Estimation of unknown parameters in nonlinear and non-Gaussian state-space models

Tanizaki, Hisashi

(Citation)
Journal of Statistical Planning and Inference,96(2):301-323

(Issue Date)
2001-07

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90000133

ESTIMATION OF UNKNOWN PARAMETERS IN NONLINEAR AND NON-GAUSSIAN STATE SPACE MODELS*

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

Key Words: *State-Space Model, Metropolis-Hastings Algorithm, Monte Carlo Optimization, Simulated Annealing, Nonparametric Density Estimation, Mode Estimation.*

AMS Subject Classification: 62M10, 62M20.

ABSTRACT: For the last decade, various simulation-based nonlinear and non-Gaussian filters and smoothers have been proposed. In the case where the unknown parameters are included in the nonlinear and non-Gaussian system, however, it is very difficult to estimate the parameters together with the state variables, because the state-space model includes a lot of parameters in general and and the simulation-based procedures are subject to the simulation errors or the sampling errors. Therefore, clearly, precise estimates of the parameters cannot be obtained (i.e., the obtained estimates may not be the global optima). In this paper, an attempt is made to estimate the state variables and the unknown parameters simultaneously, where the Monte Carlo optimization procedure is adopted for maximization of the likelihood function.

*The author would like to acknowledge the constructive comments of the anonymous referee. This research was partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Encouragement of Young Scientists (# 12730022).

1 INTRODUCTION

Since Kitagawa (1987) and Kramer and Sorenson (1988) proposed the nonlinear and non-Gaussian filter and smoother using numerical integration, the filtering and smoothing solutions have been developed, because it is difficult to apply numerical integration to the higher dimensional cases of the state vector from computational point of view. Tanizaki (1993) and Tanizaki and Mariano (1994) suggested using Monte-Carlo integration to solve filtering and smoothing problems in nonlinear and non-Gaussian cases. Carlin, Polson and Stoffer (1992), Carter and Kohn (1994, 1996) and Chib and Greenberg (1996) applied the Gibbs sampler to evaluate the smoothing means in a Bayesian framework, which smoother is extended to more general cases using both the Gibbs sampler and the Metropolis-Hastings algorithm simultaneously in Geweke and Tanizaki (1999). Furthermore, to obtain the filtering and smoothing solutions, Gordon, Salmond and Smith (1993), Kitagawa (1996) and Kitagawa and Gersch (1996) proposed the resampling filter and smoother. In Tanizaki (1996), Tanizaki and Mariano (1998) and Tanizaki (1999), rejection sampling is utilized to generate random draws directly from the filtering and smoothing densities. See Tanizaki (2000) for a survey on the simulation-based nonlinear and non-Gaussian state-space models.

Thus, various kinds of nonlinear and non-Gaussian filters have been investigated. However, in the case where unknown parameters are included in the system, estimation of the parameters are still an open question. Generally there are a lot of unknown parameters to be estimated in the state space models.[1] The traditional maximization procedures such as the simple grid search and the Newton-Raphson optimization are conventionally used for maximization of the likelihood function. However, if the number of the parameters increases, these procedures are not helpful.

Moreover, as Kitagawa (1998) pointed out, we encounter the following two problems: (i) the non-Gaussian smoother is computationally intensive and the repeated application of a numerical optimization procedure for evaluating the likelihood may make it almost impractical and (ii) the log-likelihood computed by the Monte Carlo filter is subject to a sampling error and accordingly precise maximum likelihood parameter estimates can be obtained only by using a very large number of particles or by parallel application of many Monte Carlo filters. To improve the two problems, Kitagawa (1998) proposed estimating the state vectors and the unknown parameters simultaneously, which is called

[1] For simplicity of discussion, consider the linear and normal state-space model given by $y_t = x_t\alpha_t + \epsilon_t$ and $\alpha_t = \Phi\alpha_{t-1} + \eta$, where $\epsilon_t \sim N(0, \Sigma_\epsilon)$ and $\eta \sim N(0, \Sigma_\eta)$. Let ϵ_t, α_t and η_t be $p \times 1$, $k \times 1$ and $k \times 1$ vectors, respectively. Then, the number of unknown parameters to be estimated is $p(p+1)/2$ for Σ_ϵ, k^2 for Φ and $k(k+1)/2$ for Σ_η. Thus, as p or k increases, the number of unknown parameters drastically increases.

the self-organizing filter and smoother in his paper. A vector of the unknown parameters is regarded as the state vector which movement is appropriately assumed by a researcher through the transition equation. One of the disadvantages in Kitagawa (1998) is that we need a large number of the sample size to have the stable estimates of the unknown parameters. In this paper, we consider the estimation method such that we can apply even in the small sample.

Conventionally, as for maximization, the derivative-based methods such as the Newton-Raphson optimization procedure are used, where the iterative procedure is performed to obtain an optimum. The derivative-based methods generally have a problem that there is no method of distinguishing between a local optimum and a global one. It is well known that the method of the simulated annealing can find the best among multiple solutions (see, for example, Kirkpatrick, Gelatt, Jr. and Vecchi (1983), Bohachevsky, Johnson and Stein (1986), Goffe, Ferrier and Rogers (1994) and Brooks and Morgan (1995)). In the case where the function to be maximized is known, the simulated annealing is a powerful tool. For the nonlinear and non-Gaussian state-space models, however, we cannot obtain the explicit functional form of the likelihood function. Therefore, the simulated annealing suggested by Kirkpatrick, Gelatt, Jr. and Vecchi (1983) cannot be directly applied to our problem. In this paper, the likelihood function is regarded as a kernel of the density function with respect to the unknown parameters. Based on the likelihood function, random draws of the state vectors and those of the unknown parameters are generated, where Gibbs sampling and the Metropolis-Hastings algorithm may be taken for random number generation. Note that for random number generation it might be possible to use the other sampling techniques such as rejection sampling instead of the Metropolis-Hastings algorithm. See Geman and Geman (1984), Tanner and Wong (1987), Gelfand and Smith (1990), Arnold (1993), Boswell, Gore, Patil and Taillie (1993), Smith and Roberts (1993), O'Hagan (1994), Tierney (1994), Chib and Greenberg (1995), Geweke (1996, 1997) for the sampling techniques. Through the random draws of the unknown parameters, we numerically obtain the functional form of the likelihood function with respect to the unknown parameters, where the nonparametric density estimation technique is applied. Based on the nonparametric density estimation, for each unknown parameter the mode is obtained, which corresponds to the global maximum, i.e., the maximum likelihood estimate of the parameter. See Prakasa Rao (1983), Silverman (1986), Ullah (1988), Härdle (1990) and Izenman (1991) for the nonparametric density estimation. Thus, in this paper, we propose estimating the state vectors and the unknown parameters simultaneously, using the Monte Carlo optimization procedure.

