On markov chain monte carlo methods for nonlinear and non-gaussian state-space models

Geweke, John

Tanizaki, Hisashi

ON MARKOV CHAIN MONTE CARLO METHODS FOR NONLINEAR AND NON-GAUSSIAN STATE-SPACE MODELS

JOHN GEWEKE

University of Minnesota and
Federal Reserve Bank of Minneapolis
Department of Economics
1035 Management and Economics
271 – 19th Ave. S.
Minneapolis, MN 55455, USA
(geweke@bayes.econ.umn.edu)

HISASHI TANIZAKI

Kobe University
Faculty of Economics
Rokkodai, Nadaku
Kobe 657-8501, Japan
(tanizaki@kobe-u.ac.jp)

Key Words: Nonlinear and Non-Gaussian State-Space Models, Filtering, Smoothing, Markov Chain Monte Carlo, Gibbs Sampling, Metropolis-Hastings Algorithm.

ABSTRACT

In this paper, a nonlinear and/or non-Gaussian smoother utilizing Markov chain Monte Carlo Methods is proposed, where the measurement and transition equations are specified in any general formulation and the error terms in the state-space model are not necessarily normal. The random draws are directly generated from the smoothing densities. For random number generation, the Metropolis-Hastings algorithm and the Gibbs sampling technique are utilized. The proposed procedure is very simple and easy for programming, compared with the existing nonlinear and non-Gaussian smoothing techniques. Moreover, taking several candidates of the proposal density function, we examine precision of the proposed estimator.

1 INTRODUCTION

By recent progress of computer, computer-intensive techniques have been developed, which have been in turn applied to various research fields. In this paper, some of the computer-intensive techniques are used to resolve a nonlinear and non-Gaussian smoothing problem. There are several kinds of sampling techniques, i.e., importance sampling, rejection sampling, Gibbs sampling, the Metropolis-Hastings algorithm and etc. They are classified into two groups. One is called the independence Monte Carlo and another is known as the Markov chain Monte Carlo. Both importance sampling and rejection sampling techniques are included in the independence Monte Carlo approach while Gibbs sampling (see Appendix 1.1) and the Metropolis-Hastings algorithm (see Appendix 1.2) belong to the Markov chain Monte Carlo. See Geweke (1988, 1989a, 1989b, 1996, 1997), Shao (1989) and Koop (1994) for importance sampling and Knuth (1981), Boswell, Gore, Patil and Taillie (1993), O'Hagan (1994), Chib and Greenberg (1995) and Geweke (1996, 1997) for rejection sampling,

Numerous papers deal with the nonlinear and non-Gaussian filters and smoothers. For example, see Gelb (1974), Anderson and Moore (1979), Harvey (1989), Kitagawa and Gersch (1996) and Tanizaki (1996) for nonlinear and non-Gaussian state space modeling. Recently, the sampling techniques above have been applied to the nonlinear and non-Gaussian filters and smoothers, which are classified into two approaches; one is based on the recursive algorithms and another utilizes the Markov chain Monte Carlo methods. Tanizaki and Mariano (1994), Mariano and Tanizaki (1995) and Tanizaki (1996, 1999b) applied Monte Carlo integration with importance sampling to derive nonlinear and non-Gaussian state-space models. Tanizaki (1996, 1999a), Mariano and Tanizaki (1999) and Tanizaki and Mariano (1998) utilized rejection sampling to generate random draws directly from filtering densities. Gordon, Salmond and Smith (1993) and Kitagawa (1996), Kitagawa and Gersch (1996), Tanizaki (1997) and Tanizaki and Mariano (1996) also obtained filtering means by random draws.

As for nonlinear and non-Gaussian filters and smoothers with Markov chain Monte Carlo, Carlin, Polson and Stoffer (1992) and Carter and Kohn (1994, 1996), Chib and Greenberg (1996) and De Jong and Shephard (1995) introduced the nonlinear and/or non-Gaussian state-space models with Gibbs sampling. They investigated the nonlinear state-space models in the Bayesian framework, where the nuisance parameters introduced in the state-space model are assumed to be stochastic. The state-space models that they used are quite restricted to some functional forms, because they studied the special state-space models such that it is easy to generate random draws from the underlying assumptions. For example, Carter and Kohn (1996) deals with the conditionally Gaussian state-space models and Den Jong and Shephard (1995)

approximates a non-Gaussian disturbance by a mixture of normals. Also, see Durbin and Koopman (1997), Pitt and Shephard (1997) and Shepard and Pitt (1997). In this paper, we propose the nonlinear and non-Gaussian smoother using Gibbs sampling and the Metropolis-Hastings algorithm, which would be suitable to any nonlinear and non-Gaussian state-space model. It is needless to say that the proposal density is required for the Metropolis-Hastings algorithm. Some candidates of the proposal density are investigated through some Monte Carlo studies. Thus, in this paper we attempt to solve a nonlinear and non-Gaussian smoothing problem in a general setup.

2   STATE SPACE MODEL

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

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

for $t = 1, 2, \cdots, T$. A vector y_t is observable while a vector α_t is unobserved. The error terms ϵ_t and η_t are mutually independently distributed, which are typically assumed to be normal but not necessarily in this paper. $h_t(\cdot, \cdot)$ and $f_t(\cdot, \cdot)$ are vector functions, which are assumed to be known. Let Y_s be the information set up to time s, i.e., $Y_s = \{y_1, y_2, \cdots, y_s\}$. We consider estimating the conditional expectation of α_t using information Y_s, i.e., $\mathrm{E}(\alpha_t|Y_s) \equiv \alpha_{t|s}$. Depending on t and s, the conditional expectation $\alpha_{t|s}$ is called prediction if $t > s$, filtering if $t = s$ and smoothing if $t < s$. Moreover, there are three kinds of smoothing by the relationship between t and s with $t < s$. Let k and T be the fixed nonnegative integer and the sample size, respectively. $\alpha_{k|t}$ for fixed k and $t = k+1, k+2, \cdots, T$ is called fixed-point smoothing, which is useful to estimate the initial condition of the system. $\alpha_{t|t+k}$ for fixed k and $t = 1, 2, \cdots, T - k$ is known as fixed-lag smoothing. $\alpha_{t|T}$ for $t = 1, 2, \cdots, T$ is called fixed-interval smoothing, which is helpful to investigate the past condition of the system. In this paper, we focus only on fixed-interval smoothing, i.e., $\alpha_{t|T}$.

2.1   FIXED-INTERVAL SMOOTHING

Define A_s by $A_s = \{\alpha_0, \alpha_1, \cdots, \alpha_s\}$, which is a set consisting of the state-variables up to time s. Let $p_y(y_t|\alpha_t)$ and $p_\alpha(\alpha_t|\alpha_{t-1})$ be the density functions obtained from measurement equation (1) and transition equation (2). Moreover, let $p(A_s, Y_s)$, $p_\alpha(A_s)$, $p_y(Y_s|A_s)$ and $p(A_s|Y_s)$ be the joint density of A_s and Y_s, the density of A_s, the conditional density of Y_s given A_s and the conditional density of A_s given Y_s, respectively.

We derive the conditional density function $p(A_T|Y_T)$ to obtain the fixed-interval smoothing mean, i.e., $\alpha_{t|T}$. First, note that $p(A_T, Y_T)$ is written as:

$$p(A_T, Y_T) = p_\alpha(A_T)p_y(Y_T|A_T), \tag{3}$$

where $p_\alpha(A_T)$ and $p_y(Y_T|A_T)$ are represented as:

$$p_\alpha(A_T) = \begin{cases} p_\alpha(\alpha_0) \prod_{s=1}^{T} p_\alpha(\alpha_s|\alpha_{s-1}), & \text{if } \alpha_0 \text{ is stochastic,} \\ \prod_{s=1}^{T} p_\alpha(\alpha_s|\alpha_{s-1}), & \text{otherwise,} \end{cases} \tag{4}$$

$$p_y(Y_T|A_T) = \prod_{s=1}^{T} p_y(y_s|\alpha_s). \tag{5}$$

Accordingly, from the Bayes theorem, the posterior density $p(A_T|Y_T)$ is obtained as follows:

$$\begin{aligned} p(A_T|Y_T) &= \frac{p(A_T, Y_T)}{\int p(A_T, Y_T)\mathrm{d}A_T} \\ &= \frac{p_\alpha(A_T)p_y(Y_T|A_T)}{\int p_\alpha(A_T)p_y(Y_T|A_T)\mathrm{d}A_T}. \end{aligned} \tag{6}$$

Therefore, evaluation of a function $g(\alpha_t)$ is obtained as:

$$\begin{aligned} \mathrm{E}(g(\alpha_t)|Y_T) &= \int g(\alpha_t)p(A_T|Y_T)\mathrm{d}A_T \\ &= \frac{\int g(\alpha_t)p_\alpha(A_T)p_y(Y_T|A_T)\mathrm{d}A_T}{\int p_\alpha(A_T)p_y(Y_T|A_T)\mathrm{d}A_T}, \end{aligned} \tag{7}$$

for $t = 1, 2, \cdots, T$. $g(\cdot)$ is a function, which is typically $g(\alpha_t) = \alpha_t$ or $g(\alpha_t) = (\alpha_t - \alpha_{t|T})(\alpha_t - \alpha_{t|T})'$.

