Bayesian estimation of state-space models using the Metropolis-Hastings algorithm within Gibbs sampling

Geweke, John

Tanizaki, Hisashi

(Citation)
Computational Statistics & Data Analysis,37(2):151-170

(Issue Date)
2001-08

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90000132

Bayesian Estimation of State-Space Models Using the Metropolis-Hastings Algorithm within Gibbs Sampling*

John Geweke
University of Iowa
Department of Economics
Iowa City, IA 52242-1000, USA
(john-geweke@uiowa.edu)

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

Abstract: In this paper, an attempt is made to show a general solution to nonlinear and/or non-Gaussian state space modeling in a Bayesian framework, which corresponds to an extension of Carlin, Polson and Stoffer (1992) and Carter and Kohn (1994, 1996). Using the Gibbs sampler and the Metropolis-Hastings algorithm, an asymptotically exact estimate of the smoothing mean is obtained from any nonlinear and/or non-Gaussian model. Moreover, taking several candidates of the proposal density function, we examine precision of the proposed Bayes estimator.

Key Words: State-Space Model, Bayesian Estimation, Markov chain Monte Carlo, Gibbs Sampler, Metropolis-Hastings algorithm, Proposal Density, Nonlinear and/or Non-Gaussian Smoothing.

1 Introduction

Since Kitagawa (1987) and Kramer and Sorenson (1988) proposed the nonlinear filter and smoother using numerical integration, nonlinear and/or non-Gaussian filtering and smoothing techniques have been developed. For example, Tanizaki (1996), Tanizaki and Mariano (1994) and Mariano and Tanizaki (1995) applied Monte Carlo integration with importance sampling to nonlinear and non-Gaussian state-space modeling, where a recursive algorithm of weight functions is obtained.

Carlin, Polson and Stoffer (1992) and Carter and Kohn (1994, 1996) utilize Gibbs sampling (also see Chib and Greenberg (1996)), where the smoothing means are evaluated by random draws in a Bayesian framework. Random draws of the state variables for all time periods are jointly generated, which implies that the smoothing procedure is

*The authors are grateful to two anonymous referees for helpful comments and suggestions. However, responsibility for any errors remains entirely with the authors.

formulated.[1] However, they choose the prior densities such that random draws are easily generated, or they utilize rejection sampling as well as Gibbs sampling in the case of the nonlinear system. It is known that rejection sampling is sometimes computationally inefficient. We sometimes have the case where rejection sampling does not work well, depending on the underlying assumptions on the functional form or the error terms. Thus, in their paper, the specific state-space models are taken.

Gordon, Salmond and Smith (1993), Kitagawa (1996) and Kitagawa and Gersch (1996) proposed both filtering and smoothing using the resampling procedure, where random draws from the filtering density and the smoothing density are recursively generated at each time and the random draws from the smoothing are based on those from the filtering density. In the case of smoothing, the resampling approach has the disadvantage that it takes an extremely long time computationally.

Tanizaki (1996, 1999), Tanizaki and Mariano (1998) and Mariano and Tanizaki (2000) proposed nonlinear filter and smoother utilizing rejection sampling, where random draws from the filtering density and the smoothing density are recursively obtained as in the resampling procedure. When the acceptance probability is close to zero, rejection sampling takes a long time computationally. Moreover, we have the case where the acceptance probability is zero. In such a case, rejection sampling cannot be applied.

In order to avoid these computational disadvantages of the existing procedures, Geweke and Tanizaki (1999) suggested the nonlinear and/or non-Gaussian smoother applying the Metropolis-Hastings algorithm and the Gibbs sampler simultaneously, 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. They also focus on smoothing in a non-Bayesian framework.

Utilizing the Metropolis-Hastings algorithm in addition to the Gibbs sampler, in this paper, we deal with any nonlinear and/or non-Gaussian state-space model in a Bayesian framework. Thus, this paper is an extension of Carlin, Polson and Stoffer (1992), Carter and Kohn (1994, 1996) and Geweke and Tanizaki (1999).[2] Moreover, several proposal densities are taken and compared by some Monte Carlo studies, since the critical problem of the Metropolis-Hastings algorithm is choice of the proposal density. We conclude by the root mean square criterion that the proposed procedure are not affected by choice of the proposal densities except for some cases.

[1]Carlin, Polson and Stoffer (1992) showed the filtering procedure, which has more computational burden than the smoother (usually, since the filtering estimate is based on the smoothing estimate, smoothing is more computational than filtering). Therefore, their filter cannot be practically applied.

[2]As mentioned above, Carlin, Polson and Stoffer (1992) and Carter and Kohn (1994, 1996) dealt with the special state-space models which we can easily generated random draws. In the nonlinear cases, Carlin, Polson and Stoffer (1992) utilized rejection sampling within Gibbs sampling, which is sometimes infeasible in practice. However, we can apply the estimation procedure shown in this paper to any nonlinear and/or non-Gaussian state-space models.

Geweke and Tanizaki (1999) developed the nonlinear and/or nonnormal smoother in a non-Bayesian framework, while in this paper we consider it in a Bayesian framework.

2 Metropolis Algorithm within Gibbs Sampling

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

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

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

for $t = 1, 2, \cdots, T$, where T denotes the sample size. 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. $h_t(\cdot,\cdot,\cdot)$ and $f_t(\cdot,\cdot,\cdot)$ are the vector functions, which are assumed to be known. We introduce the nuisance parameters γ and δ into the state-space model. Let Y_t be the information set up to time t, i.e., $Y_t = \{y_1, y_2, \cdots, y_t\}$. We consider estimating the conditional expectation of α_t using information Y_T, i.e., $\alpha_{t|T} \equiv \mathrm{E}(\alpha_t|Y_T)$.[3]

Consider deriving $P(\alpha_t|Y_T)$ to obtain the smoothing mean $\alpha_{t|T}$. Let us define $A_t = \{\alpha_0, \alpha_1, \cdots, \alpha_t\}$, which is a set consisting of the state-variables up to time t. Define $P_y(y_t|\alpha_t, \gamma)$ and $P_\alpha(\alpha_t|\alpha_{t-1}, \delta)$ as the density functions obtained from measurement equation (1) and transition equation (2). Denote the prior distributions of γ and δ by $P_\gamma(\gamma)$ and $P_\delta(\delta)$. Let $P(A_t, Y_t|\gamma, \delta)$, $P_\alpha(A_t|\delta)$ and $P_y(Y_t|A_t, \gamma)$ be the joint density of A_t and Y_t given γ and δ, the density of A_t given δ and the conditional density of Y_t given A_t and γ, respectively.