2 STATE SPACE MODEL

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

$$\text{(Measurement Equation)} \quad y_t = q_t(\alpha_t, \epsilon_t; \gamma), \tag{1}$$

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

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. $q_t(\cdot, \cdot; \cdot)$ and $f_t(\cdot, \cdot; \cdot)$ are vector functions, which are assumed to be known. γ and δ indicate the unknown parameters to be estimated. The error terms ϵ_t and η_t are mutually independently distributed, which are typically assumed to be normal but not necessarily. Let Y_s be the information set up to time s, i.e., $Y_s = \{y_1, y_2, \cdots, y_s\}$. Define $p_y(y_t|\alpha_t; \gamma)$ and $p_\alpha(\alpha_t|\alpha_{t-1}; \delta)$ as the density functions obtained from the measurement equation (1) and the transition equation (2). We consider estimating α_t using information Y_s. $a_{t|s} \equiv \mathrm{E}(\alpha_t|Y_s)$ is called prediction if $t > s$, filtering if $t = s$ and smoothing $t < s$, respectively.

Carlin, Polson and Stoffer (1992), Carter and Kohn (1994, 1996) and Chib and Greenberg (1996) introduced the nonlinear non-Gaussian state-space models with Gibbs sampling (see Appendix 1 for the Gibbs sampler). They investigated the nonlinear state-space models only in the Bayesian framework and in addition they studied the special state-space models such that it is easy to generate random draws from the underlying assumptions. To improve these problems, Geweke and Tanizaki (1999) proposed 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 models. See Appendix 2 for the Metropolis-Hastings algorithm. The nonlinear and non-Gaussian smoother proposed by Geweke and Tanizaki (1999) is briefly discussed as follows.

Smoothing: Let us define $A_t = \{\alpha_0, \alpha_1, \cdots, \alpha_t\}$ and $A_t^* = \{\alpha_t, \alpha_{t+1}, \cdots, \alpha_T\}$, respectively. We consider generating random draws of A_T from $p(A_T|Y_T; \gamma, \delta)$. 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 $t = 1, 2, \cdots, T$. Let us define $p_y(Y_T|A_T; \gamma)$ and $p_\alpha(A_T; \delta)$ as:

$$p_y(Y_T|A_T; \gamma) = \prod_{t=1}^{T} p_y(y_t|\alpha_t; \gamma),$$

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

Then, $p(\alpha_t|A_{t-1}, A^*_{t+1}, Y_T; \gamma, \delta)$ is given by the following conditional distribution function:

$$
\begin{aligned}
&p(\alpha_t|A_{t-1}, A^*_{t+1}, Y_T; \gamma, \delta) \\
&\quad = \frac{p_y(Y_T|A_T; \gamma)p_\alpha(A_T; \delta)}{\int p_y(Y_T|A_T; \gamma)p_\alpha(A_T; \delta)\mathrm{d}\alpha_t} \\
&\quad \propto \begin{cases} p_y(y_t|\alpha_t; \gamma)p_\alpha(\alpha_t|\alpha_{t-1}; \delta)p_\alpha(\alpha_{t+1}|\alpha_t; \delta), & \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{for } t = 1, 2, \cdots, T-1, & \\ p_y(y_t|\alpha_t; \gamma)p_\alpha(\alpha_t|\alpha_{t-1}; \delta), \qquad \text{for } t = T \text{ (i.e., endpoint)}, & \end{cases}
\end{aligned} \tag{3}
$$

Equation (3) implies that a kernel of $p(\alpha_t|A_{t-1}, A^*_{t+1}, Y_T; \gamma, \delta)$ is given by $p_y(y_t|\alpha_t; \gamma)\ p_\alpha(\alpha_t|\alpha_{t-1}; \delta)\ p_\alpha(\alpha_{t+1}|\alpha_t; \delta)$ when $t = 1, 2, \cdots, T-1$ and $p_y(y_t|\alpha_t; \gamma)\ p_\alpha(\alpha_t|\alpha_{t-1}; \delta)$ when $t = T$ (i.e., endpoint). Utilizing the Metropolis-Hastings algorithm, random draws of α_t from $p(\alpha_t|A_{t-1}, A^*_{t+1}, Y_T; \gamma, \delta)$ are generated. Define $A_{i,t}$ and $A^*_{i,t}$ as $A_{i,t} = \{\alpha_{i,0}, \alpha_{i,1}, \cdots, \alpha_{i,t}\}$ and $A^*_{i,t} = \{\alpha_{i,t}, \alpha_{i,t+1}, \cdots, \alpha_{i,T}\}$, respectively, which are the i-th random draws of A_t and A^*_t. Given (Y_T, γ, δ), we can generate random draws of A_T as follows.

(i) Take an appropriate initial value for $A_{0,T}$.

(ii) Given ($A_{i,t-1}$, $A^*_{i-1,t+1}$, Y_T, γ, δ), generate a random draw of α_t, denoted by $\alpha_{i,t}$, from $p(\alpha_t|A_{i,t-1}, A^*_{i-1,t+1}, Y_T; \gamma, \delta)$, where the Metropolis-Hastings algorithm is utilized in the case where it is intractable to generate random draws from $p(\alpha_t|A_{i,t-1}, A^*_{i-1,t+1}, Y_T; \gamma, \delta)$.

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

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

The first M random draws are excluded from consideration, taking into account stability of the Gibbs sampler and the Metropolis-Hastings algorithm. In Section 4, $M = 0.2N$ is taken. Thus, the random draws of A_T given Y_T, γ and δ, i.e., $A_{i,T}$, can be obtained using both the Gibbs sampler and the Metropolis-Hastings algorithm.