Usually, the filtering and smoothing formulas yield the recursive algorithms of the density functions and the smoothing density requires the filtering density and the one-step ahead prediction density (see Appendix 2). However, the approach introduced in this section is not the standard smoothing algorithm because it is not a recursive algorithm (see Tanizaki (1996, 1997)). That is, $p(\alpha_t|Y_T)$ is obtained as follows. Let us define $A_s^* = \{\alpha_s, \alpha_{s+1}, \cdots, \alpha_T\}$. Based on equation (6), the smoothing density $p(\alpha_t|Y_T)$ is given by:

$$\begin{aligned} p(\alpha_t|Y_T) &= \iint p(A_T|Y_T)\mathrm{d}A_{t-1}\mathrm{d}A_{t+1}^* \\ &= \frac{\iint p(A_T, Y_T)\mathrm{d}A_{t-1}\mathrm{d}A_{t+1}^*}{\int p(A_T, Y_T)\mathrm{d}A_T} \\ &= \frac{\iint p_\alpha(A_T)p_y(Y_T|A_T)\mathrm{d}A_{t-1}\mathrm{d}A_{t+1}^*}{\int p_\alpha(A_T)p_y(Y_T|A_T)\mathrm{d}A_T}. \end{aligned} \tag{8}$$

4

In the case of the standard smoothing algorithm (Appendix 2), $p(\alpha_t|Y_T)$ is recursively obtained from $p(\alpha_{t+1}|Y_T)$ for $t = T-1, T-2, \cdots, 1$. Equation (8) seems to be different from the conventional recursive smoothing algorithm (16) in Appendix 2. However, as shown in Tanizaki (1999c), equation (16) can be easily derived from equation (8).

Tanizaki (1997) made an attempt to evaluate the smoothing estimates, generating random draws of A_T from $p_\alpha(A_T)$, where $p_y(Y_T|A_T)$ is not utilized to generate the random draws and accordingly the obtained estimates sometimes show a poor performance. In this paper, we consider generating random draws of A_T, using both $p_\alpha(A_T)$ and $p_y(Y_T|A_T)$, in order to improve precision of the smoothing estimates.

We have discussed the smoothing technique. In the case of filtering, simply replace T by t in equations (3) – (7). The filtering mean $E(g(\alpha_t)|Y_t)$ is computed from equation (7), replacing T by t. Accordingly, filtering is more computer-intensive than smoothing. In the standard density-based smoothing algorithm, $p(\alpha_t|Y_{t-1})$ and $p(\alpha_t|Y_t)$ are required. After $p(\alpha_t|Y_{t-1})$ and $p(\alpha_t|Y_t)$ are computed for $t = 1, 2, \cdots, T$, $p(\alpha_t|Y_T)$ are obtained by the backward recursive algorithm. See Appendix 2 for the standard algorithms. Thus, clearly smoothing is more computer-intensive than filtering in the conventional density-based recursive algorithm. However, according to the Markov chain Monte Carlo procedure, it is easier to compute smoothing, rather than filtering.

2.2   LIKELIHOOD FUNCTION

In the case where unknown parameters are included in equations (1) and (2), the likelihood function to be maximized with respect to the unknown parameters is written as:

$$
\begin{aligned}
p(Y_T) &= \int p(A_T, Y_T)\mathrm{d}A_T \\
&= \int p_\alpha(A_T)p_y(Y_T|A_T)\mathrm{d}A_T.
\end{aligned}
$$
(9)

An alternative estimation method of the unknown parameters is known as the EM (Expectation - Maximization) algorithm, where the expected log-likelihood function is maximized, given all the observed data Y_T. See Dempster, Laird and Rubin (1977) for the EM algorithm. Accordingly, the expected log-likelihood function to be maximized is given by:

$$
\begin{aligned}
&\mathrm{E}\Big(\log(p(A_T, Y_T))|Y_T\Big) \\
&= \mathrm{E}\Big(\log(p_\alpha(A_T)p_y(Y_T|A_T))|Y_T\Big) \\
&= \int \log(p_\alpha(A_T)p_y(Y_T|A_T))p(A_T|Y_T)\mathrm{d}A_T.
\end{aligned}
$$
(10)

Shumway and Stoffer (1982) applied the EM algorithm to the state-space model in linear and normal case. For the procedure proposed in this paper, it is much easier to utilize the EM algorithm, rather than maximization of equation (9). As for the features of the EM algorithm, it is known that the convergence speed is very slow but it quickly searches the neighborhood of the true parameter value.

3   MONTE CARLO SMOOTHER

In this paper, using both the Metropolis-Hastings algorithm and Gibbs sampling, an attempt is made to generate random draws of A_T directly from $p(A_T|Y_T)$, shown in equation (6). The smoothing means are evaluated as the arithmetic averages of the random draws. Define A_s^* as a set of the state-variables after time s, i.e., $A_s^* = \{\alpha_s, \alpha_{s+1}, \cdots, \alpha_T\}$. Note that A_0^* implies A_T and A_T is equivalent to $\{A_s, A_{s+1}^*\}$. According to the Gibbs sampler, random draws of A_T from $p(A_T|Y_T)$ are based on those of α_s from $p(\alpha_s|A_{s-1}, A_{s+1}^*, Y_T)$ for $s = 1, 2, \cdots, T$. Therefore, to apply the Gibbs sampler, we need to derive the conditional density function of α_s given A_{s-1}, A_{s+1}^* and Y_T, which is represented as follows:

$$
\begin{aligned}
&p(\alpha_s|A_{s-1}, A_{s+1}^*, Y_T) \\
&= \frac{p(A_T|Y_T)}{p(A_{s-1}, A_{s+1}^*|Y_T)} \\
&= \frac{p_\alpha(A_T)p_y(Y_T|A_T)}{\displaystyle\int p_\alpha(A_T)p_y(Y_T|A_T)\mathrm{d}\alpha_s} \\
&= \begin{cases} \dfrac{p_\alpha(\alpha_s|\alpha_{s-1})p_\alpha(\alpha_{s+1}|\alpha_s)p_y(y_s|\alpha_s)}{\displaystyle\int p_\alpha(\alpha_s|\alpha_{s-1})p_\alpha(\alpha_{s+1}|\alpha_s)p_y(y_s|\alpha_s)\mathrm{d}\alpha_s}, & \text{if } s = 1, 2, \cdots, T-1, \\ \dfrac{p_\alpha(\alpha_s|\alpha_{s-1})p_y(y_s|\alpha_s)}{\displaystyle\int p_\alpha(\alpha_s|\alpha_{s-1})p_y(y_s|\alpha_s)\mathrm{d}\alpha_s}, & \text{if } s = T, \end{cases} \\
&\propto \begin{cases} p_\alpha(\alpha_s|\alpha_{s-1})p_\alpha(\alpha_{s+1}|\alpha_s)p_y(y_s|\alpha_s), & \text{if } s = 1, 2, \cdots, T-1, \\ p_\alpha(\alpha_s|\alpha_{s-1})p_y(y_s|\alpha_s), & \text{if } s = T. \end{cases} \quad (11)
\end{aligned}
$$

Here, equation (11) implies that a kernel of the conditional density function $p(\alpha_s|A_{s-1}, A_{s+1}^*, Y_T)$ is given by $p_\alpha(\alpha_s|\alpha_{s-1})p_\alpha(\alpha_{s+1}|\alpha_s)p_y(y_s|\alpha_s)$ when $s = 1$, $2, \cdots, T-1$ and $p_\alpha(\alpha_s|\alpha_{s-1})p_y(y_s|\alpha_s)$ when $s = T$ (i.e., endpoint). Note that equations (4) and (5) are used to obtain the third equality of equation (11).

Utilizing the density function $p(\alpha_s|A_{s-1}, A_{s+1}^*, Y_T)$, we consider generating random draws of A_T directly from $p(A_T|Y_T)$ by Gibbs sampling. Since it is generally intractable to generate random draws of α_s from $p(\alpha_s|A_{s-1},$

A^*_{s+1}, Y_T), the Metropolis-Hastings algorithm is applied. Let $\alpha_{i,s}$ be the i-th random draw of the state-vector at time s. Define $A_{i,s}$ and $A^*_{i,s}$ as $A_{i,s} = \{\alpha_{i,0}, \alpha_{i,1}, \cdots, \alpha_{i,s}\}$ and $A^*_{i,s} = \{\alpha_{i,s}, \alpha_{i,s+1}, \cdots, \alpha_{i,T}\}$, respectively, which are the i-th random draws of A_s and A^*_s. Let $p_*(z|x)$ be the proposal density (or the sampling density), which is the conditional distribution of z given x. For the proposal density $p_*(z|x)$, we should choose the density function such that random draws can be easily and quickly generated. Define the acceptance probability $\omega(x, z)$ as follows:

$$
\omega(x, z) = \begin{cases} \min\left(\dfrac{p(z|A_{i,s-1}, A^*_{i-1,s+1}, Y_T)p_*(x|z)}{p(x|A_{i,s-1}, A^*_{i-1,s+1}, Y_T)p_*(z|x)}, 1\right), \\ \qquad \text{if } p(x|A_{i,s-1}, A^*_{i-1,s+1}, Y_T)p_*(z|x) > 0, \\ 1, \qquad\qquad \text{otherwise.} \end{cases}
$$

Then, to generate random draws of A_T from $p(A_T|Y_T)$, the Metropolis-Hastings algorithm within Gibbs sampling is applied as follows:

(i) Take appropriate values for $\alpha_{1,0}$ and $\alpha_{0,s}$, $s = 1, 2, \cdots, T$. Typically, the smoothing estimates based on the extended Kalman filter are taken for $\alpha_{0,s}$, $s = 1, 2, \cdots, T$. $\alpha_{1,0}$ depends on the underlying assumption of the initial state α_0. That is, $\alpha_{i,0}$ is generated from $p(\alpha_0)$ if α_0 is stochastic and it is fixed as α_0 if α_0 is nonstochastic.

