as:

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

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

density for MIG. Note that the MA1 procedure does not yield the recursive algorithm, which is different from NI and MII approaches and similar to MIG approach.

The computation procedure of the new proposal is 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 (30), use equation (29) to obtain the weight functions for $i = 1, \cdots, n$ and $t = 1, \cdots, T$, and (iv) $\overline{\overline{g}}_{r|s}$ can be computed by equation (28).

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 we have to compute $P_\alpha(\alpha_t|\alpha_{t-1})$ and $P_y(y_t|\alpha_t)$ by hand, which implies much easier programming of the new proposal than NI, MII and MIG approaches. Moreover, the proposed procedure is more attractive with respect to computational time than the other procedures.

The unknown parameters can be estimated by maximum likelihood estimation. The density function of Y_T is given by:

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

Generating random draws of A_T, the likelihood function $P(Y_T)$ is evaluated as follows:

$$P(Y_T) = \frac{1}{n}\sum_{i=1}^{n} P(Y_T|A_{i,T}), \tag{31}$$

which is maximized with respect to the unknown parameters when they are included in the system.

Under some 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 MA1P, MA1F and MA2S proposed above. That is, $\Sigma_{r|s}$ is defined as:

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

and the sample variance of $\overline{\overline{\alpha}}_{r|s}$ as $\overline{\overline{\Sigma}}_{r|s}$, i.e.,

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

Then, as n goes to infinity, the following two theorems can be shown.

(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 but convergence is slow as $\sqrt{n}$.

Finally, note that, from equation (29), it might be expected that the simulation errors increase as t is large. Accordingly, when T is large, the MA1 procedure is intractable.

3.2 Monte-Carlo Approach: Type 2 (MA2)

We propose another nonlinear and nonnormal filter using a Monte-Carlo approach. For a solution to nonlinear and nonnormal state-space model, in this section, we use the random draws 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.

MA2 Prediction (MA2P): The MA2 prediction (MA2P) estimate is simple, which is obtained 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

$$P(\alpha_{t+L}|Y_t) \approx \frac{1}{n}\sum_{i=1}^{n} P_\alpha(\alpha_{t+L}|\alpha_{i,t+L-1|t}). \tag{32}$$

We can generate random draws of α_{t+L} given Y_t as follows. Pick up $\alpha_{i,t+L-1|t}$ randomly (i.e., pick up i) and generate random numbers of η_{t+L}, and we have random draws of $\alpha_{t+L|t}$ from the transition equation $\alpha_{t+L|t} = f_{t+L}(\alpha_{i,t+L-1|t}, \eta_{t+L})$, i.e., $\alpha_{i,t+L|t}$ for $i = 1, \cdots, n$.

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.

MA2 Filtering (MA2F): The MA2 filtering (MA2F) estimate is derived as follows. Suppose that the random draws $\alpha_{i,t-1|t-1}$ are available. We consider generating random numbers of α_t from the filtering density $P(\alpha_t|Y_t)$. The filtering density $P(\alpha_t|Y_t)$ given by equation (5) is represented as:

$$\begin{aligned}
P(\alpha_t|Y_t) &= \frac{\int P_y(y_t|\alpha_t)P_\alpha(\alpha_t|\alpha_{t-1})P(\alpha_{t-1}|Y_{t-1})\mathrm{d}\alpha_{t-1}}{\int P_y(y_t|\alpha_t)P_\alpha(\alpha_t|Y_{t-1})\mathrm{d}\alpha_t} \\
&= \frac{1}{P(y_t|Y_{t-1})}\int P_y(y_t|\alpha_t)P_\alpha(\alpha_t|\alpha_{t-1})P(\alpha_{t-1}|Y_{t-1})\mathrm{d}\alpha_{t-1} \\
&\approx \frac{1}{P(y_t|Y_{t-1})}\frac{1}{n}\sum_{i=1}^{n} P_y(y_t|\alpha_t)P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}) \\
&\propto \frac{1}{n}\sum_{i=1}^{n} P_y(y_t|\alpha_t)P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}),
\end{aligned} \tag{33}$$

where $P(y_t|Y_{t-1}) = \int P_y(y_t|\alpha_t)P_\alpha(\alpha_t|Y_{t-1})\mathrm{d}\alpha_t$. Thus, the filtering density $P(\alpha_t|Y_t)$ is approximately proportional to $\frac{1}{n}\sum_{i=1}^{n} P_y(y_t|\alpha_t)P_\alpha(\alpha_t|\alpha_{i,t-1|t-1})$.