Mean and Variance: Based on $A_{i,T}$, $\mathrm{E}(m(\alpha_t)|Y_T)$ is simply evaluated as the arithmetic average of $m(\alpha_{i,t})$, $i = 1, 2, \cdots, N$, which is represented by:

$$
\mathrm{E}(m(\alpha_t)|Y_T) \approx \frac{1}{N-M} \sum_{i=M+1}^{N} m(\alpha_{i,t}) \equiv \overline{m(\alpha_t)}
$$

where $m(\cdot)$ is a function. Especially, the case of $m(\alpha_t) = \alpha_t$ represents the smoothing mean, while $m(\alpha_t) = (\alpha_t - a_{t|T})(\alpha_t - a_{t|T})'$ indicates the smoothing

variance. From the basic result of the Metropolis within Gibbs sampling, we can obtain the following result:

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

for all $t = 1, 2, \cdots, T$.

Likelihood Function: In the case where equations (1) and (2) depends on the unknown parameters γ and δ, the likelihood function to be maximized is written as:

$$\begin{aligned} p(Y_T;\gamma,\delta) &= \int p(Y_T, A_T;\gamma,\delta)\mathrm{d}A_T \\ &= \int p_y(Y_T|A_T;\gamma)p_\alpha(A_T;\delta)\mathrm{d}A_T, \end{aligned} \tag{4}$$

where $p(Y_T, A_T;\gamma,\delta) \equiv p_y(Y_T|A_T;\gamma)p_\alpha(A_T;\delta)$ is defined. The difficulty in this maximization is that evaluating the integration above we need to maximize equation (4) with respect to γ and δ. In the simulation-based procedures, the integration above is evaluated by random numbers of A_T and accordingly the sampling errors are always included. In the next section, we suggest maximizing the likelihood function (4) by a Monte Carlo optimization method.

3 MAXIMUM LIKELIHOOD ESTIMATION BY MONTE CARLO OPTIMIZATION

In general, $p(Y_T;\gamma,\delta)$ in equation (4) is not obtained in an explicit functional form, because the integration in the right hand side of equation (4) cannot be evaluated analytically in the case where the system is nonlinear and non-Gaussian. Therefore, conventionally, the iterative procedure is applied for maximization of the likelihood function. That is, (i) given the fixed values for γ and δ, the likelihood function $p(Y_T;\gamma,\delta)$ is numerically estimated by evaluating the integration, (ii) based on the estimated likelihood function, the estimates of γ and δ are updated and (iii) we repeat (i) and (ii) until all the estimates are stable. In (ii), the simple grid search or the Newton-Raphson optimization procedure may be utilized to maximize the estimated likelihood function.

As an alternative optimization method, Wei and Tanner (1990), Tanner (1993) and Chan and Ledolter (1995) proposed the Monte Carlo EM algorithm to maximize the likelihood function. In the case where the Monte Carlo EM algorithm is applied to estimation of γ and δ, (i) and (ii) are modified as follows. In (i), the conditional expectation of $p(Y_T, A_T;\gamma,\delta)$ given Y_T is evaluated through the random draws of A_T given Y_T (i.e., $A_{i,T}$). In (ii), the

expectation of $p(Y_T, A_T; \gamma, \delta)$ is maximized with respect to γ and δ. In any case, since the iterative procedure is taken for estimation of γ and δ, a great amount of computation is necessary for the nonlinear and non-Gaussian filtering techniques and accordingly implementation sometimes becomes impossible. In this paper, we suggest applying a very simple optimization to the nonlinear and non-Gaussian filtering problem.

3.1 Monte Carlo Optimization

Consider maximizing a function $g(x)$ with respect to x. As it is well-known, there is no guarantee that the traditional maximization methods such as the Newton-Raphson optimization procedure give us the global optimum. The Monte Carlo optimization methods such as the simulated annealing are the useful tools which find the global optimum. Here, we apply the Metropolis-Hastings algorithm to obtain the global optimum, where $g(x)$ is regarded as a kernel of a density function (i.e., $p(x) = g(x)/\int g(x)\mathrm{d}x$ where $p(x)$ denotes the density function of x). Define the acceptance probability $\omega(x, z)$ as:

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

where $p_*(z|x)$ represents the sampling density function, which has to be appropriately specified by a researcher. Given x_{i-1}, we generate a random draw z from $p_*(\cdot|x_{i-1})$ and set $x_i = z$ with probability $\omega(x_{i-1}, z)$ and $x_i = x_{i-1}$ otherwise. Repeating the procedure for $i = 1, 2, \cdots, N$, we obtain x_i, $i = 1, 2, \cdots, N$.

We have two strategies to obtain the value of x which maximizes the function $g(x)$. For the first strategy, the function $g(x)$ is evaluated at $x = x_i$ for all $i = 1, 2, \cdots, N$ and the maximum value of $g(x)$ can be obtained out of $g(x_1)$, $g(x_2)$, $\cdots$, $g(x_N)$. For the second strategy, since we regard $g(x)$ as a kernel of a density function, we estimate the functional form of $p(x)$ by nonparametric density estimation, based on the random draws x_1, x_2, $\cdots$, x_N. See Appendix 3 for the nonparametric density estimation method. Then, the mode obtained numerically from the estimated density function of $p(x)$, say $\widehat{p}(x)$, can be taken as the value of x which maximizes the function $g(x)$. It is known that $\widehat{p}(x)$ converges to $p(x)$ but the convergence speed is at most $O(N^{1/5})$, which is very slow (see Appendix 3). We can consider that the estimated mode also converges at the same speed. Note in Section 2 that x and $g(x)$ correspond to (γ, δ) and $p(Y_T; \gamma, \delta)$, respectively.

In the case where $g(x)$ is not analytically obtained in an explicit functional form, the second strategy is useful, which utilizes the empirical density function. In this paper, since the likelihood function $p(Y_T; \gamma, \delta)$ is not analytically obtained, we apply the second strategy to the nonlinear and non-Gaussian filtering and smoothing problem.

3.2 Maximum Likelihood Estimation

The unobserved variable, which is called the state variable, is included in the state-space model. Therefore, in practice, it is very difficult to evaluate the integration in equation (4) together with maximizing equation (4) with respect to the unknown parameters, because of the sampling errors and the computational disadvantages. The second strategy discussed in Section 3.1 is a helpful tool to this problem.