(ii) Given $A_{i,s-1}$ and $A^*_{i-1,s+1}$ (i.e., the other random draws except for a random number of α_s), we utilize the Metropolis-Hastings algorithm to generate a random draw of α_s, denoted by $\alpha_{i,s}$, from the conditional density $p(\alpha_s|A_{i,s-1}, A^*_{i-1,s+1}, Y_T)$, which takes the following procedure:

   (1) Generate z from $p_*(\cdot|\alpha_{i-1,s})$ and u from the uniform distribution between zero and one.

   (2) Compute $\omega(\alpha_{i-1,s}, z)$ and choose either (a) or (b).
      (a) If $u \le \omega(\alpha_{i-1,s}, z)$, set $\alpha_{i,s} = z$.
      (b) Otherwise, set $\alpha_{i,s} = \alpha_{i-1,s}$.

(iii) Repeat (ii) for $s = 1, 2, \cdots, T$.

(iv) Repeat (ii) and (iii) for $i = 1, 2, \cdots, N$.

Note in Step (i) that the extended Kalman filter is one of the traditional nonlinear filters, where the nonlinear measurement and transition equations given by equations (1) and (2) are linearized by the first-order Taylor series expansion and the linearized system is directly applied to the standard linear recursive algorithm (see Wishner, Tabaczynski and Athans (1969), Gelb (1974), Anderson and Moore (1979) and Tanizaki and Mariano (1996)). In Step (ii),

an alternative method of random number generation is rejection sampling. That is, rejection sampling is applied to generate a random draw of α_t from $p(\alpha_t|A_{i,t-1}, A^*_{i-1,t+1}, Y_T)$. Rejection sampling has the following disadvantages: (i) it takes a long time computationally when the acceptance probability is very small, and (ii) we cannot apply rejection sampling when the supremum of the target density divided by the proposal density does not exist. Therefore, in order to avoid the above problems, in this paper we utilize the Metropolis-Hastings algorithm in Step (ii).

Thus, both the Gibbs sampler and the Metropolis-Hastings algorithm are simultaneously used to obtain the random draws $A_{i,T}$, $i = 1, 2, \cdots, N$. Based on $A_{i,T}$, $\mathrm{E}(g(\alpha_s)|Y_T)$ is simply evaluated as the arithmetic average of $g(\alpha_{i,s})$, $i = 1, 2, \cdots, N$, which is represented by:

$$\frac{1}{N-M} \sum_{i=M+1}^{N} g(\alpha_{i,s}) \equiv \overline{g(\alpha_s)},$$

where $g(\cdot)$ is a function. Especially, the case of $g(\alpha_s) = \alpha_s$ represents the smoothing mean. We may take $M = 0$ but usually 10 – 20 % of N is taken for M. From the basic result of the Metropolis within Gibbs sampling, we can obtain the following result:

$$\overline{g(\alpha_s)} \longrightarrow \mathrm{E}(g(\alpha_s)|Y_T) \equiv \int g(\alpha_s) p(\alpha_s|Y_T) \mathrm{d}\alpha_s, \quad \text{as } N \longrightarrow \infty,$$

for all $s = 1, 2, \cdots, T$. In Monte Carlo studies of the next section, we take 20% of N for M.

Note that the filtering procedure is implemented as follows. Replace T by t in the above procedure (i) – (iv). Then the random draws from the filtering density $p(\alpha_t|Y_t)$ are given by $\alpha_{i,t}$, $i = 1, 2, \cdots, N$, where t denotes the endpoint in the procedure (i) – (iv). Furthermore, in addition to the above procedure in order to obtain filtering, we should put the following step: (v) Repeat (i) – (iv) for $t = 1, 2, \cdots, T$. As for computational burden, number of iteration is given by the order of NT for smoothing and that of $\sum_{t=1}^{T} Nt = NT(T-1)/2$ for filtering. Thus, filtering is more computer-intensive than smoothing in the proposed procedure. Therefore, in this paper, we focus only on fixed-interval smoothing.

In addition, we should keep in mind as follows. Compared with the independence Monte-Carlo methods such as importance sampling and rejection sampling, it is known that the Markov chain Monte-Carlo methods need a lot of random draws from the following reasons:

- In the Markov chain Monte-Carlo methods we usually discard the first 10% – 20% random draws.

- A random draw is positively correlated with the next random draw in general.

8

- Convergence of the Gibbs sampler is very slow especially in the case where there is high correlation between α_t and α_{t-1} (see Chib and Greenberg (1995)).

To implement the procedure (i) – (iv), we need to specify the proposal density $p_*(z|x)$. In Monte Carlo studies of Section 5, smoothing estimates are compared taking several different candidates of the proposal density function $p_*(z|x)$.

3.1 LIKELIHOOD FUNCTION

To evaluate the likelihood function shown in equation (9), we need to generate random draws of A_T from $p_\alpha(A_T)$. Since $p_\alpha(A_T)$ indicates the unconditional distribution function of A_T, it is not appropriate to use the random draws generated by the procedure proposed in (i) – (iv) (remember that the random draws generated by (i) – (iv) are from $p(A_T|Y_T)$). Accordingly, we have to generate the random draws of A_T from $p_\alpha(A_T)$ to evaluate the likelihood function (9), which implies that we need extra computation. To generate random draws of A_T from $p_\alpha(A_T)$, we utilize the transition equation (2) only. In Tanizaki (1997), the unconditional random draws of A_T are utilized to derive the filtering and smoothing estimates and the likelihood function.

In order to reduce the computational disadvantage, alternatively consider applying the EM algorithm, where the random draws of A_T from $p(A_T|Y_T)$ are utilized. Thus, for estimation of unknown parameters, the conditional expectation of the log-likelihood function given by equation (10) is maximized. Using the random draws from the smoothing density, equation (10) is rewritten as follows:

$$
\begin{aligned}
& \mathrm{E}\Big(\log(p(A_T, Y_T))|Y_T\Big) \\
= {} & \mathrm{E}\Big(\log(p_\alpha(A_T)p_y(Y_T|A_T))|Y_T\Big) \\
= {} & \int \log(p_\alpha(A_T)p_y(Y_T|A_T))p(A_T|Y_T)\mathrm{d}A_T \\
\approx {} & \frac{1}{N-M}\sum_{i=M+1}^{N} \log(p_\alpha(A_{i,T})p_y(Y_T|A_{i,T})).
\end{aligned}
\tag{12}
$$

Thus, for the proposed smoother, the EM algorithm is easily applied to estimation of the unknown parameters.

4 CHOICE OF PROPOSAL DENSITY

The Metropolis-Hastings algorithm has the problem of specifying the proposal density, which is the crucial criticism. Several generic choices of the proposal

density are discussed by Tierney (1994) and Chib and Greenberg (1995). In this paper, we take four candidates for the proposal density $p_*(z|x)$.

4.1 PROPOSAL DENSITY (A)

It might be natural for the proposal density to take the density function obtained from the transition equation (2), i.e., $p_*(z|x) = p_\alpha(z|\alpha_{i,s-1})$. In this case, $p_*(z|x)$ does not depend on x, i.e., $p_*(z|x) = p_*(z)$, which is called the independence chain.

4.2 PROPOSAL DENSITY (B)

It might be possible for the proposal density to utilize the extended Kalman smoothed estimates. That is, the proposal density is represented by $p_*(z|x) = N(a_{s|T}^*, c\Sigma_{s|T}^*)$, where $a_{s|T}^*$ and $\Sigma^* s|T$ denote the first- and the second-moments (i.e., mean and variance) based on the extended Kalman smoothed estimates at time s. It is possible to choose c based on the optimal acceptance rate (for example, see Gelman, Roberts and Gilks (1996) for the optimal acceptance rate). However, we need much more extra computation if c is based on the optimal acceptance rate. Since we want to see how the proposed smoother depends on c, $c = 1, 2, 4, 16$ is taken in Monte Carlo studies of the next section.

Note that this proposal density is also the independence chain because $p_*(z|x) = p_*(z)$.

4.3 PROPOSAL DENSITY (C)

We may take the proposal density $p_*(z|x) = p_*(z - x)$, which is called the random walk chain. In this paper, we consider $p_*(z|x) = N(x, c\Sigma_{s|T}^*)$, where $c = 1, 2, 4, 16$ is examined in the next section.