Under the setup, the density of A_T and Y_T given γ and δ, i.e., $P(A_T, Y_T|\gamma, \delta)$, is written as:

$$P(A_T, Y_T|\gamma, \delta) = P_\alpha(A_T|\delta) P_y(Y_T|A_T, \gamma), \tag{3}$$

where the two densities in the right hand side are represented by:

$$P_\alpha(A_T|\delta) = \begin{cases} P_\alpha(\alpha_0|\delta) \displaystyle\prod_{t=1}^{T} P_\alpha(\alpha_t|\alpha_{t-1}, \delta), & \text{if } \alpha_0 \text{ is stochastic}, \\ \displaystyle\prod_{t=1}^{T} P_\alpha(\alpha_t|\alpha_{t-1}, \delta), & \text{otherwise}, \end{cases} \tag{4}$$

$$P_y(Y_T|A_T, \gamma) = \prod_{t=1}^{T} P_y(y_t|\alpha_t, \gamma), \tag{5}$$

where $P_\alpha(\alpha_0|\delta)$ denotes the initial density of α_0 when α_0 is assumed to be a random variable. From the Bayes theorem, the conditional distribution of A_T given Y_T, γ and δ, i.e., $P(A_T|Y_T, \gamma, \delta)$, is obtained as follows:

$$P(A_T|Y_T, \gamma, \delta) = \frac{P(A_T, Y_T|\gamma, \delta)}{\displaystyle\int P(A_T, Y_T|\gamma, \delta)\,\mathrm{d}A_T}. \tag{6}$$

[3]The conditional expectation $\alpha_{t|s} = \mathrm{E}(\alpha_t|Y_s)$ 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 $\alpha_{t|T}$.

Using the Metropolis-Hastings algorithm and the Gibbs sampler, an attempt is made to generate random draws of A_T, γ and δ directly from $P(A_T|Y_T,\gamma,\delta)$, $P(\gamma|A_T,Y_T,\delta)$ and $P(\delta|A_T,Y_T,\gamma)$. Define $A^*_{t+1}=\{\alpha_{t+1},\alpha_{t+2},\cdots,\alpha_T\}$, where $A_T=\{A_t,A^*_{t+1}\}$, $A_T=A^*_0$ and $A^*_{T+1}=\emptyset$ (an empty set). According to the Gibbs sampler, random draws of A_T from $P(A_T|Y_T,\gamma,\delta)$ are based on those of α_t from $P(\alpha_t|A_{t-1},A^*_{t+1},Y_T,\gamma,\delta)$ for $t=1,2,\cdots,T$. To perform the Gibbs sampler, therefore, we need to obtain the conditional density function of α_t given A_{t-1}, A^*_{t+1}, Y_T, γ and δ, which is derived from equation (6) and is shown in the following equation:

$$
\begin{aligned}
&P(\alpha_t|A_{t-1},A^*_{t+1},Y_T,\gamma,\delta)\\
&\quad=\frac{P(A_T|Y_T,\gamma,\delta)}{\displaystyle\int P(A_T|Y_T,\gamma,\delta)\mathrm{d}\alpha_t}\\
&\quad\propto\begin{cases}P_y(y_t|\alpha_t,\gamma)P_\alpha(\alpha_t|\alpha_{t-1},\delta)P_\alpha(\alpha_{t+1}|\alpha_t,\delta), & \text{if } t=1,2,\cdots,T-1,\\ P_y(y_t|\alpha_t,\gamma)P_\alpha(\alpha_t|\alpha_{t-1},\delta), & \text{if } t=T \text{ (endpoint).}\end{cases}
\end{aligned}
\tag{7}
$$

The posterior density of γ given A_T, Y_T and δ, i.e., $P(\gamma|A_T,Y_T,\delta)$, and that of δ given A_T, Y_T and γ, i.e., $P(\delta|A_T,Y_T,\gamma)$, are given by:

$$P(\gamma|A_T,Y_T,\delta)\propto P_y(Y_T|A_T,\gamma)P_\gamma(\gamma), \tag{8}$$

$$P(\delta|A_T,Y_T,\gamma)\propto P_\alpha(A_T|\delta)P_\delta(\delta). \tag{9}$$

Thus, all the kernels of $P(\alpha_t|A_{t-1},\ A^*_{t+1},\ Y_T,\ \gamma,\ \delta)$ for $t=1,2,\cdots,T$, $P(\gamma|A_T,\ Y_T,\ \delta)$ and $P(\delta|A_T,\ Y_T,\ \gamma)$ are obtained. Utilizing the three kernels given by (7) – (9), we consider evaluating the smoothing mean by generating random draws of A_T, γ and δ directly from the conditional density $P(A_T,\gamma,\delta|Y_T)$. The Metropolis-Hastings algorithm within the Gibbs sampler is applied to random number generation. From the three posterior densities (7) – (9), the smoothing random draws are generated as follows:

(i) Take appropriate values[4] for γ, δ and α_t, $t=0,1,\cdots,T$.

(ii) Generate a random draw of α_t from $P(\alpha_t|A_{t-1},A^*_{t+1},Y_T,\gamma,\delta)$ for $t=1,2,\cdots,T$.

(iii) Generate a random draw of γ from $P(\gamma|A_T,Y_T,\delta)$.

(iv) Generate a random draw of δ from $P(\delta|A_T,Y_T,\gamma)$.

(v) Repeat (ii) – (iv) N times to obtain N random draws of A_T, δ and γ.

In Steps (ii) – (v), the random draws of A_T, δ and γ are updated, which sampling method is called the Gibbs sampler. See Geman and Geman (1984), Tanner and Wong (1987), Gelfand, Hills, Racine-Poon and Smith (1990), Gelfand and Smith (1990), Carlin and

[4]Typically, the smoothed estimates based on the extended Kalman filter are taken for α_t, $t=0,1,\cdots,T$. 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)). Moreover, α_0 is generated from the initial density $P_\alpha(\alpha_0|\delta)$ if α_0 is stochastic and it is fixed as α_0 if α_0 is nonstochastic.

Polson (1991), Geweke (1996, 1997) and so on for the Gibbs sampler. We may change the order of Steps (ii) – (iv). Moreover, in Steps (ii) – (iv), generally it is intractable to generate the random draws of α_t for $t = 1, 2, \cdots, T$, γ and δ. In this case, there are two ways to generate random draws, i.e., one is rejection sampling[5] and another is the Metropolis-Hastings algorithm (see Appendix 1 for the Metropolis-Hastings algorithm). It is well known that rejection sampling takes a long time computationally when the acceptance probability is close to zero and also rejection sampling cannot be applied when the acceptance probability is zero. Therefore, in order to avoid these computational disadvantages of rejection sampling, in this paper we suggest applying the Metropolis-Hastings algorithm in Steps (ii) — (iv), when it is not feasible to generate the random draws from the corresponding posterior densities. Thus, the Gibbs sampler and the Metropolis-Hastings algorithm are combined in order to obtain the smoothing random draws from any state-space model.