Let us define $\theta \equiv (\gamma, \delta)$. Note in equation (4) that maximizing the marginal density of Y_T with respect to θ is equivalent to maximizing the joint density of Y_T and A_T with respect to θ, because clearly we have the following equality:

$$\frac{\partial \int p(Y_T, A_T; \theta) \mathrm{d}A_T}{\partial \theta} = \int \frac{\partial p(Y_T, A_T; \theta)}{\partial \theta} \mathrm{d}A_T,$$

which implies that the solution of θ obtained by solving $\max_{\theta} \int p(Y_T, A_T; \theta) \mathrm{d}A_T$ is equivalent to that by $\max_{\theta} p(Y_T, A_T; \theta)$. Accordingly, the maximum value of θ obtained from the joint density $p(Y_T, A_T; \theta)$ is equal to that from the marginal density $p(Y_T; \theta)$. The maximization problem in this paper implies that the joint density of Y_T and A_T can be maximized with respect to θ.

As mentioned in Section 3.1, the proportions of $p(Y_T, A_T; \gamma, \delta)$ in the integration of equation (4) can be regarded as the density functions of γ and δ, i.e.,

$$p(\gamma | A_T, Y_T, \delta) \propto p(Y_T, A_T; \gamma, \delta) \propto p_y(Y_T | A_T; \gamma), \tag{5}$$

$$p(\delta | A_T, Y_T, \gamma) \propto p(Y_T, A_T; \gamma, \delta) \propto p_\alpha(A_T; \delta). \tag{6}$$

Using (3), (5) and (6), we generate random draws of A_T, γ and δ simultaneously from $p(Y_T, A_T; \gamma, \delta)$ through the Gibbs sampler. As for the random number generation of α_t from (3), that of γ from (5) and that of δ from (6), importance resampling, rejection sampling or the Metropolis-Hastings algorithm may be adopted in the case where it is not easy to generate the random draws.

Note as follows. In equation (5), random draws of γ from $p(\gamma | A_T, Y_T, \delta)$ might be based on $p(\gamma_1 | A_T, Y_T, \delta, \gamma_2)$ and $p(\gamma_2 | A_T, Y_T, \delta, \gamma_1)$ through the Gibbs sampler, where $\gamma = (\gamma_1, \gamma_2)$. Similarly, in equation (6), it is possible to obtain random draws of δ by blocking $\delta = (\delta_1, \delta_2)$.

Based on the random draws of γ and δ, we construct the density functions by nonparametric density estimation and obtain the modes from the empirical densities of γ and δ, which are denoted by $\widehat{\gamma}$ and $\widehat{\delta}$, respectively. The estimated modes (i.e., $\widehat{\gamma}$ and $\widehat{\delta}$) correspond to the maximum likelihood estimates of γ and δ. Therefore, in order to compute the smoothing estimate of A_T (i.e., $a_{t|T}$ for $t = 1, 2, \cdots, T$), given $\widehat{\gamma}$ and $\widehat{\delta}$, again we generate random draws of A_T from $p(\alpha_t | A_{t-1}, A^*_{t+1}, Y_T; \widehat{\gamma}, \widehat{\delta})$, $t = 1, 2, \cdots, T$. That is, replacing (γ, δ) by $(\widehat{\gamma}, \widehat{\delta})$, we

can perform the procedures (i) – (iv) discussed in Section 2. Accordingly, the procedure suggested above indicates the two-step estimator, where the first step obtains the modes based on the empirical densities of γ and δ and the second step computes the state estimates given the estimated modes $(\widehat{\gamma}, \widehat{\delta})$. As discussed in Section 3.1, since the empirical densities of γ and δ approach the true underlying densities, $\widehat{\gamma}$ and $\widehat{\delta}$ go to the true parameter values. However, we should keep in mind that the convergence speed is quite slow.

The correspondence between the maximum likelihood estimate and the Bayesian estimate is given as follows. As mentioned above, random draws of A_T, γ and δ are simultaneously generated from the three density functions (3), (5) and (6) through the Gibbs sampler and the Metropolis-Hastings algorithm. The densities (5) and (6) reduce to the posterior densities given the diffuse priors. Therefore, the arithmetic averages obtained from the random draws of γ and δ correspond to the Bayesian estimates, while the modes are the maximum likelihood estimates.

4 MONTE CARLO STUDIES

In this section, we compare the maximum likelihood estimator suggested in Section 3 and the Bayesian estimator with the diffuse priors, utilizing the state-space models shown in Simulations I – IV.

Simulation I (Linear and Normal Model): Consider the scalar system: $y_t = \alpha_t + \epsilon_t$ and $\alpha_t = \beta\alpha_{t-1} + \eta_t$. The initial value α_0 and the error terms ϵ_t and η_t, $t = 1, 2, \cdots, T$, are assumed to be distributed as:

$$\alpha_0 \sim N(0,1), \quad \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\epsilon^2 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix} \right).$$

$\beta = 0.5, 0.9, 1.0$ and $\sigma_\epsilon^2 = \sigma_\eta^2 = 1$ are taken for the true parameter values.

For Gibbs sampling, $p(\alpha_t|A_{t-1},\ A_{t+1}^*,\ Y_T; \beta\ , \sigma_\eta^2,\ \sigma_\epsilon^2)$, $p(\beta|A_T,\ Y_T; \sigma_\eta^2,\ \sigma_\epsilon^2)$, $p(\sigma_\eta^2|A_T,\ Y_T; \beta,\ \sigma_\epsilon^2)$ and $p(\sigma_\epsilon^2|A_T,\ Y_T; \beta,\ \sigma_\eta^2)$ are obtained as:

$$\begin{aligned} &p(\alpha_t|A_{t-1}, A_{t+1}^*, Y_T; \beta, \sigma_\eta^2, \sigma_\epsilon^2) \\ &\quad \propto \begin{cases} \exp\Big(-\dfrac{1}{2\sigma_\epsilon^2}(y_t - \alpha_t)^2 - \dfrac{1}{2\sigma_\eta^2}(\alpha_t - \beta\alpha_{t-1})^2 \\ \qquad -\dfrac{1}{2\sigma_\eta^2}(\alpha_{t+1} - \beta\alpha_t)^2\Big), & \text{for } t = 1, 2, \cdots, T-1, \\ \exp\Big(-\dfrac{1}{2\sigma_\epsilon^2}(y_t - \alpha_t)^2 - \dfrac{1}{2\sigma_\eta^2}(\alpha_t - \beta\alpha_{t-1})^2\Big), & \text{for } t = T, \end{cases} \end{aligned} \tag{7}$$