4.4 PROPOSAL DENSITY (D)

The alternative candidate for the proposal density is taken in this paper, which is based on approximation of the log-kernel. Let $q(z) = \log(p(z))$, where $p(z)$ may denote a kernel of the target density, i.e.,

$$p(z) = \begin{cases} p_\alpha(z|\alpha_{i,s-1})p_\alpha(\alpha_{i-1,s+1}|z)p_y(y_s|z), & \text{for } s = 1, 2, \cdots, T-1, \\ p_\alpha(z|\alpha_{i,s-1})p_y(y_s|z), & \text{for } s = T \text{ (i.e., endpoint).} \end{cases}$$

First, we take a univariate case of z for simplicity of discussion. Consider approximating the log-kernel $q(z)$ around x by the second-order Taylor series expansion, which is given by:

$$q(z) \approx q(x) + q'(x)(z - x) + \frac{1}{2}q''(x)(z - x)^2, \tag{13}$$

where x corresponds to $\alpha_{i-1,s}$ in the procedure (i) – (iv). $q'(\cdot)$ and $q''(\cdot)$ denote the first- and the second-derivatives. Depending on the signs of $q'(x)$ and $q''(x)$, we have the following cases, i.e., Cases 1 – 4. Let $p_{*i}(z|x)$. be the proposal density $p_*(z|x)$ in Case i.

Case 1: $q''(x) < 0$: Equation (13) is written by the following quadratic form:

$$q(z) \approx q(x) - \frac{1}{2}\left(-q''(x)\right)\left(z - (x - \frac{q'(x)}{q''(x)})\right)^2 + d(x),$$

where $d(x)$ is an appropriate function of x. In the above equation, the second term which depends on z is equivalent to the exponential part of the normal density because of $q''(x) < 0$. Therefore, in Case 1, the proposal density $p_{*1}(z|x)$ is taken as the following normal distribution with mean $x - q'(x)/q''(x)$ and variance $(-q''(x))^{-1}$.

Case 2: $q''(x) \geq 0$ and $q'(x) < 0$: Perform linear approximation of $q(z)$ in this case. Let x_1^* be the smaller value than the nearest mode with $x_1^* < x$ (for choice of x_1^*, we will discuss later). Then, $q(z)$ is approximated by a line passing between x_1^* and x, which is written as follows:

$$q(z) \approx q(x_1^*) + \frac{q(x_1^*) - q(x)}{x_1^* - x}(z - x_1^*).$$

From the second term of the above equation, the proposal density is represented as the following distribution with $x_1^* < z$:

$$p_{*2}(z|x) = \begin{cases} \lambda \exp(-\lambda(z - x_1^*)), & \text{if } x_1^* < z, \\ 0, & \text{otherwise}, \end{cases}$$

where λ is given by:

$$\lambda = \left|\frac{q(x_1^*) - q(x)}{x_1^* - x}\right|.$$

Let the above exponential density be $f_2(z|x)$. Thus, z is generated by $z = w + x_1^*$, where w denotes a random variable from the exponential distribution with parameter λ.

Case 3: $q''(x) \geq 0$ and $q'(x) > 0$: Similarly, perform linear approximation of $q(z)$ in this case. Let x_2^* be the larger value than the nearest mode

11

with $x < x_2^*$ (for choice of x_1^*, we will discuss later). Approximation of $q(z)$ is exactly equivalent to that of Case 2, which is given by:

$$p_{*3}(z|x) = \begin{cases} \lambda \exp(-\lambda(x_2^* - z)), & \text{if } z < x_2^*, \\ 0, & \text{otherwise.} \end{cases}$$

z is represented by $z = x_2^* - w$, where w denotes a random variable from the exponential distribution with parameter λ.

Case 4: $q''(x) \geq 0$ and $q'(x) = 0$: In this case, we approximate $q(z)$ as a uniform distribution at the neighborhood of x. As for the range of the uniform distribution, we utilize the two values x_1^* and x_2^*, which satisfies $x_1^* < x < x_2^*$. Thus, in Case 4, the proposal density $p_{*4}(z|x)$ is obtained by the uniform distribution on the interval between x_1^* and x_2^*:

Thus, for the univariate random variable of z, all the possible cases are given by Cases 1 – 4.

Remarks:

(i) x_1^* and x_2^* should be set by a researcher. Variance of the exponential distribution with parameter λ is given by $1/\lambda^2$. Utilizing the variance, we take the following values for x_1^* and x_2^*:

$$x_1^* = \text{ (the nearest mode which is smaller than } x) \ - \frac{1}{\lambda},$$
$$x_2^* = \text{ (the nearest mode which is larger than } x) \ + \frac{1}{\lambda},$$

where $1/\lambda$ implies the standard error.

(ii) The above approximation of the target density $p(z)$ shown in Cases 1 – 4 leads to the following decomposition of the proposal density:

$$p_*(z|x) = p_{*1}(z|x)I_1(x) + p_{*2}(z|x)I_2(x) + p_{*3}(z|x)I_3(x) + p_{*4}(z|x)I_4(x),$$

where $I_i(x)$, $i = 1, 2, 3, 4$, are the following indicator functions:

$$I_1(x) = \begin{cases} 1, & \text{if } q''(x) < 0 \ \text{ (i.e., if } x \text{ is in Case 1),} \\ 0, & \text{otherwise,} \end{cases}$$

$$I_2(x) = \begin{cases} 1, & \text{if } q''(x) \geq 0 \text{ and } q'(x) < 0 \ \text{ (i.e., if } x \text{ is in Case 2),} \\ 0, & \text{otherwise,} \end{cases}$$

$$I_3(x) = \begin{cases} 1, & \text{if } q''(x) \geq 0 \text{ and } q'(x) > 0 \ \text{ (i.e., if } x \text{ is in Case 3),} \\ 0, & \text{otherwise,} \end{cases}$$

$$I_4(x) = \begin{cases} 1, & \text{if } q''(x) \geq 0 \text{ and } q'(x) = 0 \quad \text{(i.e., if } x \text{ is in Case 4)}, \\ 0, & \text{otherwise}. \end{cases}$$

$p_{*1}(z|x)$ is normal, $p_{*2}(z|x)$ and $p_{*3}(z|x)$ are based on an exponential distribution and $p_{*4}(z|x)$ is uniform, where the subscript denotes the corresponding case. Suppose that the previous random draw x is in one of Cases $1 - 4$. The random draw z based on x may be in either of Cases $1 - 4$. That is, the generated random draw may move from one case to another, which implies that the irreducibility condition of the Gibbs sampler is guaranteed.

(iii) The procedure discussed above can be extended to multivariate cases. Now, suppose that z is a $k \times 1$ vector. Let $H(z)$ be the second derivative of $q(z)$ with respect to z. $H(z)$ is rewritten as $H(z) = C\Lambda C'$, where Λ is diagonal and $CC' = C'C = I$. Define $v = C'z$. Let $p_v(z)$ be the kernel of the probability density function of v, which is rewritten as: $p_v(v) = p(Cv)$. Define $q_v(v) = \log(p_v(v))$. Then, we have $q_v(v) = q(Cv)$ and note that

$$\frac{\partial q_v(z)}{\partial v} = C'\frac{\partial q(z)}{\partial z},$$

$$\frac{\partial^2 q_v(z)}{\partial v \partial v'} = C'\frac{\partial^2 q(z)}{\partial z \partial z'}C = C'H(z)C = \Lambda.$$

Construct approximations to $p_v(v_j)$, $j = 1, 2, \cdots, k$, separately using the univariate method above. Generate $v_1, v_2, \cdots, v_k$ independently from these distributions. Finally, we use $z = Cv$ to obtain a random draw of z.

5 MONTE CARLO STUDIES

In this section, we compare the extended Kalman smoother and the proposed smoother taking the several proposal densities. The simulation procedure is as follows:

(i) Generating random numbers of ϵ_t and η_t for $t = 1, 2, \cdots, T$ based on the underlying assumptions, we obtain a set of data y_t and α_t, $t = 1, 2, \cdots, T$, from equations (1) and (2), where $T = 20, 40, 100$ is taken.

(ii) Given Y_T, perform the extended Kalman smoother and the proposed smoother, where we take $(N, M) = (1000, 200), (5000, 1000)$.

(iii) Repeat (i) and (ii) G times and compare the root mean square error (RMSE) for each estimator, which is defined as: $\text{RMSE} = (1/T) \sum_{s=1}^{T} \text{MSE}_{s|T}^{1/2}$, where $\text{MSE}_{s|T} = (1/G) \sum_{g=1}^{G} (\overline{\alpha}_{s|T}^{(g)} - \alpha_s^{(g)})^2$ and $\overline{\alpha}_{s|T}$ takes the extended Kalman smoothed estimate or the proposed nonlinear smoothed

estimate while α_s denotes the artificially simulated state-variable in Simulation Procedure (i). The superscript (g) denotes the g-th simulation run and we take $G = 1000$.

Under the above setup, we examine the following state-space models.

Simulation 1: Linear and Normal Model (Tables 1 and 5): Consider the linear and normal system: $y_t = \alpha_t + \epsilon_t$ and $\alpha_t = \delta\alpha_{t-1} + \eta_t$, where $\delta = 0.5, 0.9, 1.0$ is taken. The initial value α_0 and the error terms ϵ_t and η_t, $t = 1, 2, \cdots, T$, are assumed to be: $\alpha_0 \sim N(0, 1)$ and $(\epsilon_t, \eta_t)' \sim N(0, I_2)$, where I_2 denotes a 2×2 identity matrix. In Simulations 2 and 3. we consider the same assumptions as in Simulation 1.