Let $g(\cdot)$ be a function. Consider estimating $\mathrm{E}(g(\alpha_t)|Y_T)$, based on the N random draws of α_t for $t = 1, 2, \cdots, T$. The smoothing estimate of $\mathrm{E}(g(\alpha_t)|Y_T)$ is simply obtained as the arithmetic average of the $N-M$ random draws of $g(\alpha_t)$. Especially, the case of $g(\alpha_t) = \alpha_t$ represents the smoothing mean (i.e., $\alpha_{t|T}$), while the case of $g(\alpha_t) = (\alpha_t - \alpha_{t|T})(\alpha_t - \alpha_{t|T})'$ gives us the smoothing variance. Usually, we ignore the first M random draws from consideration, because of convergence of the Markov chain Monte Carlo methods.

3 Choice of Proposal Density

The Metropolis-Hastings algorithm has the problem of specifying the proposal density as in Appendix 1, which is the crucial criticism. Several generic choices of the proposal density are discussed by Tierney (1994) and Chib and Greenberg (1995). Let $P_{*\alpha}(z|x)$, $P_{*\gamma}(z|x)$ and $P_{*\delta}(z|x)$ be the proposal densities of α_t, γ and δ, respectively.

3.1 On the Proposal Density $P_{*\alpha}(z|x)$

For the proposal density $P_{*\alpha}(z|x)$, we can consider the following candidates, which are utilized in Section 4.

3.1.1 Proposal Density I

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

[5] Carlin, Polson and Stoffer (1992) utilizes rejection sampling in the case of the nonlinear system. As mentioned above, rejection sampling is sometimes computationally inefficient. Accordingly, we sometimes have the case such that rejection sampling does not work well, depending on the underlying assumptions on the functional form or the error terms. Improving this issue, in this paper, we introduce the estimation procedure which can be applied to any state-space model.

3.1.2 Proposal Density II

It is possible for the proposal density to utilize the extended Kalman smoothed estimates, i.e., $P_{*\alpha}(z|x) = N(a^*_{t|T}, c\Sigma^*_{t|T})$, where $a^*_{t|T}$ and $\Sigma^*_{t|T}$ denote the first- and the second-moments (i.e., mean and variance) based on the extended Kalman smoothed estimates at time t. This proposal density is also the independence chain. c is an appropriate constant value. In Monte Carlo studies of the next section, $c = 1, 2, 4, 16$ is taken.

3.1.3 Proposal Density III

We may take the proposal density called the random walk chain, i.e., $P_{*\alpha}(z|x) = P_{*\alpha}(z - x)$. In this paper, we consider the proposal density as $P_{*\alpha}(z|x) = N(x, c\Sigma^*_{t|T})$. As defined in Section 3.1.2, c takes an appropriate constant value.

3.1.4 Proposal Density IV

The alternative proposal density is based on approximation of the log-kernel (see Geweke and Tanizaki (1999)). Let $q(z) = \log(P(z))$, where $P(z)$ may denote the kernel which corresponds to equation (7). Approximating the log-kernel $q(z)$ around x by the second-order Taylor series expansion, $q(z)$ is represented as:

$$q(z) \approx q(x) + q'(x)(z - x) + \frac{1}{2}q''(x)(z - x)^2, \tag{10}$$

where $q'(\cdot)$ and $q''(\cdot)$ denote the first- and the second-derivatives. Depending on the sign of $q''(x)$, we have the following four cases, i.e., Cases 1 – 4.

Case 1: $q''(x) < 0$: Equation (10) is written by:

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

where $d(x)$ is an appropriate function of x. The second term in the right hand side is equivalent to the exponential part of the normal density. Therefore, $P_{*\alpha}(z|x)$ is taken as $N(\mu, \sigma^2)$, where $\mu = x - q'(x)/q''(x)$ and $\sigma^2 = (-q''(x))^{-1}$.

Case 2: $q''(x) \geq 0$ and $q'(x) < 0$: Perform linear approximation of $q(z)$. Let x_1^* be the nearest mode with $x_1^* < x$. Then, $q(z)$ is approximated by a line passing between x_1^* and x, which is written as: $q(z) \approx q(x_1^*) + \dfrac{q(x_1^*) - q(x)}{x^* - x}(z - x_1^*)$. From the second term in the right hand side, the proposal density is represented as the exponential distribution with $z > x_1^* - d$, i.e., $P_{*\alpha}(z|x) = \lambda \exp\Big(-\lambda(z - (x_1^* - d))\Big)$ if $x_1^* - d < z$ and $P_{*\alpha}(z|x) = 0$ otherwise, where $\lambda = \left|\dfrac{q(x_1^*) - q(x)}{x_1^* - x}\right|$ and d is a positive value (see Footnote 6 for d). Thus, z is generated by $z = w + (x_1^* - d)$, where w follows 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 nearest mode with $x < x_2^*$. Approximation of $q(z)$ is exactly equivalent to that of Case 2. Taking into account $z < x_2^* + d$, the proposal density is written as: $P_{*\alpha}(z|x) = \lambda \exp\Big(-\lambda((x_2^* + d) - z)\Big)$ if $z < x_2^* + d$ and $P_{*\alpha}(z|x) = 0$ otherwise. Thus, z is generated by $z = (x_2^* + d) - w$, where w is the exponential distribution with parameter λ.

Case 4: $q''(x) \geq 0$ **and** $q'(x) = 0$: In this case, $q(z)$ is approximated as a uniform distribution at the neighborhood of x. As for the range of the uniform distribution, we utilize the two appropriate values x_1^* and x_2^*, which satisfies $x_1^* < x < x_2^*$. When we have two modes, x_1^* and x_2^* may be taken as the modes. Thus, the proposal density $P_{*\alpha}(z|x)$ is obtained by the uniform distribution on the interval between x_1^* and x_2^*, or possibly the interval between $(x_1^* - d)$ and $(x_2^* + d)$.

Thus, for approximation of the kernel, all the possible cases are given by Cases 1 – 4 depending on the signs of $q'(\cdot)$ and $q''(\cdot)$. For positive d, the generated random draw may move from one case to another,[6] which implies that the irreducibility condition of the Gibbs sampler is guaranteed. Moreover, applying the procedure above to each element of the state vector, Proposal IV is easily extended to multivariate cases.

3.2 On the Proposal Densities $P_{*\gamma}(z|x)$ and $P_{*\delta}(z|x)$

For the proposal densities $P_{*\gamma}(z|x)$ and $P_{*\delta}(z|x)$, in the case where the random draws are easily generated from the posterior density, we do not need to perform the Metropolis-Hastings algorithm. Therefore, we can generate random draws directly from the posterior density in such a case.