$$p(\beta|A_T, Y_T; \sigma_\eta^2, \sigma_\epsilon^2) \propto \exp\Big(-\frac{1}{2\sigma_\eta^2}\sum_{t=1}^{T}(\alpha_t - \beta\alpha_{t-1})^2\Big), \tag{8}$$

$$p(\sigma_\eta^2|A_T, Y_T; \beta, \sigma_\epsilon^2) \propto \Big(\frac{1}{\sigma_\eta^2}\Big)^{T/2} \exp\Big(-\frac{1}{2\sigma_\eta^2}\sum_{t=1}^{T}(\alpha_t - \beta\alpha_{t-1})^2\Big), \tag{9}$$

$$p(\sigma_\epsilon^2|A_T, Y_T; \beta, \sigma_\eta^2) \propto \Big(\frac{1}{\sigma_\epsilon^2}\Big)^{T/2} \exp\Big(-\frac{1}{2\sigma_\epsilon^2}\sum_{t=1}^{T}(y_t - \alpha_t)^2\Big). \tag{10}$$

Equations (7) and (8) reduce to normal distributions and equations (9) and (10) can be converted to chi-square distributions with $T+2$ degrees of freedom. Accordingly, the random draws from the above density functions are easily generated.

Simulation II (ARCH Model): We consider the state-space model: $y_t = \alpha_t + \epsilon_t$ and $\alpha_t = (\beta_0 + \beta\alpha_{t-1}^2)^{1/2}\eta_t$ for $\beta_0 > 0$ and $0 \leq \beta < 1$. In this simulation study, $\beta_0 = 1 - \beta$ is taken. $\beta_0 = 1 - \beta$ in the transition equation implies that the unconditional variance of α_t is normalized to be one. The initial value α_0 and the error terms ϵ_t and η_t, $t = 1, 2, \cdots, T$, are assumed to be distributed as:

$$\alpha_0 \sim N(0,1), \quad \begin{pmatrix}\epsilon_t \\ \eta_t\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}\sigma_\epsilon^2 & 0\\ 0 & 1\end{pmatrix}\right).$$

$\beta = 0.5, 0.9$ and $\sigma_\epsilon^2 = 1$ are taken. The transition equation indicates the first-order ARCH process, while the measurement equation consists of the ARCH process and the error term. See Engle (1982) and Bollerslev, Engle and Nelson (1994) for the ARCH model.

To implement Gibbs sampling, $p(\alpha_t|A_{t-1},\ A_{t+1}^*,\ Y_T; \beta,\ \sigma_\epsilon^2)$ and $p(\beta|A_T,\ Y_T; \sigma_\epsilon^2)$ are represented as:

$$p(\alpha_t|A_{t-1}, A_{t+1}^*, Y_T; \beta, \sigma_\epsilon^2)$$
$$\propto \begin{cases} \dfrac{1}{(\beta_0 + \beta\alpha_t^2)^{1/2}} \exp\Big(-\dfrac{1}{2\sigma_\epsilon^2}(y_t - \alpha_t)^2 - \dfrac{\alpha_t^2}{2\sigma_\eta^2(\beta_0 + \beta\alpha_{t-1}^2)} \\ \qquad\qquad -\dfrac{\alpha_{t+1}^2}{2\sigma_\eta^2(\beta_0 + \beta\alpha_t^2)}\Big), \quad \text{for } t = 1, 2, \cdots, T-1, \\ \exp\Big(-\dfrac{1}{2\sigma_\epsilon^2}(y_t - \alpha_t)^2 - \dfrac{\alpha_t^2}{2\sigma_\eta^2(\beta_0 + \beta\alpha_{t-1}^2)}\Big), \quad \text{for } t = T, \end{cases} \tag{11}$$

$$p(\beta|A_T, Y_T; \sigma_\epsilon^2) \propto \frac{1}{\prod_{t=1}^{T}(\beta_0 + \beta\alpha_{t-1}^2)^{1/2}} \exp\Big(-\frac{1}{2}\sum_{t=1}^{T}\frac{\alpha_t^2}{\beta_0 + \beta\alpha_{t-1}^2}\Big), \tag{12}$$

where $\beta_0 = 1 - \beta$ and $0 \leq \beta < 1$ by the assumptions. Since $p(\sigma_\epsilon^2|A_T,\ Y_T; \beta)$ is given by equation (10), random draws of σ_ϵ^2 are easily obtained. However, it is difficult to generate random numbers from $p(\alpha_t|A_{t-1},\ A_{t+1}^*,\ Y_T; \beta,\ \sigma_\epsilon^2)$ and $p(\beta|A_T,\ Y_T; \sigma_\epsilon^2)$. In this case, as discussed in Section 3.1, the Metropolis-Hastings algorithm is helpful, where the proposal densities are taken as $p_*(\alpha_t) = p_\alpha(\alpha_t|\alpha_{t-1}; \delta)$ and $p_*(\beta) = U(0,1)$, respectively (note that $U(a,b)$ denotes the uniform distribution between a and b).

Simulation III (Stochastic Volatility Model): Next, we take the following stochastic volatility model: $y_t = \exp(\alpha_t/2)\epsilon_t$ and $\alpha_t = \beta\alpha_{t-1} + \eta_t$ for $0 \leq \beta < 1$. The initial value α_0 and the error terms ϵ_t and η_t, $t = 1, 2, \cdots, T$, are assumed to be distributed as:

$$\alpha_0 \sim N(0,1), \quad \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix} \right).$$

We consider the cases of $\beta = 0.5, 0.9$ and $\sigma_\eta^2 = 1$. See Ghysels, Harvey and Renault (1996) for the stochastic volatility model.

In this simulation study, $p(\beta|A_T,\ Y_T; \sigma_\eta^2)$ is equivalent to equation (8) with the constraint $0 \leq \beta < 1$ while $p(\sigma_\eta^2|A_T,\ Y_T; \beta)$ is given by equation (9). Therefore, $p(\beta|A_T,\ Y_T; \sigma_\eta^2)$ reduces to the truncated normal distribution function. Since it is difficult to obtain a random draw from a truncated normal distribution in the case where the normal distribution is truncated around tails (see Geweke (1991)), in this simulation study we apply the Metropolis-Hastings algorithm to random number generation, where $p_*(\beta) = U(0,1)$ is utilized for the proposal density.