Simulation 2: Stochastic Volatility Model (Tables 2 and 5): We take an example of the following univariate system: $y_t = \exp\left(0.5\alpha_t\right)\epsilon_t$ and $\alpha_t = \delta\alpha_{t-1} + \eta_t$, where $\delta = 0.5, 0.9$ is taken. This is called the stochastic volatility model, which is described in De Jong and Shephard (1995) and Ghysels, Harvey and Renault(1996).

Simulation 3: ARCH(1) Model (Table 3): Next, we examine the following state-space model: $y_t = \alpha_t + \epsilon_t$ and $\alpha_t = (1 - \delta + \delta\alpha_{t-1}^2)^{1/2}\eta_t$, where $\delta = 0.5, 0.9$ is taken. Note that in Simulation 2 the unconditional variance of α_t is assumed to be one. This is called the ARCH (Autoregressive Conditional Heteroscedasticity) model (see Engle (1982), Harvey (1989) and Harvey and Streibel (1998)).

Simulation 4: Markov Switching Model (Table 4): Finally, we consider the following k-variate state-space model: $y_t = x_t\alpha_t + \epsilon_t$ and $\alpha_{i,t} = \delta_{i,1}^{\alpha_{i,t-1}}(1-\delta_{i,2})^{1-\alpha_{i,t-1}} + \eta_{i,t}$ for $i = 1, 2, \cdots, k$, which is very close to the Markov switching model (see Hamilton (1994)). x_t and α_t are k-dimensional vectors, i.e., $x_t = (x_{1,t}, x_{2,t}, \cdots, x_{k,t})$ and $\alpha_t = (\alpha_{1,t}, \alpha_{2,t}, \cdots, \alpha_{k,t})'$. The state variable $\alpha_{i,t}$ represents the state variable which takes a binary number, i.e., zero or one. $x_{i,t}$, $i = 1, 2, \cdots, k$, are the exogenous variables, which are generated by $x_{i,t} \sim N(0, 1)$, $i = 1, 2, \cdots, k$, in this simulation study. δ_{ij} for $i = 1, 2, \cdots, k$ and $j = 1, 2$ are the parameters, which are assumed to be known. The error ϵ_t is assumed to be distributed as $\epsilon_t \sim N(0, 1)$, while the error $\eta_{i,t}$ has a Bernoulli distribution with parameter $\delta_{i,1}^{\alpha_{i,t-1}}(1 - \delta_{i,2})^{1-\alpha_{i,t-1}}$ for $i = 1, 2, \cdots, k$. Thus, it is assumed from the transition equation that the present state depends on the past state. Each probability is represented as:

$$\text{Prob}(\alpha_{i,t} = 1|\alpha_{i,t-1} = 1) = \delta_{i,1}, \quad \text{Prob}(\alpha_{i,t} = 1|\alpha_{i,t-1} = 0) = 1 - \delta_{i,2},$$
$$\text{Prob}(\alpha_{i,t} = 0|\alpha_{i,t-1} = 1) = 1 - \delta_{i,1}, \quad \text{Prob}(\alpha_{i,t} = 0|\alpha_{i,t-1} = 0) = \delta_{i,2},$$

14

for $i = 1, 2, \cdots, k$. Thus, depending on $(\alpha_{1,t}, \alpha_{2,t}, \cdots, \alpha_{k,t})$, we have 2^k cases, i.e., 2^k states. In this simulation study, we examine the case $k = 2$.

The results are in Tables 1 – 4, where the parameter δ is assumed to be known. Since in Simulation 1 the system is linear and normal in this case, it is expected that the Kalman smoothed estimate is better than any other estimator. Note that the proposed procedure is the simulation-based estimator, which includes the simulation errors. Also, we can expect in Simulations 2 – 4 that the proposed nonlinear and non-Gaussian procedure obtains smaller RMSE's than the extended Kalman smoothed estimator as N goes to infinity.

Each number in the tables denotes RMSE, defined above. RMSE indicates a measure of precision of the corresponding estimator. KS indicates the Kalman smoothed estimator in Table 1 and the extended Kalman smoothed estimator in Tables 2 – 4. In each table, (A) – (D) correspond to the proposal densities discussed in Section 4. Proposal Density (A) utilizes the transition equation (2). Proposal Densities (B) and (C) are based on the extended Kalman smoothed estimates, where large c implies large variance in the proposal density. $c = 1, 2, 4, 16$ is taken in this paper. Proposal Density (D) is obtained by approximating a kernel of the target density function. The four kinds of the proposal densities are investigated in Simulations I – III but Proposal Density (A) is examined in Simulation IV.

The results obtained from Tables 1 – 4 are summarized as follows.

- For all the proposal densities (A) – (D), RMSE decreases as N increases.

- Proposal Density (A) shows a good performance for all the simulation studies, which implies that (A) is the best choice for any nonlinear and non-Gaussian state-space model.

- For Proposal Densities (B) and (C), the value of c such that RMSE is the smallest depends on the functional form of the state-space model. That is, in the case of Proposal Density (B), $c = 1, 2$ in Table 1, $c = 2, 4$ in Table 2 and $c = 4$ for $\delta = 0.5$ and $c = 16$ for $\delta = 0.9$ in Table 3 should be taken. In the case of Proposal Density (C), however, we should choose $c = 4$ in Table 1, $c = 1, 2, 4$ in Table 2, and $c = 16$ in Table 3, respectively.

- For Proposal Density (D) in Tables 1 and 3, $\delta = 0.5, 0.9$ shows a good performance. In Table 2, however, $\delta = 0.5$ is quite good while $\delta = 0.9$ is very poor. Note that in Simulation 1 the exact random draw of α_t can be generated from from $p(\alpha_t | A_{i,t-1}, A^*_{i-1,t+1}, Y_T)$ through Proposal Density (D), i.e., the target density $p(\alpha_t | A_{i,t-1}, A^*_{i-1,t+1}, Y_T)$ is proportional to the proposal density (D). In Simulations 2 and 3, large δ indicates that the unconditional distribution of the state-variable has fat tails, compared with normal distribution. Thus, the exponential approximation of

Table 1: Simulation 1: Linear and Normal Model

T	δ	N \ c	(A)	(B)				(C)				(D)	KS
				1	2	4	16	1	2	4	16		
	0.5	1000	.708	.706	.707	.707	.708	.709	.708	.708	.709	.706	.705
		5000	.706	.706	.706	.706	.706	.706	.706	.706	.706	.706	
20	0.9	1000	.688	.688	.689	.690	.691	.691	.691	.691	.691	.688	.686
		5000	.686	.687	.687	.688	.688	.688	.688	.688	.688	.688	
	1.0	1000	.678	.678	.679	.680	.681	.681	.681	.681	.682	.678	.675
		5000	.675	.677	.677	.677	.678	.678	.678	.678	.679	.678	
	0.5	1000	.711	.710	.711	.711	.712	.713	.712	.712	.713	.710	.710
		5000	.710	.710	.710	.710	.711	.711	.710	.710	.710	.710	
40	0.9	1000	.690	.689	.690	.691	.692	.692	.692	.692	.693	.689	.688
		5000	.688	.688	.689	.689	.689	.689	.689	.689	.690	.689	
	1.0	1000	.679	.678	.678	.679	.682	.681	.681	.681	.682	.678	.676
		5000	.677	.677	.678	.678	.678	.678	.678	.678	.678	.677	
	0.5	1000	.708	.706	.706	.707	.708	.709	.708	.708	.709	.706	.706
		5000	.706	.706	.706	.706	.706	.706	.706	.706	.706	.706	
100	0.9	1000	.684	.683	.684	.684	.686	.686	.685	.686	.688	.683	.682
		5000	.683	.683	.683	.683	.683	.683	.683	.683	.683	.683	
	1.0	1000	.673	.672	.672	.673	.675	.675	.674	.674	.676	.672	.670
		5000	.671	.671	.671	.671	.672	.672	.672	.672	.672	.671	

Table 2: Simulation 2: Stochastic Variance Model

T	δ	N \ c	(A)	(B)				(C)				(D)	KS
				1	2	4	16	1	2	4	16		
	0.5	1000	.895	.896	.895	.895	.899	.897	.897	.896	.899	.894	1.149
20		5000	.893	.894	.894	.894	.896	.895	.895	.895	.895	.894	
	0.9	1000	.935	.951	.946	.951	.959	.942	.942	.947	.957	1.031	2.091
		5000	.932	.942	.936	.937	.941	.936	.936	.937	.938	1.004	
	0.5	1000	.906	.908	.907	.907	.909	.909	.908	.908	.911	.906	1.158
40		5000	.904	.905	.905	.905	.906	.905	.905	.905	.906	.904	
	0.9	1000	.936	.952	.948	.949	.955	.941	.942	.945	.958	1.046	2.214
		5000	.932	.938	.936	.936	.939	.934	.935	.935	.938	1.024	
	0.5	1000	.908	.910	.909	.909	.912	.911	.911	.910	.913	.908	1.161
100		5000	.907	.907	.907	.907	.907	.907	.907	.907	.908	.907	
	0.9	1000	.937	.951	.948	.950	.965	.942	.944	.950	.962	1.046	2.266
		5000	.933	.937	.935	.935	.939	.934	.934	.935	.939	1.026	