However, generally it is quite rare to have the above case. When it is infeasible to generate random draws from the posterior density, in order to perform the Metropolis algorithm we may take the following proposal densities for γ and δ. As the first strategy, we may take the proposal density as the prior density, i.e., $P_{*\gamma}(z|x) = P_\gamma(z)$ and $P_{*\delta}(z|x) = P_\delta(z)$, which corresponds to the Proposal Density I in Section 3.1.1. Second, it might be also possible to apply the Proposal Density IV (Section 3.1.4) to $P_{*\gamma}(z|x)$ and $P_{*\delta}(z|x)$. As another candidate of the proposal densities $P_{*\gamma}(z|x)$ and $P_{*\delta}(z|x)$, when γ and δ lie on an interval, we may generate uniform random draws between the interval.

3.3 Discussion

We should keep in mind some remarks when we choose the proposal density, which are discussed as follows.

[6] As an example, consider the unimodal density in which we have Cases 2 and 3. Let x^* be the mode. We have Case 2 in the right hand side of x^* and Case 3 in the left hand side of x^*. In the case of $d = 0$, we have the random draws generated from either of Case 2 or 3. In this situation, the generated random draw does not move from one case to another. In the case of $d > 0$, however, the distribution in Case 2 can generate a random draw in Case 3. That is, for positive d, the generated random draw may move from one case to another. In Section 4, we take $d = 1/\lambda$, which is a standard error of the exponential distribution with parameter λ.

When the one-step prediction density $P_\alpha(\alpha_t|\alpha_{t-1};\delta)$ is far from the posterior density $P(\alpha_t|A_{t-1},A^*_{t+1},Y_T;\gamma,\delta)$, shown in equation (7), i.e., when $P_\alpha(\alpha_t|\alpha_{t-1};\delta)$ is far from $P_y(y_t|\alpha_t;\gamma)$, Proposal Density I is not appropriate. In the case where the measurement equation is highly nonlinear, it might be expected that we often have this case.

In Proposal II, we utilize the first- and second-moments of the extended Kalman smoother. The peak and range of $P(\alpha_t|Y_T)$ are not known in general, but mean and variance of the state-variable can be estimated by the extended Kalman smoothed algorithm, even if the extended Kalman smoother indicates the biased estimator. It might be appropriate to consider that the extended Kalman smoothed estimates are not too far from the true state mean values. Therefore, the Metropolis algorithm would be improved by utilizing the proposal density based on the extended Kalman smoother.

In Proposal Densities II and III, moreover, c should be equal to or greater than one. The support of the proposal distribution should include that of the target density (see, for example, Geweke (1996)). For Proposal II, $N(a^*_{t|T},c\Sigma^*_{t|T})$ is used to cover $P(\alpha_t|A_{t-1},A^*_{t+1},Y_T;\gamma,\delta)$. The support of $N(a^*_{t|T},c\Sigma^*_{t|T})$ should include that of the posterior density $P(\alpha_t|A_{t-1},A^*_{t+1},Y_T;\gamma,\delta)$. Since the support of $P(\alpha_t|A_{t-1},A^*_{t+1},Y_T;\gamma,\delta)$ is not known in general, it might be safe that the proposal density $P_{*\alpha}(z|x)$ should be more broadly distributed than the posterior density $P(\alpha_t|A_{t-1},A^*_{t+1},Y_T;\gamma,\delta)$. Therefore, it is appropriate that $c\geq 1$ should be chosen. Otherwise, Proposal Density II could be break down, especially for the multimodal posterior density. Similarly, for Proposal Density III, when c is small, it takes a long time for the chain to travel over the support of the target density. Accordingly, c should not be too small for Proposal Density III.

Proposal Density IV gives us a poor approximation in both tails of the target density. In the tails, the random draws generated from the proposal density result in the outliers, because we sometimes have the case where the approximated density in the tails has too large variance, compared with the target density.

For the other possible candidate, in the case where the state-variable α_t lies on an interval, a uniform distribution between the interval might be taken as the proposal density. However, this case would not be usual in practice.

4 Monte Carlo Studies

In this section, the simulation-based smoothers with Proposal Densities I – IV are investigated. Five state space models are examined, i.e., Simulations A – E. For simplicity, we consider the cases where γ is not included in equation (1) while δ is in equation (2). The simulation procedure is as follows:

(i) Generate random numbers of ϵ_t and η_t for $t=1,2,\cdots,T$, based on the underlying assumptions of the error terms. Given δ and the random draws of ϵ_t and η_t, we obtain a set of data y_t and α_t, $t=1,2,\cdots,T$, from equations (1) and (2), where $T=100$ and $\delta=0.5,0.9$.

(ii) Given Y_T, perform the Bayes estimator shown in Section 2 in order to obtain the state mean of α_t (i.e., $\alpha_{t|T}$) and the Bayes mean of δ (say, $\overline{\delta}$), where we take $(M,N)=$

(3000, 8000), (5000, 10000). As for the prior density of δ, the diffuse prior is chosen for Simulations A and B and the uniform prior is taken for Simulations C – D.

(iii) Repeat (i) and (ii) G times, where $G = 1000$ is taken.

(1) For comparison of the state-variable α_t, compute the root mean square error (RMS) for each estimator, which is defined as:

$$\mathrm{RMS} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{G}\sum_{g=1}^{G}(\alpha_{t|T}^{(g)} - \alpha_t^{(g)})^2\right)^{1/2}.$$

$\alpha_{t|T}$ in the equation above takes the estimated state mean, while α_t denotes the artificially simulated state-variable. The superscript (g) denotes the g-th simulation run. That is, $\alpha_t^{(g)}$ indicates the simulated state-variable at time t in the g-th simulation run.

(2) For comparison of the parameter δ, compute the arithmetic average (AVE) and the root mean square error (RMS) of $\overline{\delta}$, i.e., $\mathrm{AVE} = \dfrac{1}{G}\sum_{g=1}^{G}\overline{\delta}^{(g)}$ and $\mathrm{RMS} = \left(\dfrac{1}{G}\sum_{g=1}^{G}(\overline{\delta}^{(g)} - \delta)^2\right)^{1/2}$. In Simulations C and D, the standard error (SER) is used instead of RMS.

Under the above setup, in this section, we examine several types of state-space models, i.e., ARCH Model, Stochastic Volatility Model, Structural Change Model, Shifted-Mean Model and Nonstationary Growth Model.

Simulation A (ARCH Model): Consider the nonlinear system: $y_t = \alpha_t + \epsilon_t$ and $\alpha_t = (1-\delta+\delta\alpha_{t-1}^2)^{1/2}\eta_t$, where $\delta = 0.5, 0.9$ is taken.[7] $\alpha_0 \sim N(0,1)$ and $\begin{pmatrix}\epsilon_t\\ \eta_t\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}\begin{pmatrix}1&0\\0&1\end{pmatrix}\right)$ are assumed, which are also used in Simulations B – D. The uniform distribution is taken for the prior density of δ, i.e., $P_\delta(\delta) = 1$ for $0 < \delta < 1$. See Engle (1982) and Bollerslev, Engle and Nelson (1994) for the ARCH model.