Simulation IV (Logistic Model): The following logistic type of nonlinear state-space model is taken as follows: $y_t = \exp(\alpha_t)/(\exp(\alpha_t) + \exp(\epsilon_t))$ and $\alpha_t = \exp(\alpha_{t-1})/(\exp(\alpha_{t-1}) + \exp(\eta_t))$. The initial value α_0 and the error terms ϵ_t and η_t, $t = 1, 2, \cdots, T$, are assumed to be distributed as:

$$\alpha_0 \sim U(0,1), \quad \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\epsilon^2 & 0 \\ 0 & \sigma_\eta^2 \end{pmatrix} \right).$$

In this simulation study, $\sigma_\epsilon^2 = \sigma_\eta^2 = 1$ is examined for the true parameter values. Tanizaki (1993, 1996) took the logistic model as an example.

In order to perform Gibbs sampling, the conditional density functions $p(\alpha_t|A_{t-1},\ A_{t+1}^*,\ Y_T; \sigma_\eta^2,\ \sigma_\epsilon^2)$, $p(\sigma_\eta^2|A_T,\ Y_T; \sigma_\epsilon^2)$ and $p(\sigma_\epsilon^2|A_T,\ Y_T; \sigma_\eta^2)$ are represented as:

$$p(\alpha_t|A_{t-1}, A_{t+1}^*, Y_T; \sigma_\eta^2, \sigma_\epsilon^2)$$

$$\propto \begin{cases} \dfrac{1}{(1-\alpha_t)\alpha_t} \exp\Big(-\dfrac{1}{2\sigma_\epsilon^2}(\log(\dfrac{1}{y_t} - 1) + \alpha_t)^2 \\ \quad -\dfrac{1}{2\sigma_\eta^2}(\log(\dfrac{1}{\alpha_t} - 1) + \alpha_{t-1})^2 - \dfrac{1}{2\sigma_\eta^2}(\log(\dfrac{1}{\alpha_{t+1}} - 1) + \alpha_t)^2\Big), \\ \qquad\qquad\qquad\qquad\qquad \text{for } t = 1, 2, \cdots, T-1, \\ \dfrac{1}{(1-\alpha_t)\alpha_t} \exp\Big(-\dfrac{1}{2\sigma_\epsilon^2}(\log(\dfrac{1}{y_t} - 1) + \alpha_t)^2 \\ \qquad\qquad -\dfrac{1}{2\sigma_\eta^2}(\log(\dfrac{1}{\alpha_t} - 1) + \alpha_{t-1})^2\Big), \qquad \text{for } t = T, \end{cases} \tag{13}$$

$$p(\sigma_\eta^2|A_T, Y_T; \sigma_\epsilon^2) \propto (\frac{1}{\sigma_\eta^2})^{T/2} \exp\Big(-\frac{1}{2\sigma_\eta^2}\sum_{t=1}^{T}(\log(\frac{1}{\alpha_t} - 1) + \alpha_{t-1})^2\Big), \tag{14}$$

$$p(\sigma_\epsilon^2|A_T, Y_T; \sigma_\eta^2) \propto (\frac{1}{\sigma_\epsilon^2})^{T/2} \exp\Big(-\frac{1}{2\sigma_\epsilon^2}\sum_{t=1}^{T}(\log(\frac{1}{y_t}-1)+\alpha_t)^2\Big). \tag{15}$$

Since equations (14) and (15) can be transformed into the chi-square distribution with $T+2$ degrees of freedom, both σ_η^2 and σ_ϵ^2 are easily generated. However, it is not easy to generate a random draw of α_t from equation (13) and accordingly the Metropolis-Hastings algorithm can be applied to random number generation.

Simulation Procedure: For both the maximum likelihood estimator and the Bayesian estimator, we compare the estimates of the unknown parameters (i.e., β, σ_ϵ^2 and σ_η^2). Note that the correspondence between (γ, δ) in the system (1) and (2) and (β, σ_ϵ^2, σ_η^2) in the above simulation studies are: $\gamma = (\sigma_\epsilon^2)$ and $\delta = (\beta, \sigma_\eta^2)$. The simulation procedure is as follows: (i) generating random numbers of ϵ_t and η_t for $t = 1, 2, \cdots, T$, we obtain a set of data y_t and α_t, $t = 1, 2, \cdots, T$, from equations (1) and (2), where $T = 40, 100, 200$ is taken, (ii) given Y_T, perform each estimator, (iii) repeat (i) and (ii) G times, where $G = 1000$ is taken in this paper, and (iv) compute the arithmetic average (AVE), the root mean square error (RMS) and the 0.25th, 0.50th and 0.75th quantiles (25%, 50% and 75%) based on the G estimates for each estimator.

Results and Discussion: The results are in Tables 1 – 9, where N denotes the number of random draws. Usually, the first M random draws are excluded from consideration because of stability of the random draws. In this paper, we take M as 20% of N. Therefore, all the values in the tables are obtained from $0.8N$ random numbers. We take the cases of $N = 2000$, 5000, 10000. Tables 1, 3, 5 and 7 represent the results from the maximum likelihood estimation suggested in this paper, while Tables 2, 4, 6 and 8 are obtained from the Bayesian estimation with the diffuse priors.

In Simulation I, a normal and linear state-space model is examined (see Tables 1 and 2). In Tables 1 and 2, there is not significant difference for $N = 2000$, 5000, 10000. That is, the number of random draws does not affect precision of the estimates. AVE indicates the bias between the true parameter value and the estimated one. As the sample size T increases, AVE's approach the true parameter values and RMS's decrease. In this simulation study, for almost all the cases, the maximum likelihood estimates are less biased than the Bayesian estimates.

In Simulation II (Tables 3 and 4), the maximum likelihood estimator is better than the Bayesian estimator only in the case of $\widehat{\sigma}_\epsilon^2$ in $\beta = 0.5$. In the other cases, the latter performs better than the former with respect to AVE and RMS criteria. However, in both Simulations III (Tables 5 and 6) and IV (Tables 7 and 8), the former is better than the latter for almost all the cases.