Table 3: Simulation 3: ARCH(1) Model

T	δ	$N \setminus c$	(A)	(B)				(C)				(D)	KS
				1	2	4	16	1	2	4	16		
	0.5	1000	.680	.684	.680	.679	.680	.683	.681	.681	.681	.679	.702
20		5000	.679	.682	.679	.679	.679	.679	.679	.679	.679	.679	
	0.9	1000	.530	.610	.585	.563	.539	.545	.541	.535	.533	.534	.675
		5000	.527	.598	.572	.554	.536	.535	.534	.532	.531	.532	
	0.5	1000	.683	.685	.682	.682	.683	.685	.683	.683	.683	.681	.704
40		5000	.681	.683	.681	.681	.681	.681	.681	.681	.681	.681	
	0.9	1000	.522	.589	.564	.543	.526	.534	.528	.527	.522	.526	.650
		5000	.520	.580	.555	.538	.523	.526	.524	.522	.521	.521	
	0.5	1000	.683	.685	.682	.682	.682	.685	.683	.683	.683	.681	.706
100		5000	.681	.682	.681	.681	.681	.682	.681	.681	.681	.681	
	0.9	1000	.519	.582	.558	.539	.521	.530	.525	.522	.520	.523	.644
		5000	.517	.572	.548	.531	.519	.521	.520	.518	.517	.519	

Table 4: Simulation 4: Two-Variate Model ($k = 2$)

T	δ_{11}	δ_{12}	δ_{21}	δ_{22}	N	(A)		KS	
						$\alpha_{1,t}$	$\alpha_{2,t}$	$\alpha_{1,t}$	$\alpha_{2,t}$
	0.2	0.5	0.2	0.8	1000	.446	.444	.461	.489
					5000	.446	.440		
20	0.2	0.5	0.5	0.8	1000	.446	.449	.460	.460
					5000	.445	.448		
	0.2	0.8	0.5	0.8	1000	.445	.449	.487	.460
					5000	.443	.448		
	0.2	0.5	0.2	0.8	1000	.447	.436	.461	.476
					5000	.446	.434		
40	0.2	0.5	0.5	0.8	1000	.447	.445	.460	.457
					5000	.447	.445		
	0.2	0.8	0.5	0.8	1000	.439	.445	.479	.458
					5000	.437	.445		
	0.2	0.5	0.2	0.8	1000	.446	.435	.460	.473
					5000	.445	.433		
100	0.2	0.5	0.5	0.8	1000	.446	.446	.459	.458
					5000	.445	.445		
	0.2	0.8	0.5	0.8	1000	.436	.445	.474	.458
					5000	.434	.444		

the kernel does not work well, which is because the exponential approximation easily generates outliers from the approximation method of the kernel.

- In Table 4, we consider the two-variate cases, where a discrete type of state variable is taken and Proposal (A) is compared with KS. For both cases, the proposed nonlinear and nonnormal smoother with Proposal (A) performs better than KS in the sense of RMSE.

Accordingly, it might be easily concluded that the proposed nonlinear and non-Gaussian smoother performs well. For the proposal density, Proposal Density (A) might be recommended, because the other candidates of the proposal density are useful in the limited situations or they have *ad hoc* assumptions. That is, choice of c depends on trial and error for Proposal Densities (B) and (C), which implies that choice of c is quite *ad hoc.* Moreover, Proposal Density (D) is poor when the target distribution has fat tails.

Next, we consider estimating the unknown parameter δ, which is assumed to be fixed in Tables 1 – 3. In Table 5, a comparison between the true parameter and the estimate of δ is shown for KS and the proposed smoother, where $(N, M) = (5000, 1000)$ is taken and the two proposal densities (A) and (D) are examined. That is, given the artificially simulated data y_t for $t = 1, 2, \cdots, T$, the unknown parameter δ is estimated. Note that δ indicates the true value. AVE and RMSE represent the arithmetic average and the root mean square error obtained from the G estimates of δ. 25%, 50% and 75% give us the 0.25th, 0.50th and 0.75th quantiles from the G estimates. The maximization of the expected log-likelihood function (12) is performed by a simple grid search, in which the function is maximized by changing the parameter value of δ by 0.01. The results obtained from Table 5 are summarized as follows.

- For Proposal Density (A) of Simulation 1, AVE is close to the true parameter value for most of the cases and RMSE is the smallest of the three. For Proposal Density (D), both AVE and RMSE shows a good performance for $\delta = 0.9, 1.0$ but not for $\delta = 0.5$. We obtain the result that the parameter estimates are overestimated for (A) and (D) but underestimated for KS.

- — in Simulation 2 of KS indicates that the maximum likelihood estimation cannot be performed, because the innovation form of the likelihood function (see equation (17) in Appendix 2) does not depend on the unknown parameter δ under the assumption of $E(\alpha_0) = 0$. (A) is better than (D) with respect to AVE and RMSE. That is, AVE of (A) is closer to the true parameter value and (A) shows smaller RMSE than (D).

18

Table 5: Estimation of Unknown Parameter ($N = 5000$)

Simu-lation	δ		$T = 20$ (A)	(D)	KS	$T = 40$ (A)	(D)	KS	$T = 100$ (A)	(D)	KS
		AVE	0.531	0.510	0.359	0.561	0.648	0.428	0.600	0.722	0.472
	0.5	RMSE	0.085	0.458	0.353	0.102	0.318	0.239	0.129	0.257	0.144
		25%	0.500	0.450	0.170	0.510	0.610	0.300	0.530	0.690	0.400
		50%	0.510	0.680	0.420	0.530	0.720	0.470	0.590	0.740	0.490
		75%	0.550	0.810	0.610	0.605	0.810	0.610	0.660	0.790	0.570
		AVE	0.916	0.880	0.767	0.920	0.932	0.845	0.925	0.945	0.878
	0.9	RMSE	0.036	0.207	0.271	0.033	0.075	0.133	0.035	0.055	0.065
1		25%	0.900	0.840	0.670	0.900	0.910	0.800	0.910	0.930	0.850
		50%	0.910	0.930	0.840	0.910	0.950	0.880	0.920	0.950	0.890
		75%	0.920	0.980	0.930	0.930	0.970	0.925	0.940	0.970	0.920
		AVE	1.010	0.963	0.885	1.006	0.989	0.951	1.001	0.997	0.981
	1.0	RMSE	0.031	0.152	0.232	0.021	0.054	0.105	0.013	0.019	0.040
		25%	1.000	0.930	0.830	1.000	0.970	0.930	0.990	0.990	0.970
		50%	1.010	0.990	0.950	1.010	1.000	0.980	1.000	1.000	0.990
		75%	1.020	1.040	1.010	1.020	1.020	1.010	1.010	1.010	1.000
		AVE	0.545	0.671	—	0.574	0.746	—	0.638	0.829	—
	0.5	RMSE	0.093	0.254	—	0.116	0.297	—	0.170	0.341	—
		25%	0.500	0.510	—	0.510	0.565	—	0.550	0.790	—
		50%	0.510	0.610	—	0.540	0.790	—	0.630	0.850	—
2		75%	0.550	0.850	—	0.610	0.890	—	0.710	0.890	—
		AVE	0.917	0.926	—	0.922	0.946	—	0.931	0.959	—
	0.9	RMSE	0.032	0.077	—	0.035	0.067	—	0.040	0.064	—
		25%	0.900	0.900	—	0.900	0.920	—	0.910	0.950	—
		50%	0.910	0.940	—	0.910	0.960	—	0.930	0.960	—
		75%	0.930	0.990	—	0.940	0.980	—	0.950	0.980	—

Thus, we sometimes have the infeasible case for KS as in Simulation 2. In addition, even in the linear and normal case (i.e., Simulation 1), the estimate of δ is underestimated and less efficient for all $\delta = 0.5, 0.9, 1.0$ while (A) performs better than any other estimator.

6   SUMMARY

Using rejection sampling, Tanizaki (1996, 1999a), Tanizaki and Mariano (1998) and Mariano and Tanizaki (1999) proposed a nonlinear and non-Gaussian filter, where random draws are recursively generated from the filtering density at each time. Rejection sampling is very efficient to generate random draws from the target density function. However, it is well known that (i) rejection sampling takes a lot of time computationally when the acceptance probability is very small and (ii) it cannot be applied when the acceptance probability does not exist.

Carlin, Polson and Stoffer (1992), Carter and Kohn (1994, 1996) and De Jong and Shephard (1995) proposed the nonlinear and/or non-Gaussian state-space modeling with Gibbs sampler in a Bayesian framework. The state-space models which they investigated cannot be applied to all the general nonlinear and non-Gaussian state-space models. That is, they dealt with the state-space models such that we can easily generate random draws from the target density. In such a sense, their models are quite specific.

Improving these disadvantages of rejection sampling and Gibbs sampling, in this paper, we have proposed the nonlinear and non-Gaussian smoother which can be applied to any functional form of the measurement and transition equations in the state-space model. The proposed nonlinear and non-Gaussian smoother utilizes both the Metropolis-Hastings algorithm and the Gibbs sampler. The estimation procedure is given by (i) – (iv) in Section 3. In general, the problem of the Metropolis-Hastings algorithm is to choose the proposal density $p_*(z|x)$. By the Monte Carlo studies, the proposed smoother is compared taking several proposal density functions (i.e., Proposal Densities (A) – (D)) shown in Section 4. As a result, for all the simulation studies, Proposal Density (A) is the best assumption in the sense of RMSE, which is obtained from the transition equation (2). The reasons why Proposal Density (A) is the best are summarized as follows.