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 to obtain a random observation from $P(\alpha_t|Y_t)$. In general, the distribution which we want to sample is dominated by the $P_\alpha(\alpha_t|\alpha_{i,t-1|t-1})$ density. Therefore, for $t = 1, \cdots, T$, we choose $\alpha_{i,t-1|t-1}$ with probability $1/n$ (i.e., we choose i), sample the $P_\alpha(\alpha_t|\alpha_{i,t-1|t-1})$ random variable and accept it with probability $w_1(\alpha_t; y_t)$, 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 random numbers generated from $P(\alpha_t|Y_t) \propto \frac{1}{n}\sum_{i=1}^{n} P_y(y_t|\alpha_t)P_\alpha(\alpha_t|\alpha_{i,t-1|t-1})$ are obtained

MA2 Smoothing (MA2S): Finally, the MA2 smoothing (MA2S) 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 MA2F procedure shown above.

In order to obtain the MA2S estimate, note that each component in equation (6) is transformed as follows:

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

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

$P(\alpha_{t+1}|Y_t)$ 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:

$$\begin{aligned} P(\alpha_t|Y_T) &\approx \frac{1}{P(y_t|Y_{t-1})} \left(\frac{1}{n} \sum_{i=1}^{n} P_y(y_t|\alpha_t) P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}) \right) \left(\frac{1}{n} \sum_{j=1}^{n} \frac{P_\alpha(\alpha_{j,t+1|T}|\alpha_t)}{P(\alpha_{j,t+1|T}|Y_t)} \right) \\ &\propto \left(\frac{1}{n} \sum_{i=1}^{n} P_y(y_t|\alpha_t) P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}) \right) \left(\frac{1}{n} \sum_{j=1}^{n} \frac{P_\alpha(\alpha_{j,t+1|T}|\alpha_t)}{P(\alpha_{j,t+1|T}|Y_t)} \right) \\ &\propto \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{q_{j,t}} P_y(y_t|\alpha_t) P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}) P_\alpha(\alpha_{j,t+1|T}|\alpha_t), \end{aligned} \tag{34}$$

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

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

(ii) $$\sum_{j=1}^{n} 1/q_{j,t} = 1.$$

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 $\frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{1}{q_{j,t}}$ $P_y(y_t|\alpha_t) P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}) P_\alpha(\alpha_{j,t+1|T}|\alpha_t)$. Therefore, based on $P_y(y_t|\alpha_t) P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}) P_\alpha(\alpha_{j,t+1|T}|\alpha_t)$,

we can generate the random numbers of α_t from $P(\alpha_t|Y_T)$ for $t = T-1, T-2, \cdots, 1$, which is the 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 $w_1(\alpha_t; y_t)$ and $w_2(\alpha_t; \alpha_{j,t+1|T})$ satisfy:

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

Thus, the random draw of α_t given information Y_T can be generated.[16]

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 $t = T-1, T-2, \cdots, 1$, which 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 $0 \leq w_1(\alpha_t; y_t)w_2(\alpha_t; \alpha_{j,t+1|T}) \leq 1$, $P_y(y_t|\alpha_t) \propto w_1(\alpha_t; y_t)$, $P_\alpha(\alpha_{j,t+1|T}|\alpha_t) \propto w_2(\alpha_t; \alpha_{j,t+1|T})$, and $\alpha_{i,T|T}$ is the random draw from the filtering density. Thus, the random draw of α_t given information Y_T can be generated.