Table 1: Simulation I — Maximum Likelihood Estimation

β		T	40			100			200		
		N	2000	5000	10000	2000	5000	10000	2000	5000	10000
0.5	$\widehat{\beta}$	AVE	0.444	0.443	0.442	0.467	0.468	0.468	0.474	0.467	0.467
		RMS	0.233	0.233	0.235	0.181	0.179	0.181	0.156	0.154	0.153
		25%	0.307	0.295	0.292	0.353	0.349	0.351	0.358	0.352	0.355
		50%	0.461	0.463	0.455	0.446	0.454	0.447	0.445	0.440	0.438
		75%	0.599	0.595	0.600	0.573	0.569	0.575	0.574	0.577	0.564
	$\widehat{\sigma}_\eta^2$	AVE	0.778	0.753	0.732	0.928	0.886	0.864	1.031	0.993	0.983
		RMS	0.586	0.588	0.598	0.574	0.553	0.560	0.551	0.531	0.511
		25%	0.312	0.296	0.273	0.438	0.415	0.397	0.588	0.555	0.569
		50%	0.641	0.585	0.566	0.861	0.820	0.774	0.952	0.903	0.898
		75%	1.164	1.127	1.110	1.362	1.307	1.290	1.477	1.395	1.383
	$\widehat{\sigma}_\epsilon^2$	AVE	0.958	0.975	0.990	0.922	0.968	0.988	0.893	0.925	0.940
		RMS	0.567	0.553	0.547	0.523	0.500	0.498	0.497	0.480	0.455
		25%	0.482	0.483	0.507	0.489	0.576	0.597	0.484	0.563	0.597
		50%	0.936	0.979	1.003	0.931	1.003	1.015	0.946	0.981	0.999
		75%	1.358	1.373	1.395	1.317	1.330	1.371	1.269	1.280	1.285
0.9	$\widehat{\beta}$	AVE	0.820	0.819	0.820	0.860	0.859	0.860	0.880	0.881	0.881
		RMS	0.151	0.152	0.151	0.086	0.086	0.085	0.050	0.050	0.049
		25%	0.761	0.766	0.765	0.821	0.820	0.822	0.859	0.859	0.857
		50%	0.854	0.854	0.853	0.874	0.875	0.876	0.889	0.889	0.888
		75%	0.910	0.909	0.908	0.914	0.914	0.914	0.912	0.914	0.913
	$\widehat{\sigma}_\eta^2$	AVE	1.308	1.300	1.282	1.168	1.168	1.161	1.081	1.075	1.072
		RMS	0.675	0.663	0.644	0.480	0.466	0.469	0.317	0.307	0.300
		25%	0.844	0.862	0.864	0.824	0.842	0.826	0.875	0.867	0.863
		50%	1.274	1.268	1.262	1.118	1.122	1.111	1.057	1.052	1.048
		75%	1.720	1.703	1.673	1.448	1.453	1.426	1.252	1.256	1.245
	$\widehat{\sigma}_\epsilon^2$	AVE	0.680	0.684	0.682	0.826	0.835	0.837	0.925	0.926	0.931
		RMS	0.539	0.531	0.533	0.384	0.371	0.369	0.257	0.247	0.239
		25%	0.303	0.321	0.310	0.614	0.622	0.633	0.765	0.771	0.776
		50%	0.635	0.646	0.660	0.845	0.851	0.848	0.933	0.923	0.929
		75%	0.973	0.961	0.973	1.058	1.056	1.067	1.082	1.079	1.088
1.0	$\widehat{\beta}$	AVE	0.904	0.903	0.903	0.959	0.959	0.959	0.981	0.981	0.981
		RMS	0.133	0.135	0.135	0.056	0.056	0.056	0.026	0.026	0.026
		25%	0.878	0.881	0.879	0.949	0.948	0.948	0.976	0.977	0.976
		50%	0.935	0.935	0.936	0.972	0.972	0.972	0.986	0.986	0.986
		75%	0.962	0.961	0.961	0.983	0.983	0.983	0.991	0.991	0.991
	$\widehat{\sigma}_\eta^2$	AVE	1.595	1.600	1.573	1.309	1.306	1.301	1.150	1.148	1.145
		RMS	0.914	0.917	0.888	0.556	0.544	0.542	0.321	0.321	0.315
		25%	1.095	1.098	1.080	0.979	0.985	0.968	0.958	0.951	0.949
		50%	1.549	1.545	1.536	1.237	1.246	1.230	1.127	1.118	1.123
		75%	2.055	2.050	1.982	1.585	1.559	1.583	1.324	1.311	1.312
	$\widehat{\sigma}_\epsilon^2$	AVE	0.607	0.599	0.600	0.798	0.801	0.801	0.907	0.907	0.909
		RMS	0.568	0.569	0.569	0.360	0.353	0.352	0.227	0.226	0.221
		25%	0.252	0.254	0.229	0.609	0.609	0.615	0.766	0.768	0.773
		50%	0.550	0.555	0.569	0.813	0.805	0.804	0.909	0.902	0.907
		75%	0.870	0.853	0.856	1.002	0.989	0.998	1.041	1.040	1.039