Simulation B (Stochastic Volatility Model): Take an example of the following state space model: $y_t = \exp(0.5\alpha_t)\,\epsilon_t$ and $\alpha_t = \delta\alpha_{t-1} + \eta_t$, where $\delta = 0.5, 0.9$ is taken. The uniform prior $P_\delta(\delta) = 1$ for $0 < \delta < 1$ is assumed. See Ghysels, Harvey and Renault (1996) for the stochastic volatility model.

Simulation C (Structural Change Model): The data generating process is given by: $y_t = d_t + \delta\alpha_t + \epsilon_t$ and $\alpha_t = \delta\alpha_{t-1} + \eta_t$, where $\delta = 0.5, 0.9$, but the estimated system is: $y_t = \alpha_t + \epsilon_t$ and $\alpha_t = \delta\alpha_{t-1} + \eta_t$, where $d_t = 1$ for $t = 21, 22, \cdots, 40$, $d_t = -1$ for

[7]Note from the transition equation that the unconditional variance of α_t is assumed to be one. In this paper, $\delta = 0.5, 0.9$ is examined. α_t is distributed with large tails in the case of $\delta = 0.9$, because the forth-moment of the ARCH(1) process does not exist when $\delta^2 > 1/\sqrt{3}$. See Engle (1982) and Bollerslev, Engle and Nelson (1994).

$t = 61, 62, \cdots, 80$ and $d_t = 0$ otherwise. The diffuse prior is assumed for δ.[8] This model corresponds to the case where the sudden shifts occur at time periods 21, 41, 61 and 81.

Simulation D (Shifted-Mean Model): The data generating process is given by: $y_t = d_t + \epsilon_t$, but the estimated system is: $y_t = \alpha_t + \epsilon_t$ and $\alpha_t = \delta\alpha_{t-1} + \eta_t$, where d_t is defined in Simulation C. The diffuse prior is assumed for δ.

Simulation E (Nonstationary Growth Model): The system is: $y_t = \alpha_t^2/20 + \epsilon_t$ and $\alpha_t = \delta_1\alpha_{t-1} + \delta_2\alpha_{t-1} / (1 + \alpha_{t-1}^2) + \delta_3 \cos(1.2(t-1)) + \eta_t$, where the true parameter values are taken as: $(\delta_1, \delta_2, \delta_3) = (0.5, 25, 8)$. ϵ_t, η_t and α_0 are mutually independently distributed as: $\epsilon_t \sim N(0, 1)$, $\eta_t \sim N(0, 10)$ and $\alpha_0 \sim N(0, 10)$. The diffuse prior is assumed for δ. This model is examined in Kitagawa (1987, 1996) and Carlin et al. (1992).

In Simulation A, the transition equation follows the first-order autoregressive conditional heteroscedasticity model (i.e., ARCH(1) model), while the measurement equation consists of the state-variable and the error term. Simulation B is called the stochastic volatility model, where the transition equation follows the AR(1) process and the measurement equation denotes a nonlinear function of the state-variable and the error term. In Simulations C and D, the estimated model is different from the data generating process. Simulation E is called the nonstationary growth model, which is nonlinear in both the transition and measurement equations.

Results and Discussion: The results are in Tables 1 – 6 for the suggested Bayes estimator and Table 7 for the extended Kalman smoothed estimator (EK) and the importance resampling smoother (IR). In Table 7, EK denotes the standard linear recursive Kalman smoothed estimator for Simulations C and D and the extended Kalman smoothed estimator for Simulations A, B and E. IR is briefly discussed in Appendix 2. I, II, III and IV indicate the simulation-based nonlinear smoothers with the Proposal Densities I – IV shown in Section 3.1. For Proposal Densities II and III, we examine $c = 1, 2, 4, 16$. $\delta = 0.5, 0.9$ is chosen for Simulations A and B. For EK and IR in Table 7, δ is not estimated and it is assumed to be constant, because EK and IR are not Bayes estimators. For the simulation-based nonlinear estimator discussed in this paper, as mentioned above, the uniform prior of δ is assumed for Simulations A and B and the diffuse prior of δ is for Simulations C – E. For each estimator, we compare RMS for α_t and both AVE and RMS (or SER) for δ.

First of all, in Tables 1 – 5, to check convergence diagnostics on the Metropolis within the Gibbs sampler, the cases of $(M, N) = (3000, 8000), (5000, 10000)$ are examined. For the burnin period M, there are some diagnostic tests, which are discussed in Geweke (1992) and Mengersen, Robert and Guihenneuc-Jouyaux (1999). However, since their tests are applicable in the case of one sample path, we cannot utilize them. Because G

[8] When the diffuse prior is assumed for δ (i.e., $P_\delta(\delta)$ is constant for δ), the posterior density of δ is given by $\delta \sim N(\widehat{\delta}, s_\delta^2)$, where $\widehat{\delta}$ and s_δ^2 are the ordinary least squares estimates, i.e., $\widehat{\delta} = (\sum_{t=1}^T \alpha_{t-1}^2)^{-1} \sum_{t=1}^T \alpha_{t-1}\alpha_t$ and $s_\delta^2 = (\sum_{t=1}^T \alpha_{t-1}^2)^{-1}$.

Table 1: Simulation A (ARCH Model): $T = 100$ and $N - M = 5000$

		$\delta = 0.5$			$\delta = 0.9$		
		α_t	δ		α_t	δ	
	c	RMS	AVE	RMS	RMS	AVE	RMS
$M = 3000$							
I		.683	.447	.147	.530	.700	.247
II	1	.684	.453	.147	.572	.859	.090
II	2	.683	.454	.146	.555	.740	.221
II	4	.683	.452	.147	.540	.720	.232
II	16	.683	.446	.147	.530	.715	.231
III	1	.683	.448	.147	.532	.708	.239
III	2	.683	.447	.147	.531	.708	.238
III	4	.683	.447	.148	.530	.705	.240
III	16	.683	.444	.147	.530	.700	.244
IV		.682	.448	.145	.530	.706	.237
$M = 5000$							
I		.683	.446	.148	.530	.699	.248
II	1	.684	.453	.148	.573	.854	.097
II	2	.683	.454	.147	.555	.743	.217
II	4	.683	.451	.146	.540	.723	.230
II	16	.683	.446	.146	.530	.714	.231
III	1	.683	.446	.147	.532	.707	.239
III	2	.683	.447	.147	.531	.707	.239
III	4	.683	.447	.146	.531	.703	.242
III	16	.683	.446	.146	.530	.701	.243
IV		.682	.450	.144	.530	.705	.239

Table 2: Simulation B (Stochastic Volatility Model): $T = 100$ and $N - M = 5000$