The computational procedure is as follows. From the density functions of ϵ_t and η_t, we have

$$
\begin{aligned}
&P(\alpha_t|Y_t) \propto \frac{1}{n}\sum_{i=1}^{n} w_1(\alpha_t; y_t)P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}) \quad \text{for the filtering density, and} \\
&P(\alpha_t|Y_T) \propto \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{1}{q_{j,t}} w_1(\alpha_t; y_t)w_2(\alpha_t; \alpha_{j,t+1|T})P(\alpha_t|\alpha_{i,t-1|t-1}) \quad \text{for the smoothing density,}
\end{aligned}
$$

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.

Thus, the random draws $\alpha_{i,r|s}$ for $(r, s) = (t+L, t), (t, t), (t, T)$ are generated. Let $g^*_{r|s}$ be the MA2 estimate of $g_{r|s} \equiv E\big(g(\alpha_r)|Y_s\big)$.

$$
g^*_{r|s} = \frac{1}{n}\sum_{i=1}^{n} g(\alpha_{i,r|s}), \tag{35}
$$

which is called the Monte-Carlo predictor (MA2P) when $(r, s) = (t+L, t)$, the Monte-Carlo filter (MA2F) when $(r, s) = (t, t)$ and the Monte-Carlo smoother (MA2S) when $(r, s) = (t, T)$.

[16] In equation (34), 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:

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

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.

The likelihood function is given by

$$P(Y_T) = \prod_{t=1}^{T} \frac{1}{n} \sum_{i=1}^{n} P_y(y_t|\alpha_{i,t|t-1}). \tag{36}$$

Note that we have

$$\int P_y(y_t|\alpha_t)P(\alpha_t|Y_{t-1})\mathrm{d}\alpha_t = \frac{1}{n} \sum_{i=1}^{n} P_y(y_t|\alpha_{i,t|t-1}).$$

The features of the above procedure are: (i) we do not need the 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 generated from the posterior density $P(\alpha_r|Y_s)$, and (iii) programming is very easy, compared with the other nonlinear filters by numerical integration (Kitagawa (1987) and Kramer and Sorenson (1988)), Monte-Carlo integration with importance sampling (Tanizaki (1993), Tanizaki and Mariano (1994) and Mariano and Tanizaki (1995)), Monte-Carlo integration with Gibbs sampling (Carlin, Polson and Stoffer (1992)) and Monte-Carlo stochastic simulations.

We show two examples to implement the above procedure.

Example 1 (Linear Model):

$$\begin{aligned} &\text{Measurement Equation:} && y_t = Z_t\alpha_t + \epsilon_t, \\ &\text{Transition Equation:} && \alpha_t = T_t\alpha_{t-1} + \eta_t, \\ & && \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H_t & 0 \\ 0 & Q_t \end{pmatrix} \right), \end{aligned} \tag{37}$$

where Z_t, T_t, H_t and Q_t are assumed to be known.

For prediction, we pick up $\alpha_{i,t|t}$ 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 up $\alpha_{i,t-1|t-1}$ 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., $\frac{1}{n}\sum_{i=1}^{n} N(\Sigma_t\mu_{i,t}, \Sigma_t)$, where $\mu_{i,t} = Q_t^{-1}T_t\alpha_{i,t-1|t-1} + Z_t'H_t^{-1}y_t$ and $\Sigma_t^{-1} = Z_t'H_t^{-1}Z_t + Q_t^{-1}$. Note that $\alpha_{i,t-1|t-1}$ is chosen at random. Thus, the random number of α_t given Y_t is obtained from $\frac{1}{n}\sum_{i=1}^{n} N(\Sigma_t\mu_{i,t}, \Sigma_t)$.