(i) In the case of Monte Carlo studies, the true state-space model is known. However, in the case where the underlying model is not known, note that (A) is not necessarily the best proposal density.

(ii) Proposal Densities (B) and (C) utilize the extended Kalman smoothed estimates and they sometimes show a good performance depending on c, where choice of c becomes another problem in practice.

(iii) Moreover, for Simulations 2 and 3, in the case where the extended Kalman smoothed estimates are away from the true values, the smoothing estimates obtained from (B) and (C) become poor.

(iv) Proposal Density (D) is good in almost all the cases but it is very poor when the target density $p(\alpha_s | A_{s-1}, A_{s+1}^*, Y_T)$ has fat tails in practice as shown in the case $\delta = 0.9$ of Simulation 2.

Thus, in this paper we have proposed the nonlinear and non-Gaussian smoothing procedure in a general form, using the Metropolis-Hastings algorithm and Gibbs sampler. Furthermore, the candidates of the proposal density required to perform the Metropolis-Hastings algorithm have been examined.

APPENDICES

APPENDIX 1: MARKOV CHAIN MONTE CARLO METHODS

Appendix 1.1: Gibbs Sampling

Geman and Geman (1984), Tanner and Wong (1987), Gelfand, Hills, Racine-Poon and Smith (1990), Gelfand and Smith (1990), Carlin and Polson (1991), Zeger and Karim (1991) and so on developed the Gibbs sampling theory. Carlin, Polson and Stoffer (1992), Carter and Kohn (1994, 1996) and De Jong and Shephard (1995) applied to the Gibbs sampler to the nonlinear and/or non-Gaussian state-space models. The Gibbs sampling theory is concisely described as follows.

Consider two random variables x and y in order to make things easier. Two conditional density functions $p(x|y)$ and $p(y|x)$ are assumed to be known, which denote the conditional distribution function of x given y and that of y given x. Suppose that we can easily generate random draws of x from $p(x|y)$ and those of y from $p(y|x)$, respectively. Pick up an arbitrary value for x, i.e., $x^{(0)}$, and we take the following procedure: (i) given $x^{(i-1)}$, generate a random draw of y, i.e., $y^{(i)}$, from $p(y|x^{(i-1)})$, (ii) given $y^{(i)}$, generate a random draw of x, i.e., $x^{(i)}$, from $p(x|y^{(i)})$, and (iii) repeat the procedure for $i = 1, 2, \cdots, N$. From the convergence theory of the Gibbs sampler, as N goes to infinity, we can regard $x^{(N)}$ and $y^{(N)}$ as random draws from $p(x, y)$, which is a joint density function of x and y. The basic result of the Gibbs sampler is as follows.

$$\frac{1}{N-M} \sum_{i=M+1}^{N} g(x^{(i)}, y^{(i)}) \longrightarrow \int g(x, y) p(x, y) \mathrm{d}x \mathrm{d}y, \quad \text{as } N \longrightarrow \infty,$$

where $g(\cdot, \cdot)$ is a function. We may take $M = 0$ but usually 10 – 20 % of N is taken for M. We have considered the bivariate case, but it is easily extended to the multivariate cases.

If x is highly correlated with y, it is known that convergence of the Gibbs sampler is very slow. Let x, y and z be the random vectors. Take an example of the case where x is highly correlated with y. If we generate random draws from $p(x|y, z)$, $p(y|x, z)$ and $p(z|x, y)$, convergence of the Gibbs sampler is slow. Therefore, in this case, without separating x and y, random number generation from $p(x, y|z)$ and $p(z|x, y)$ yields better random draws from the joint density $p(x, y, z)$.

In the state space model, from the structure of the transition equation (2), α_t is highly correlated with α_{t-1}, α_{t-2}, $\cdots$. Accordingly, in the case of the state space model, it is known that convergence of the Gibbs sampler is very slow. However, blocking of the random vectors is very *ad hoc* in practical exercises. And sometimes it is not feasible since we have the case where $p(x, y|z)$ is not available.

Appendix 1.2: Metropolis-Hastings Algorithm

Smith and Roberts (1993), Tierney (1994), Chib and Greenberg (1995, 1996) and Geweke (1996) discussed the Metropolis-Hastings algorithm, which is the random number generation method such that we can generate random draws from any density function.

Consider generating a random draw of z from $p(z)$, which is called the target density function. When it is hard to generate random draws from the target density $p(\cdot)$, the Metropolis-Hastings algorithm can be applied to random number generation. The Metropolis-Hastings algorithm utilizes another distribution function $p_*(z|x)$, which is called the proposal density. The proposal density has to satisfy the following conditions: (i) we can quickly and easily generate random draws from the proposal density and (ii) the proposal density should be distributed with the same range as the target density.

In order to perform the Metropolis-Hastings algorithm, first let us define the acceptance probability $\omega(x, z)$ as:

$$
\omega(x, z) = \begin{cases} \min\left(\dfrac{p(z)p_*(x|z)}{p(x)p_*(z|x)}, 1\right), & \text{if } p(x)p_*(z|x) > 0, \\ 1, & \text{otherwise.} \end{cases}
$$

Using $\omega(x, z)$ defined above, the Metropolis-Hastings algorithm can be implemented as follows.

(i) Take an initial value of x, which is denoted by $x^{(0)}$.

(ii) Given $x^{(i-1)}$, generate a random draw z from $p_*(\cdot|x^{(i-1)})$ and a uniform random draw u from the interval between zero and one.

(iii) Compute $\omega(x^{(i-1)}, z)$ and choose either (a) or (b).

    (a) If $u \leq \omega(x^{(i-1)}, z)$, set $x^{(i)} = z$.

    (b) Otherwise, set $x^{(i)} = x^{(i-1)}$.

(iv) Repeat (ii) and (iii) for $i = 1, 2, \cdots, N$.

(v) Take $x^{(N)}$ as a random draw from $p(x)$ for sufficiently large N.

The basic result of the Metropolis-Hastings algorithm is as follows.

$$\frac{1}{N-M} \sum_{i=M+1}^{N} g(x^{(i)}) \longrightarrow \int g(x)p(x)\mathrm{d}x, \quad \text{as } N \longrightarrow \infty,$$

where $g(\cdot)$ is a function. Usually the first M random draws are excluded from consideration.

    For choice of the proposal density $p_*(z|x)$, the proposal density should not have too large variance and too small variance (see, for example, Chib and Greenberg (1995)). That is, the proposal density should be chosen so that the chain travels over the support of the target density. This may fail to occur, with a consequent undersampling of low probability regions, if the chain is near the mode and if candidates are drawn too close to the current value (see Chib and Greenberg (1996)). For a functional form of the proposal density $p_*(z|x)$, we may take $p_*(z|x) = p_*(z - x)$, called the random walk chain, or $p_*(z|x) = p_*(z)$, called the independence chain.

    Note that $p(z)$ is not necessarily a probability density function, i.e., it is possibly a kernel of the target density function, because of the form of the acceptance probability $\omega(x, z)$. Remember that we need the ratio of the target and proposal densities to derive $\omega(x, z)$.

APPENDIX 2: STATE SPACE MODEL

The recursive density algorithm on filtering is known as follows (for example, see Kitagawa (1987), Harvey (1989) and Kitagawa and Gersch (1996)):

$$p(\alpha_t|Y_{t-1}) = \int p_\alpha(\alpha_t|\alpha_{t-1})p(\alpha_{t-1}|Y_{t-1})\mathrm{d}\alpha_{t-1}, \tag{14}$$

$$p(\alpha_t|Y_t) = \frac{p_y(y_t|\alpha_t)p(\alpha_t|Y_{t-1})}{\int p_y(y_t|\alpha_t)p(\alpha_t|Y_{t-1})\mathrm{d}\alpha_t}, \tag{15}$$

where the initial condition is given by:

$$p(\alpha_1|Y_0) = \begin{cases} \int p_\alpha(\alpha_1|\alpha_0)p(\alpha_0)\mathrm{d}\alpha_0, & \text{if } \alpha_0 \text{ is stochastic,} \\ p_\alpha(\alpha_1|\alpha_0), & \text{otherwise.} \end{cases}$$

Thus, based on the two densities $p_\alpha(\alpha_t|\ \alpha_{t-1})$ and $p_y(y_t|\ \alpha_t)$, equation (14) yields $p(\alpha_t|\ Y_{t-1})$ given $p(\alpha_{t-1}|Y_{t-1})$ and equation (15) yields $p(\alpha_t|Y_t)$ given $p(\alpha_t|Y_{t-1})$. Thus, repeating prediction and updating for all t, the filtering densities $p(\alpha_t|Y_t)$, $t = 1, 2, \cdots, T$, can be recursively obtained. Note that equation (14) corresponds to one-step ahead prediction algorithm.