		$\delta = 0.5$			$\delta = 0.9$		
		α_t	δ		α_t	δ	
	c	RMS	AVE	RMS	RMS	AVE	RMS
$M = 3000$							
I		.916	.422	.156	.937	.865	.081
II	1	.917	.421	.156	.942	.865	.081
II	2	.916	.422	.156	.939	.864	.081
II	4	.916	.421	.156	.939	.864	.081
II	16	.917	.419	.157	.942	.864	.082
III	1	.917	.420	.157	.938	.865	.081
III	2	.916	.420	.157	.938	.864	.082
III	4	.917	.420	.157	.939	.864	.081
III	16	.917	.419	.157	.941	.864	.081
IV		.916	.423	.155	1.021	.856	.084
$M = 5000$							
I		.916	.423	.155	.937	.865	.081
II	1	.917	.420	.156	.941	.865	.081
II	2	.916	.421	.156	.939	.865	.081
II	4	.916	.421	.155	.939	.864	.082
II	16	.917	.419	.157	.941	.864	.080
III	1	.917	.420	.157	.938	.864	.081
III	2	.916	.419	.157	.938	.865	.081
III	4	.917	.419	.157	.939	.864	.082
III	16	.917	.420	.156	.941	.864	.081
IV		.916	.423	.155	1.020	.856	.084

Table 3: Simulation C (Structural Change Model): $T = 100$ and $N - M = 5000$

		$\delta = 0.5$			$\delta = 0.9$		
		α_t	δ		α_t	δ	
	c	RMS	AVE	SER	RMS	AVE	SER
$M = 3000$							
I		.726	.784	.059	.690	.978	.008
II	1	.726	.785	.058	.690	.978	.008
II	2	.726	.785	.059	.690	.978	.008
II	4	.726	.784	.059	.690	.978	.008
II	16	.726	.784	.059	.691	.978	.008
III	1	.727	.784	.059	.691	.978	.008
III	2	.726	.784	.059	.691	.978	.008
III	4	.726	.784	.059	.691	.978	.008
III	16	.726	.784	.059	.691	.978	.008
IV		.726	.785	.059	.690	.978	.007
$M = 5000$							
I		.726	.784	.059	.690	.978	.008
II	1	.726	.785	.059	.690	.978	.008
II	2	.726	.785	.059	.690	.978	.008
II	4	.726	.784	.059	.691	.978	.008
II	16	.727	.784	.059	.691	.978	.008
III	1	.727	.784	.059	.691	.978	.008
III	2	.726	.784	.059	.691	.978	.008
III	4	.726	.784	.059	.691	.978	.008
III	16	.727	.784	.059	.691	.978	.008
IV		.726	.785	.059	.691	.978	.008

Table 4: Simulation D (Shifted-Mean Model): $T = 100$ and $N - M = 5000$

		$M = 3000$			$M = 5000$		
		α_t	δ		α_t	δ	
	c	RMS	AVE	SER	RMS	AVE	SER
I		.524	.365	.130	.524	.365	.130
II	1	.524	.365	.130	.524	.365	.130
II	2	.524	.365	.130	.524	.365	.130
II	4	.524	.365	.130	.524	.365	.130
II	16	.524	.363	.130	.524	.364	.130
III	1	.524	.364	.130	.524	.364	.130
III	2	.524	.364	.130	.524	.364	.130
III	4	.524	.364	.130	.524	.364	.130
III	16	.524	.363	.130	.524	.363	.130
IV		.524	.366	.130	.524	.366	.131

simulation runs are performed in this paper, we have G test statistics when we apply the diagnostic tests. It is difficult to evaluate G testing results at the same time. Therefore, given $N - M = 5000$, we consider using the alternative approach to see if $M = 3000$ is sufficient. If the cases of $M = 3000$ are close to those of $M = 5000$, we can conclude that $M = 3000$ is sufficiently large for Simulations A – E. From Tables 1 – 5, $M = 3000$ and $M = 5000$ are very close to each other. In Table 6, the acceptance probabilities on average are shown, which are obtained in the Metropolis algorithm (see Appendix 1). If the acceptance probabilities on average are close to zero, the chain does not travel over the support of the target density. However, we obtain the result that the Metropolis algorithm works because the acceptance probabilities on average are not close to zero for Simulations A – E. Thus, in Tables 1 – 5 the cases of $M = 3000$ are very close to those of $M = 5000$ and in Table 6 the acceptance probabilities are far from zero. Therefore, we can conclude that $M = 3000$ is large enough, which implies that $M = 5000$ is also sufficiently large. Accordingly, hereafter we focus on the cases of $M = 5000$.

Simulation A (the ARCH(1) model with a white noise) is shown in Table 1. For $\delta = 0.5$, RMS of α_t does not depend on the proposal density, because RMSs of α_t are very close to each other when $\delta = 0.5$. However, in the case of $\delta = 0.9$, from RMSs of α_t, Proposal Densities I, III and IV are better than Proposal Density II. For the cases of $\delta = 0.9$ and Proposal Density II, AVEs of δ are close to the true parameter value but the case of $c = 1$ and Proposal Density II is not realistic because in the case of $c = 1$ and Proposal Density II the RMSs of δ are too small, compared with the others. Thus, we can see that the support of the target density is not included in that of Proposal Density II, i.e., the target density is far from Proposal Density II.

In Table 2, for the case of $\delta = 0.5$, all the Proposal Densities I – IV are very similar, although AVEs of Proposal Densities I and IV are slightly close to the true value, compared with Proposal Densities II and III. However, when $\delta = 0.9$, Proposal Density IV turns

Table 5: Simulation E (Nonstationary Growth Model): $T = 100$ and $N - M = 5000$

	c	$\delta_1 = .5$ AVE	RMS	$\delta_2 = 25$ AVE	RMS	$\delta_3 = 8$ AVE	RMS
$M = 3000$							
I		.477	.071	26.776	3.578	1.421	7.014
II	1	.499	.112	27.727	16.320	2.814	5.509
II	2	.493	.100	28.485	14.090	2.494	5.801
II	4	.486	.088	29.159	12.295	2.168	6.110
II	16	.493	.079	28.340	10.543	1.766	6.542
III	1	.477	.081	25.607	6.618	2.228	6.065
III	2	.483	.072	26.085	4.958	1.930	6.359
III	4	.490	.065	26.407	3.786	1.727	6.565
III	16	.499	.064	26.362	4.389	1.719	6.618
IV		.422	.137	25.593	8.173	2.438	5.997
$M = 5000$							
I		.478	.070	26.864	3.632	1.350	7.067
II	1	.498	.112	27.800	16.353	2.794	5.521
II	2	.495	.103	28.094	14.734	2.477	5.815
II	4	.488	.088	29.193	12.767	2.153	6.126
II	16	.493	.080	28.213	11.087	1.732	6.578
III	1	.480	.078	25.473	6.069	2.169	6.121
III	2	.485	.070	26.086	4.449	1.885	6.413
III	4	.493	.062	26.460	3.459	1.691	6.604
III	16	.500	.061	26.409	3.876	1.698	6.652
IV		.421	.137	25.808	8.228	2.447	5.990