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 $P(\alpha_t|Y_T)$ is represented as average of the normal densities with mean $\Sigma_t^*\mu_{ij,t}^*$ and variance Σ_t^*, i.e., $\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{1}{q_{j,t}} N(\Sigma_t^*\mu_{ij,t}^*, \Sigma_t^*)$, where $\mu_{ij,t}^* = Q_t^{-1}T_t\alpha_{i,t-1|t-1} + Z_t'H_t^{-1}y_t + \alpha_{j,t+1|T}Q_{t+1}^{-1}T_{t+1}$ and $\Sigma_t^{*-1} = Z_t'H_t^{-1}Z_t + Q_t^{-1} + T_{t+1}'Q_{t+1}^{-1}T_{t+1}$. $\alpha_{i,t-1|t-1}$ and $\alpha_{j,t+1|T}$ are picked up with probability $1/n$ and probability $1/q_{j,t}$, respectively, and the random number of α_t given Y_T is obtained. The endpoint case is the same as filtering.

Example 2 (Nonlinear Model): In the case where the transition equation is nonlinear and the measurement equation is linear, we have the following state-space model:

$$\begin{aligned} &\text{Measurement Equation:} && y_t = Z_t\alpha_t + \epsilon_t, \\ &\text{Transition Equation:} && \alpha_t = f_{1,t}(\alpha_{t-1}) + \eta_t, \\ & && \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H_t & 0 \\ 0 & Q_t \end{pmatrix} \right), \end{aligned} \tag{38}$$

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

For prediction, given $\alpha_{i,t|t}$, generating normal random draws of η_{t+L} for $L = 1, 2, \cdots$, we have the random draw of α_{t+L} given Y_t recursively, using $N\big(f_{1,t+L}(\alpha_{i,t+L-1|t}), Q_{t+L}\big)$ 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 average of the normal densities with mean $\Sigma_t\mu_{i,t}$ and variance Σ_t, i.e., $\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} N(\Sigma_t\mu_{i,t}, \Sigma_t)$, where $\mu_{i,t} = Q_t^{-1} f_{1,t}(\alpha_{i,t-1|t-1}) + Z_t' H_t^{-1} y_t$ and $\Sigma_t^{-1} = Z_t' H_t^{-1} Z_t + Q_t^{-1}$. Picking up $\alpha_{i,t-1|t-1}$ at random, the random number of 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 (34) cannot be obtained explicitly. Therefore, the random numbers from the posterior density are generated utilizing rejection sampling. We pick up $\alpha_{i,t-1|t-1}$ 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

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

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

In the case where the transition equation is linear and the measurement equation is nonlinear, we have the following state-space model:

$$\begin{aligned} &\text{Measurement Equation:} && y_t = h_{1,t}(\alpha_t) + \epsilon_t, \\ &\text{Transition Equation:} && \alpha_t = T_t\alpha_{t-1} + \eta_t, \\ & && \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H_t & 0 \\ 0 & Q_t \end{pmatrix}\right), \end{aligned} \tag{39}$$

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

For prediction, given $\alpha_{i,t|t}$, generating normal random draws of η_{t+L} for $L = 1, 2, \cdots$, we have the random number of α_{t+L} given Y_t recursively, using $N\big(T_{t+L}\alpha_{i,t+L-1|t}, Q_{t+L}\big)$ 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 $P_{\alpha_*}(\alpha_t|y_t, \alpha_{t-1})$ cannot be obtained explicitly. Therefore, we may apply rejection sampling. We pick up $\alpha_{i,t-1|t-1}$ randomly. Then, we 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 $w_1(\alpha_t; y_t) = \exp\left(-\dfrac{1}{2}\big(y_t - h_{1,t}(\alpha_t)\big)' H_t^{-1}\big(y_t - h_{1,t}(\alpha_t)\big)\right)$.

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 $P_{\alpha_*}(\alpha_t|y_t, \alpha_{t-1}, \alpha_{t+1})$ cannot be obtained explicitly. Therefore, the random numbers from the posterior density (34) are generated utilizing rejection sampling. We generate a random draw from the normal density with mean $\Sigma_t^*\mu_t^*$ and variance Σ_t^*, i.e., $N(\Sigma_t^*\mu_t^*, \Sigma_t^*)$, where $\mu_t^* = Q_t^{-1}T_t\alpha_{t-1} + \alpha_{t+1}Q_{t+1}^{-1}T_{t+1}$ and $\Sigma_t^{*-1} = Q_t^{-1} + T_{t+1}'Q_{t+1}^{-1}T_{t+1}$. Picking up $\alpha_{i,t-1|t-1}$ and $\alpha_{j,t+1|T}$ at random, the random number of α_t given Y_T is obtained 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 $w_1(\alpha_t; y_t)$. The endpoint case is the same as filtering.