The density-based fixed-interval smoothing algorithm is represented as (see, for example, Kitagawa (1987), Harvey (1989), Kitagawa and Gersch (1996) and so on):

$$p(\alpha_t|Y_T) = p(\alpha_t|Y_t) \int \frac{p(\alpha_{t+1}|Y_T)p_\alpha(\alpha_{t+1}|\alpha_t)}{p(\alpha_{t+1}|Y_t)} \mathrm{d}\alpha_{t+1}, \tag{16}$$

for $t = T - 1, T - 2, \cdots, 1$. From the density functions $p_\alpha(\alpha_{t+1}|\alpha_t)$, $p(\alpha_t|Y_t)$ and $p(\alpha_{t+1}|Y_t)$, we can obtain the density-based fixed-smoothing algorithm (16), which is a back ward recursion from $p(\alpha_{t+1}|Y_T)$ to $p(\alpha_t|Y_T)$. Thus, the fixed-interval smoothing is derived together with the filtering algorithm given by equations (14) and (15). Note that the smoothing density at time T (i.e., the endpoint case in the smoothing algorithm) is equivalent to the filtering density at time T. In Tanizaki (1999c), it is shown that equation (16) is equivalent to equation (8), which implies that equation (16) can be derived from equation (8).

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

$$\begin{aligned} p(Y_T) &= \prod_{t=1}^{T} p(y_t|Y_{t-1}) \\ &= \prod_{t=1}^{T} \left(\int p_y(y_t|\alpha_t)p(\alpha_t|Y_{t-1})\mathrm{d}\alpha_t \right), \end{aligned} \tag{17}$$

which is called the innovation form of the likelihood function. $p(y_t|Y_{t-1})$ in equation (17) corresponds to the denominator of equation (15).

REFERENCES

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

Atkinson, A.C., Koopman, S.J. and Shephard, N., 1997, " Detecting Shocks: Outliers and Breaks in Time Series, " *Journal of Econometrics*, Vol.80, pp.387 – 422.

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

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

Carlin, B.P., Polson, N.G. and Stoffer, D.S., 1992 " A Monte Carlo Approach to Nonnormal and Nonlinear State Space Modeling, " *Journal of the American Statistical Association*, Vol.87, No.418, pp.493 – 500.

Carter, C.K. and Kohn, R., 1994, " On Gibbs Sampling for State Space Models, " *Biometrika*, Vol.81, No.3, pp.541 – 553.

Carter, C.K. and Kohn, R., 1996, " Markov Chain Monte Carlo in Conditionally Gaussian State Space Models, " *Biometrika*, Vol.83, No.3, pp.589 – 601.

Chib, S. and Greenberg, E., 1995, " Understanding the Metropolis-Hastings Algorithm, " *The American Statistician*, Vol.49, No.4, pp.327 – 335.

Chib, S. and Greenberg, E., 1996, " Markov Chain Monte Carlo Simulation Methods in Econometrics, " *Econometric Theory*, Vol.12, No.4, pp.409 – 431.

De Jong, P. and Shephard, N., 1995, " The Simulation Smoother for Time Series Models, " *Biometrika*, Vol.82, No.2, pp.339 – 350.

Dempster, A.P., Laird, N.M. and Rubin, D.B., 1977, " Maximum likelihood from Incomplete Data via the EM Algorithm, " *Journal of the Royal Statistical Society*, Ser.B, Vol.39, pp.1 – 38 (with discussion).

Durbin, J. and Koopman, S.J., 1997, " Monte Carlo Maximum Likelihood Estimation for Non-Gaussian State Space Models, " *Biometrika*, Vol.84, No.3, pp.669 – 684.

Engle, R.F., 1982, " Autoregressive Conditional Heteroscedasticity with Estimates of Variance of U.K. Inflation, ", *Econometrica*, Vol.50, pp.987 – 1008.

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

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

Gelman, A., Roberts, G.O. and Gilks, W.R., 1996, " Efficient Metropolis Jumping Rules, " in *Bayesian Statistics 5*, edited by Bernardo, J.M., Berger, J.O, David, A.P., and Smith, A.F.M., Oxford University Press, pp.599 – 607.

Geman, S. and Geman D., 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, " Antithetic Acceleration of Monte Carlo Integration in Bayesian Inference, " *Journal of Econometrics*, Vol.38, No.1/2, pp.73 – 89.

Geweke, J., 1989a, " Modeling with Normal Polynomial Expansions, " in *Economic Complexity: Chaos, Sunspots, Bubbles and Nonlinearity*, edited by Barnett, W.A., Geweke, J. and Shell, K., Cambridge University Press.

Geweke, J., 1989b, " Bayesian Inference in Econometric Models Using Monte Carlo Integration, " *Econometrica*, Vol.57, No.6, pp.1317 – 1339.

Geweke, J., 1996, " Monte Carlo Simulation and Numerical Integration, " in *Handbook of Computational Economics*, Vol.1, edited by Amman, H.M., Kendrick, D.A. and Rust, J., pp.731 – 800, North-Holland.

Geweke, J., 1997, " Posterior Simulators in Econometrics, " in *Advances in Economics and Econometrics: Theory and Applications*, Vol.3, edited by Kreps, D. and Wallis, K.F., pp.128 – 165, Cambridge University Press.

Ghysels, E., Harvey, A.C. and Renault, E., 1996, " Stochastic Volatility," in *Handbook of Statistics*, Vol.14, edited by Maddala, G.S. and Rao, C.R., pp.119 – pp.191, North-Holland.

Gordon, N.J., Salmond, D.J. and Smith, A.F.M., 1993, " Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation, " *IEE Proceedings-F*, Vol.140, No.2, pp.107 – 113.

Hamilton, J.D., 1994 *Time Series Analysis*, Princeton University Press.

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

Harvey, A.C. and Streibel, M., 1998, " Testing for a Slowly Changing Level with Special Preference to Stochastic Volatility, " *Journal of Econometrics*, Vol.87, No.1, pp.167 – 189.

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).

Kitagawa, G., 1996, " Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State-Space Models, " *Journal of Computational and Graphical Statistics*, Vol.5, No.1, pp.1 – 25.

Kitagawa, G. and Gersch, W., 1996, *Smoothness Priors Analysis of Time Series* (Lecture Notes in Statistics, No.116), Springer-Verlag.

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.

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

Mariano, R.S. and Tanizaki, H., 1999, " Simulation-Based Inference in Nonlinear State-Space Models: Application to Testing Permanent Income Hypothesis, " in *Simulation-Based Inference in Econometrics: Methods and Applications*, edited by Mariano, R.S., Weeks, M. and Schuermann, T., Cambridge University Press, forthcoming.

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

Pitt, M.K. and Shephard, N., 1997, " Filtering via Simulation: Auxiliary Particle Filters, " Unpublished Manuscript (http:// www.nuff.ox.ac.uk/ economics/papers).

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

Shephard, N. and Pitt, M.K., 1997, " Likelihood Analysis of Non-Gaussian Measurement Time Series, " *Biometrika*, Vol.84, No.3, pp.653 – 667.

Shumway, R.H. and Stoffer, D.S., 1982, " An Approach to Time Series Smoothing and Forecasting Using the EM Algorithm, " *Journal of Time Series Analysis*, Vol.3, pp.253 – 264.

Smith, A.F.M. and Roberts, G.O., 1993, " Bayesian Computation via Gibbs Sampler and Related Markov Chain Monte Carlo Methods, " *Journal of the Royal Statistical Society*, Ser.B, Vol.55, No.1, pp.3 – 23.

Tanizaki, H., 1996, *Nonlinear Filters: Estimation and Applications* (Second, Revised and Enlarged Edition), Springer-Verlag.

Tanizaki, H., 1997, " Nonlinear and Nonnormal Filters Using Monte Carlo Methods, " *Computational Statistics and Data Analysis*, No.25, No.4, pp.417 – 439.

Tanizaki, H., 1999a, " On the Nonlinear and Nonnormal Filter Using Rejection Sampling, " *IEEE Transactions on Automatic Control*, Vol.44, No.2, pp.314 – 319.

Tanizaki, H., 1999b, " Nonlinear and Nonnormal Filter Using Importance Sampling: Antithetic Monte-Carlo Integration, " *Communications in Statistics, Simulation and Computation*, Vol.28, No.2, forthcoming.

Tanizaki, H., 1999c, " Nonlinear and Non-Gaussian State-Space Modeling with Monte Carlo Techniques: A Survey and Comparative Study, " in *Handbook of Statistics*, edited by Rao, C.R. and Shanbhag, D.N.,

North-Holland, forthcoming (http://ht.econ.kobe-u.ac.jp/~tanizaki/cv/papers/survey.pdf).

Tanizaki, H. and Mariano, R.S., 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 (in *Econometric Inference Using Simulation Techniques*, Chap.12, edited by van DijK, H.K., Manfort, A. and Brown, B.W., pp.245 – 261, John Wiley & Sons, 1995).

Tanizaki, H. and Mariano, R.S., 1996, " Nonlinear Filters Based on Taylor Series Expansions, " *Communications in Statistics, Theory and Methods*, Vol.25, No.6, pp.1261 – 1282.

Tanizaki, H. and Mariano, R.S., 1998, " Nonlinear and Non-Gaussian State-Space Modeling with Monte Carlo Simulations, " *Journal of Econometrics*, Vol.83, No.1,2, pp.263 – 290.

Tanner, M.A. and Wong, W.H., 1987, " The Calculation of Posterior Distributions by Data Augmentation, " *Journal of the American Statistical Association*, Vol.82, No.398, pp.528 – 550 (with discussion).

Tierney, L., 1994, " Markov Chains for Exploring Posterior Distributions, " *The Annals of Statistics*, Vol.22, No.4, pp.1701 – 1762.

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

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