Table 6: Acceptance Probability on Average: $T = 100$ and $M = 5000$

	$c \setminus \delta$	A 0.5	A 0.9	B 0.5	B 0.9	C 0.5	C 0.9	D 0.5	D 0.9	E
I		.619	.674	.690	.586	.532	.479	.629	—	.217
II	1	.823	.673	.607	.276	.734	.698	.849	.784	.357
II	2	.738	.672	.513	.226	.625	.557	.704	.639	.327
II	4	.579	.643	.398	.175	.482	.418	.535	.482	.281
II	16	.315	.440	.214	.095	.257	.219	.283	.254	.189
III	1	.698	.772	.597	.348	.653	.604	.678	.648	.564
III	2	.603	.699	.489	.260	.549	.495	.578	.543	.481
III	4	.498	.611	.383	.189	.439	.388	.468	.434	.397
III	16	.298	.413	.211	.097	.249	.214	.271	.245	.246
IV		.924	.859	.927	.945	1.000	1.000	1.000	—	.811

Table 7: Extended Kalman Smoother (EK) and Importance Resampling Smoother (IR)

— RMS of α_t and $T = 100$ —

Simulation	δ	EK	IR $N = 50$	IR $N = 100$
A	0.5	.706	.718	.720
	0.9	.644	.586	.554
B	0.5	1.161	.941	.926
	0.9	2.266	1.003	.968
C	0.5	.751	.805	.779
	0.9	.692	.801	.750
D	0.5	.523	.547	.535
	0.9	.529	.550	.545

out to be the worst estimator of the four proposal densities, although Proposal Densities I – III still show a good performance. It is well known that the stochastic volatility model has fat tails. As for Proposal Density IV, when the $(i-1)$-th random draw is in the tails, the i-th random draw has an extremely large variance. In this case, Proposal Density IV does not work, compared with the other proposal densities.

Table 3 shows the results of Simulation C (Structural Change Model), where the data generating process is different from the estimated model. We often have this case in practice. For all the proposal densities, the obtained results are very similar. It is natural that AVEs of δ are different from the true values, because the estimated model is not the true one. For δ, therefore, SER (instead of RMS) is shown in this table. All the values are very close to each other. We can see no difference among the proposal densities.

In Simulation D (Shifted-Mean Model) of Table 4, similarly we examine the case where the estimated model is not the true one. The transition equation is assumed to be the first-order autoregressive process and the autoregressive coefficient is estimated by the Bayesian approach. Also, for δ, SERs are shown in the table. As a result, we cannot find any difference among Proposal Densities I – IV, because all the values are similar.

In Table 5, using the nonstationary growth model, the three parameters $(\delta_1, \delta_2, \delta_3)$ are estimated by the Bayesian approach. For δ_1, AVEs are close to the true value and RMSs of I and III are smaller than those of II and IV. For δ_2, AVEs of I, III and IV are close to the true value but those of II are slightly larger and RMSs of II are larger than the others. For δ_3, all the AVEs are underestimated and RMSs are also large. Thus, it is seen that δ_1 and δ_2 are correctly estimated.

As a result, Proposal Densities I and III show a good performance for almost all the simulation studies. Proposal Density IV is also quite good except for the case $\delta = 0.9$ of Simulation B. Proposal Density II is very bad in $\delta = 0.9$ of Simulation A and δ_2 of Simulation E. Thus, Proposal Densities II and IV are inferior to Proposal Densities I and III in the sense that Proposal Densities II and IV sometimes show a bad performance.

Furthermore, as mentioned above, Table 6 represents the acceptance probabilities on average, which are obtained in the Metropolis-Hastings algorithm (see Appendix 1). The acceptance probability equal to zero implies that the chain stays at the same point, i.e., all the random draws generated from the proposal density are discarded. Conversely, the acceptance probability which is equal to one indicates that the proposal density is the same distribution as the target density. From Table 6, Proposal Density IV is close to one, compared with the other proposal densities, because Proposal Density IV approximates the target density. As for Proposal Densities II and III, the acceptance probabilities on average are very different, depending on c. As c is large, the acceptance rate decreases.

Next, in Table 7 we compare the suggested Bayesian procedure with the extended Kalman smoother (EK) and the importance resampling smoother (IR), where given fixed δ the estimates of the state mean are examined in the RMS criterion. See Appendix 2 for IR. Precision of EK depends on nonlinearity and nonnormality of the system, but IR approaches the true state mean as N goes to infinity. However, under the same computational burden as the suggested Bayesian approach, IR shows a very poor performance from Table 7. For each time period, the order of computation is given by N for the Bayesian approach discussed in this paper and N^3 for IR (see Appendix 2). That is, $N = 50$

for IR is more computer-intensive than $N = 10000$ for the Bayes estimator. Remember that $N = 8000$, 10000 is used in Tables 1 – 5. We can find that the Bayes estimator performs much better than IR from Tables 1 – 5 and 7. Sometimes, EK is better than IR (for example, $\delta = 0.5$ of A, C and D). In Table 7, δ is not estimated, where the true values are utilized for Simulations A and B but the cases of $\delta = 0.5, 0.9$ are computed for Simulations C and D. However, the parameter δ is unknown in general and therefore it should be estimated. It is extremely time-consuming for IR to estimate the parameter δ and the state means $\alpha_{1|T}$, $\alpha_{2|T}$, $\cdots$, $\alpha_{T|T}$ simultaneously using the maximum likelihood estimation method, because the iterative procedure such as the Newton-Raphson optimization method has to be taken for estimation of δ, where the i-th iteration of δ is updated after generating all the random draws of the state variables given the parameter value in $(i-1)$-th iteration. Thus, judging from computational cost and estimation of the unknown parameter, the Bayesian approach discussed in this paper is preferred to EK and IR.

5 Summary

Carlin, Polson and Stoffer (1992) and Carter and Kohn (1994, 1996) and Chib and Greenberg (1996) 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 or they considered the case where rejection sampling works well. In this paper, we have shown the nonlinear and non-Gaussian smoother using both Gibbs sampling and the Metropolis-Hastings algorithm, which would be suitable to any nonlinear and non-Gaussian state-space model. Thus, under the Bayesian approach we have introduced the nonlinear and non-Gaussian smoothing procedure in more general formulation than the existing studies.

Moreover, it is known that choice of the proposal density is a critical problem to the Metropolis-Hastings algorithm. In this paper, therefore, several types of the proposal density functions have been investigated. As a result from Monte Carlo studies, for choice of the proposal density, we have obtained the very similar RMSs among Proposal Densities I – IV except for a few cases (i.e., the case $\delta = 0.9$ and small c of Proposal Density II in Simulation A, and the case $\delta = 0.9$ of Proposal Density IV in Simulation B). In such a sense, the proposed procedure is quite robust to choice of the proposal density, but use of the transition equation might be recommended for safety because Proposal Density I has shown a good performance for all the simulation studies examined in this paper.