In the case where both the transition equation and the measurement equation are nonlinear, we have the following state-space model:

$$\begin{aligned} &\text{Measurement Equation:} && y_t = h_{1,t}(\alpha_t) + \epsilon_t, \\ &\text{Transition Equation:} && \alpha_t = f_{1,t}(\alpha_{t-1}) + \eta_t, \\ & && \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H_t & 0 \\ 0 & Q_t \end{pmatrix}\right), \end{aligned} \tag{40}$$

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

For prediction, given $\alpha_{i,t|t}$, generating normal random draws of η_{t+L} for $L = 1, 2, \cdots$, we have the random draw of α_{t+L} given Y_t recursively, using $N\big(f_{1,t+L}(\alpha_{i,t+L-1|t}), Q_{t+L}\big)$ 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 $P_{\alpha_*}(\alpha_t|y_t, \alpha_{t-1})$ cannot be obtained explicitly. Therefore, we may apply rejection sampling. We pick up $\alpha_{i,t-1|t-1}$ randomly. Then, we generate a normal random draw $N\left(f_{1,t}(\alpha_{t-1}), Q_t\right)$ 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 $P_{\alpha_*}(\alpha_t|y_t, \alpha_{t-1}, \alpha_{t+1})$ cannot be obtained explicitly. Therefore, the random numbers from the posterior density are generated utilizing rejection sampling. We pick up $\alpha_{i,t-1|t-1}$ and $\alpha_{j,t+1|T}$ randomly. Then, we generate a random draw from $N\left(f_{1,t}(\alpha_{i,t-1|t-1}), Q_t\right)$ and accept it with probability $w_1(\alpha_t; y_t)w_2(\alpha_t; \alpha_{j,t+1|T})$. The endpoint case is the same as filtering.

4 Numerical Examples: Monte-Carlo Experiments

In this section, we compare the estimates shown in Table 1.

Table 1: Nonlinear and Nonnormal State-Space Modeling

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
$\overline{\overline{g}}_{r\|s}$	$\overline{\overline{\alpha}}_{r\|s}$	MA1P	MA1F	MA1S	Section 3.1
$g^*_{r\|s}$	$\alpha^*_{r\|s}$	MA2P	MA2F	MA2S	Section 3.2

(1): Note that $\alpha_{r|s} \equiv E(\alpha_r|Y_s)$.

(2): Denote $a_{r|s}$ by the Kalman filter estimate or the extended Kalman filter estimate.

We perform four Monte-Carlo experiments: Section 4.1 Linear and Normal Model, Section 4.2 Nonlinear Model, and Section 4.3 Nonlinear Model.

4.1 Linear and Normal Model

Consider the following linear scalar system:

$$\begin{aligned}
&\text{Measurement Equation:} && y_t = \alpha_t + \epsilon_t, \\
&\text{Transition Equation:} && \alpha_t = \delta\alpha_{t-1} + \eta_t, \\
& && \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right), \\
& && \alpha_0 \sim N(0, 1),
\end{aligned} \tag{41}$$

where $t = 1, \cdots, T$. We compare the following estimates: the Kalman (K) estimate $a_{r|s}$[17] , the NI estimate $\widetilde{\alpha}_{r|s}$[18] , the MII estimate $\widehat{\alpha}_{r|s}$[19] , the MIG estimate $\overline{\alpha}_{r|s}$[20] , and the MA1 estimate $\overline{\overline{\alpha}}_{r|s}$. Note that the Kalman filter estimate $a_{t|t}$ gives us the optimal solution in this example because the system (41) is a linear and normal model.

The simulation procedure is as follows:

(i) Generating standard normal random numbers of 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 NIP, NIF, NIS, MIIP, MIIF and MIIS, $n = 250$ for MIGP, MIGF and MIGS, $n = 2500$ for MA1P, MA1F and MA1S, and $n = 1000$ for MA2P, MA2F and MA2S in order to take an equal 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, which are defined as:

$$\text{BIAS}(\alpha_{r|s}) = \frac{1}{G}\sum_{g=1}^{G}(\alpha_{r|s}^{(g)} - \alpha_r^{(g)}), \qquad \text{RMSE}(\alpha_{r|s}) = \left(\frac{1}{G}\sum_{g=1}^{G}(\alpha_{r|s}^{(g)} - \alpha_r^{(g)})^2\right)^{1/2},$$

where $\alpha_{r|s}$ takes $a_{r|s}$, $\widetilde{\alpha}_{r|s}$, $\widehat{\alpha}_{r|s}$, $\overline{\alpha}_{r|s}$ or $\overline{\overline{\alpha}}_{r|s}$. The superscript (g) denotes the g-th simulation run, and we take $G = 1000$. That is, $\alpha_r^{(g)}$ denotes the simulated state variable at time r in the g-th simulation run.

It is expected in this experiment that K is better than any other estimator because the system is linear and normal.

[17] 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.

[18] 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).

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

[20] We assume that the nuisance parameters are distributed as $\gamma_t \sim \text{IG}\left(\frac{1}{2}, \frac{1}{2}\right)$ and $\lambda_t \sim \text{IG}\left(\frac{1}{2}, \frac{1}{2}\right)$. The assumptions on the nuisance parameters are remained for all the experiments from Section 4.1 to Section 4.3. We take $J = 10$.

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	0.0105	0.0045	0.0012
	MA2	0.0086	0.0047	0.0004
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	1.6105	1.1480	1.3319
	MA2	1.3004	0.8175	0.6939

The results are in Table 2a. NI is very close to K. MIG and MA1 are poor estimators, which diverge from the true value as t increases. K, NI and MII are slightly better than MA2 with respect to BIAS and RMSE but not too different.

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	0.425	0.250	0.260	0.090	0.290	0.460	0.610	0.710
	MA2	0.374	0.214	0.248	0.095	0.240	0.410	0.540	0.620
$\delta = 1.0$	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	0.999	0.042	0.042	0.980	0.990	1.000	1.010	1.025
	MA2	0.963	0.106	0.113	0.830	0.920	0.980	1.030	1.060

Table 2b shows the estimation results of unknown parameter δ. Since MIG gives us the poor results in Table 2a and a computational burden of MIG is too large, hereafter we do not consider MIG. Also, the result of MA1 is not reliable from Table 2a. Therefore, we do not pay too much attension to MA1.

MA2 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 MA2 is biased, compared with that of K, NI and MII.

4.2 Nonlinear Model I

Consider the following nonlinear scalar system:

$$\begin{aligned} &\text{Measurement Equation:} && y_t = \alpha_t + \epsilon_t, \\ &\text{Transition Equation:} && \alpha_t = (1 - \delta + \delta\alpha_{t-1}^2)^{1/2}\eta_t, \\ & && \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right), \\ & && \alpha_0 \sim N(0,1), \end{aligned} \tag{42}$$

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[21] , while the measurement equation consists of the ARCH(1) term and the error. The measurement equation is linear but the transition equation is nonlinear. In this experiment, therefore, it might be expected that K does not show a good performance.

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

We compare precision of the estimated state variables for K, NI, MII, MA1 and MA2. The results are in Table 3a. NI, MII and MA2 perform better than the other estimators, while NI, MII and MA2 give us the similar results (Especially, as it is expected, NI, MII and MA2 are better than K). It is natural that K is not good, because for K the transition equation is linearized with respect to α_{t-1} and η_t. Similarily, MA1 is the worst estimator, which diverges from the true value as t increases.

[21] In this Monte-Carlo experiment, the unconditional variance of α_t is assumed to be one.