Appendices

Appendix 1: 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. Let us define $P_*(z|x)$ as the proposal density and the acceptance probability as: $\omega(x,z) = \min\left(\dfrac{P(z)P_*(x|z)}{P(x)P_*(z|x)}, 1\right)$ if $P(x)P_*(z|x) > 0$ and $\omega(x,z) = 1$ otherwise.

Random number generation by 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 (say z) from $P_*(\cdot|x^{(i-1)})$ and a uniform random draw (say u) from the interval between zero and one.

(iii) Set $x^{(i)} = z$ if $u \leq \omega(x^{(i-1)}, z)$ and set $x^{(i)} = x^{(i-1)}$ otherwise.

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

Then, $x^{(N)}$ is taken as a random draw from $P(x)$ for sufficiently large N.

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 $\omega(x,z)$. Furthermore, 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. See, for example, Geweke (1992) and Mengersen, Robert and Guihenneuc-Jouyaux (1999) for the MCMC convergence diagnostics.

Appendix 2: Importance Resampling Smoother (IR)

The density-based recursive algorithm on filtering is given by:

$$P(\alpha_t|Y_{t-1}) = \int P_\alpha(\alpha_t|\alpha_{t-1},\delta)P(\alpha_{t-1}|Y_{t-1})\mathrm{d}\alpha_{t-1}, \tag{11}$$

$$P(\alpha_t|Y_t) = \frac{P_y(y_t|\alpha_t,\gamma)P(\alpha_t|Y_{t-1})}{\displaystyle\int P_y(y_t|\alpha_t,\gamma)P(\alpha_t|Y_{t-1})\mathrm{d}\alpha_t}, \tag{12}$$

where the initial condition is given by:

$$P(\alpha_1|Y_0) = \begin{cases} \displaystyle\int P_\alpha(\alpha_1|\alpha_0,\delta)P(\alpha_0)\mathrm{d}\alpha_0, & \text{if } \alpha_0 \text{ is stochastic,} \\ P_\alpha(\alpha_1|\alpha_0,\delta), & \text{if } \alpha_0 \text{ is nonstochastic.} \end{cases}$$

The density-based recursive algorithm on smoothing utilizes both the one-step ahead prediction density $P(\alpha_{t+1}|Y_t)$ and the filtering density $P(\alpha_t|Y_t)$, which is represented by:

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

for $t = T-1, T-2, \cdots, 1$. Given $P(\alpha_t|Y_t)$ and $P(\alpha_{t+1}|Y_t)$, the smoothing algorithm shown above is a backward recursion from $P(\alpha_{t+1}|Y_T)$ to $P(\alpha_t|Y_T)$.

Let $\alpha_{i,t|s}$ be the i-th random draw of α_t from $P(\alpha_t|Y_s)$. Equation (11) indicates one-step ahead random draw as follows:

$$\alpha_{i,t|t-1} = f_t(\alpha_{i,t-1|t-1}, \eta_{i,t}, \delta), \tag{13}$$

for $i = 1, 2, \cdots, N$. Given $\alpha_{i,t|t-1}$, $i = 1, 2, \cdots, N$, the filtering density (12) is approximated as:

$$P(\alpha_{i,t|t-1}|Y_t) \approx \frac{P_y(y_t|\alpha_{i,t|t-1}, \gamma)}{\sum_{j=1}^N P_y(y_t|\alpha_{j,t|t-1}, \gamma)}, \tag{14}$$

which implies that the one-step ahead prediction random draw $\alpha_{i,t|t-1}$ is taken as a filtering random draw with probability $P(\alpha_{i,t|t-1}|Y_t)$. By resampling, we have $\alpha_{i,t|t}$ for $i = 1, 2, \cdots, N$ and $t = 1, 2, \cdots, T$. Thus, $\alpha_{i,t|t}$ for $i = 1, 2, \cdots, N$ is recursively obtained.

Suppose that $\alpha_{i,t|t}$ and $\alpha_{j,t+1|T}$ are available for $i, j = 1, 2, \cdots, N$. Next, consider generating ($\alpha_{1,t|T}$, $\alpha_{2,t|T}$, $\cdots$, $\alpha_{N,t|T}$) given ($\alpha_{1,t+1|T}$, $\alpha_{2,t+1|T}$, $\cdots$, $\alpha_{N,t+1|T}$). Equation (14) evaluated at $\alpha_t = \alpha_{i,t|t}$ is approximated as:

$$P(\alpha_{i,t|t}|Y_T) \approx \frac{1}{N} \sum_{j=1}^N \frac{P_\alpha(\alpha_{j,t+1|T}|\alpha_{i,t|t}, \delta)}{\sum_{m=1}^N P_\alpha(\alpha_{j,t+1|T}|\alpha_{m,t|t}, \delta)}, \tag{15}$$

where in the denominator of equation (15) the one-step ahead prediction density , i.e., $P(\alpha_{j,t+1|T}|Y_t)$, is also approximately evaluated as:

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

Equation (15) implies that the filtering random draw $\alpha_{i,t|t}$ is taken as a smoothing random draw with probability $P(\alpha_{i,t|t}|Y_T)$.

For time t, we need to compute the two summations with respect to j and m in order to obtain $P(\alpha_{i,t|t}|Y_T)$ for $i = 1, 2, \cdots, N$. Accordingly, in equation (15), the order of computation is given by N^3 for each time period t. For the Markov chain Monte Carlo procedure discussed in this paper, the order of computation is N for time t, which is much less computer-intensive than the resampling procedure in the smoothing algorithm.

References

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

Bollerslev, T., Engle, R.F. and Nelson, D.B., 1994, " ARCH Models, " in *Handbook of Econometrics*, Vol.4, edited by Engle, R.F. and McFadden, D.L., pp.2959 – 3038, 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.

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.

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., 1992, " Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments, " in *Bayesian Statistics, Vol.4*, pp.169 – 193 (with discussion), edited by Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M., Oxford University Press.

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.

Geweke, J. and Tanizaki, H., 1999, " On Markov Chain Monte Carlo Methods for Nonlinear and Non-Gaussian State-Space Models, " *Communications in Statistics, Simulation and Computation*, Vol.28, No.4, pp.867 – 894.

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.

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.

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

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., 2000, " Simulation-Based Inference in Nonlinear State-Space Models: Application to Testing Permanent Income Hypothesis, " in *Simulation-Based Inference in Econometrics: Methods and Applications*, Chap.9, pp.218 – 234, edited by Mariano, R.S., Weeks, M. and Schuermann, T., Cambridge University Press.

Mengersen, K.L., Robert, C.P. and Guihenneuc-Jouyaux, C., 1999, " MCMC Convergence Diagnostics: A Reviewww, " in *Bayesian Statistics, Vol.6*, pp.514 – 440 (with discussion), edited by Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M., Oxford University Press.

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

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.