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	0.703	0.299	0.362	0.100	0.595	0.820	0.920	0.970
	MA2	0.465	0.350	0.351	0.010	0.075	0.500	0.800	0.930
$\delta = 0.9$	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	0.900	0.148	0.148	0.770	0.880	0.950	0.990	0.990
	MA2	0.805	0.258	0.275	0.440	0.730	0.910	0.990	0.990

In Table 3b, the estimates of δ are compared for K, NI, MII, MA1 and MA2. Each likelihood function is maximized by a simple grid search, where we take $\delta = 0.00, 0.01, 0.02, \cdots, 0.99$. The estimates by MA2 is the closest to the true value in the bias criterion, except for MA1 (we do not consider MA1 because it is not reliable from the results in Table 3a). In the case of $\delta = 0.9$, the estimates by MA2 give us the closest to the true value in average and the smallest RMSE.

4.3 Nonlinear Model II

Consider the following nonlinear scalar system:

$$\begin{aligned} &\text{Measurement Equation:} \quad y_t = \frac{1}{1+\exp(-\alpha_t + \epsilon_t)}, \\ &\text{Transition Equation:} \quad \alpha_t = \frac{1}{1+\exp(-\alpha_{t-1} + \eta_t)}, \\ &\qquad \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right), \\ &\qquad \alpha_0 \sim N(0,1), \end{aligned} \tag{43}$$

where $t = 1, \cdots, T$. In this experiment, both the measurement and the transition equations are nonlinear[22] .

[22] The measuremant equation can be transformed into a linear function in the state variable, which is shown as:

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

For K, equation (43), 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	−0.0009	−0.0009	−0.0009
	MA2	−0.0016	−0.0016	−0.0012
RMSE	K	0.2041	0.2135	0.2151
	NI	0.2017	0.1974	0.1972
	MII	0.2035	0.1992	0.1990
	MA1	0.2018	0.1975	0.1974
	MA2	0.2017	0.1975	0.1972

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

5 Summary

In this paper, the two Monte-Carlo procedures using the simulation technique are proposed; (i) one utilizes the density function of y_t given α_t and the random draws from the transition equation, and (ii) another requires the random draws only. Both procedures improve the other nonlinear filters developed in the past from the following three points; computational time, simplicity of computer programming and no *ad hoc* assumptions.

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

The problems of the MII procedure developed by Tanizaki (1993), Tanizaki and Mariano (1994) and Mariano and Tanizaki (1995) are: (i) the importance density $P_I(\alpha_t)$ has to be appropriately chosen by a researcher and (ii) we need to derive the densities $P_\alpha(\alpha_t|\alpha_{t-1})$ and $P_y(y_t|\alpha_t)$ by hand, which is similar to the NI approach.

The MIG approach (Carlin, Polson and Stoffer (1992)) also has some problems: (i) we need to assume the distributions of nuisance parameters which is also an ad hoc assumption, (ii) use of the Gibbs sampler leads to a great amount of data storage, and (iii) the MIG also takes a lot of time computationally, i.e., $T \times n \times J$.

The MA1 approach improves some of these problems, i.e., we have to compute $P_y(y_t|\alpha_t)$ by hand but not $P_\alpha(\alpha_t|\alpha_{t-1})$, and we do not need *ad hoc* assumptions such as choice of the nodes for the NI procedure and choice of the importance density for the MII approach. 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.

As an alternative simple procedure, we have proposed the MA2 estimator. By random draws only, we evaluate estimates of prediction, filtering and smoothing. For the random number generation, rejection sampling may be adopted, which is a method of the random number generation from any distribution. MA2 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 3.2.)

The Monte-Carlo experiments have shown the result that MA2 is very close to NI and MII. In addition to simplicity of computer programming and no *ad hoc* assumptions, precision of the estimates is also close to NI procedure. Accordingly, the MA2 approach proposed in this paper might be recommended for nonlinear and nonnormal 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.

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 H. Tanizaki (1995), " Prediction of Final Data with Use of Preliminary and/or Revised Data, " forthcoming in *Journal of Forecasting*.

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.

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

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.

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.