Table 2: Simulation I — Bayesian Estimation

β		T	40			100			200		
		N	2000	5000	10000	2000	5000	10000	2000	5000	10000
0.5	$\widehat{\beta}$	AVE	0.343	0.342	0.342	0.417	0.414	0.411	0.455	0.448	0.440
		RMS	0.248	0.247	0.246	0.152	0.151	0.151	0.114	0.107	0.116
		25%	0.208	0.205	0.204	0.341	0.338	0.337	0.385	0.387	0.375
		50%	0.360	0.365	0.365	0.431	0.423	0.418	0.458	0.446	0.438
		75%	0.484	0.478	0.482	0.507	0.501	0.492	0.530	0.509	0.508
	$\widehat{\sigma}_\eta^2$	AVE	1.363	1.361	1.359	1.233	1.264	1.281	1.158	1.228	1.236
		RMS	0.557	0.531	0.516	0.439	0.432	0.428	0.380	0.395	0.390
		25%	1.073	1.087	1.096	0.964	1.014	1.054	0.902	0.987	1.025
		50%	1.298	1.313	1.323	1.207	1.234	1.258	1.135	1.203	1.214
		75%	1.584	1.598	1.584	1.472	1.479	1.477	1.393	1.434	1.430
	$\widehat{\sigma}_\epsilon^2$	AVE	1.078	1.081	1.082	0.943	0.913	0.898	0.936	0.858	0.861
		RMS	0.425	0.393	0.379	0.345	0.324	0.308	0.328	0.325	0.312
		25%	0.787	0.817	0.817	0.717	0.682	0.688	0.721	0.646	0.680
		50%	1.050	1.051	1.045	0.917	0.907	0.891	0.933	0.867	0.865
		75%	1.334	1.317	1.306	1.152	1.126	1.091	1.152	1.051	1.048
0.9	$\widehat{\beta}$	AVE	0.758	0.758	0.758	0.830	0.829	0.829	0.868	0.865	0.866
		RMS	0.202	0.201	0.201	0.105	0.106	0.106	0.059	0.061	0.062
		25%	0.690	0.689	0.691	0.786	0.783	0.785	0.839	0.837	0.838
		50%	0.789	0.791	0.791	0.844	0.842	0.840	0.877	0.874	0.874
		75%	0.862	0.860	0.860	0.889	0.889	0.887	0.904	0.901	0.903
	$\widehat{\sigma}_\eta^2$	AVE	1.850	1.855	1.853	1.468	1.480	1.485	1.240	1.265	1.249
		RMS	0.989	0.987	0.982	0.620	0.617	0.626	0.393	0.408	0.389
		25%	1.485	1.505	1.503	1.171	1.203	1.209	1.014	1.047	1.047
		50%	1.809	1.823	1.819	1.432	1.457	1.455	1.216	1.226	1.227
		75%	2.158	2.158	2.140	1.716	1.746	1.728	1.447	1.460	1.442
	$\widehat{\sigma}_\epsilon^2$	AVE	0.841	0.840	0.840	0.831	0.823	0.819	0.916	0.885	0.907
		RMS	0.390	0.374	0.369	0.332	0.330	0.329	0.253	0.253	0.247
		25%	0.583	0.589	0.590	0.623	0.624	0.607	0.764	0.739	0.746
		50%	0.796	0.810	0.815	0.836	0.819	0.812	0.916	0.887	0.923
		75%	1.068	1.034	1.028	1.018	1.011	1.003	1.081	1.036	1.062
1.0	$\widehat{\beta}$	AVE	0.877	0.877	0.877	0.952	0.952	0.951	0.979	0.978	0.978
		RMS	0.164	0.164	0.164	0.065	0.065	0.066	0.029	0.030	0.030
		25%	0.842	0.844	0.843	0.939	0.938	0.938	0.972	0.972	0.971
		50%	0.915	0.914	0.914	0.967	0.967	0.967	0.985	0.985	0.985
		75%	0.948	0.949	0.948	0.981	0.981	0.981	0.991	0.991	0.991
	$\widehat{\sigma}_\eta^2$	AVE	2.132	2.136	2.142	1.575	1.583	1.586	1.288	1.305	1.296
		RMS	1.294	1.300	1.303	0.726	0.731	0.729	0.401	0.412	0.407
		25%	1.704	1.711	1.704	1.251	1.262	1.263	1.088	1.109	1.102
		50%	2.055	2.067	2.081	1.516	1.537	1.551	1.270	1.282	1.290
		75%	2.486	2.485	2.505	1.855	1.854	1.850	1.456	1.472	1.469
	$\widehat{\sigma}_\epsilon^2$	AVE	0.776	0.774	0.771	0.818	0.813	0.811	0.937	0.887	0.903
		RMS	0.408	0.401	0.397	0.332	0.327	0.327	0.440	0.222	0.220
		25%	0.535	0.538	0.539	0.613	0.618	0.619	0.781	0.756	0.767
		50%	0.751	0.738	0.726	0.819	0.805	0.804	0.916	0.880	0.903
		75%	0.987	0.971	0.975	1.009	0.994	0.994	1.056	1.014	1.037

Table 3: Simulation II — Maximum Likelihood Estimation

β		T	40			100			200		
	$\backslash N$		2000	5000	10000	2000	5000	10000	2000	5000	10000
0.5	$\widehat{\beta}$	AVE	0.394	0.350	0.326	0.422	0.396	0.369	0.484	0.433	0.423
		RMS	0.271	0.267	0.270	0.277	0.275	0.273	0.272	0.274	0.271
		25%	0.171	0.164	0.164	0.167	0.163	0.154	0.236	0.197	0.186
		50%	0.309	0.258	0.231	0.393	0.337	0.299	0.488	0.405	0.402
		75%	0.613	0.517	0.466	0.652	0.621	0.567	0.719	0.656	0.637
	$\widehat{\sigma}_\epsilon^2$	AVE	1.096	1.088	1.092	1.026	1.027	1.016	1.040	1.013	1.006
		RMS	0.574	0.558	0.542	0.333	0.319	0.313	0.264	0.248	0.243
		25%	0.706	0.721	0.719	0.805	0.814	0.807	0.856	0.840	0.839
		50%	1.047	1.012	1.033	1.018	1.016	1.003	1.017	1.005	0.988
		75%	1.382	1.368	1.360	1.229	1.222	1.202	1.196	1.173	1.166
0.9	$\widehat{\beta}$	AVE	0.453	0.412	0.396	0.470	0.416	0.400	0.577	0.493	0.477
		RMS	0.528	0.558	0.568	0.530	0.572	0.581	0.449	0.523	0.534
		25%	0.172	0.160	0.162	0.151	0.126	0.123	0.261	0.147	0.138
		50%	0.414	0.326	0.300	0.447	0.325	0.301	0.693	0.534	0.491
		75%	0.742	0.684	0.649	0.782	0.732	0.708	0.845	0.811	0.794
	$\widehat{\sigma}_\epsilon^2$	AVE	0.635	0.627	0.627	0.597	0.560	0.556	0.663	0.603	0.591
		RMS	0.559	0.561	0.557	0.494	0.521	0.521	0.432	0.478	0.479
		25%	0.318	0.320	0.330	0.391	0.361	0.370	0.459	0.409	0.406
		50%	0.574	0.582	0.582	0.581	0.551	0.545	0.652	0.587	0.590
		75%	0.878	0.843	0.849	0.790	0.739	0.729	0.854	0.793	0.780

Table 4: Simulation II — Bayesian Estimation

β		T	40			100			200		
	$\backslash N$		2000	5000	10000	2000	5000	10000	2000	5000	10000
0.5	$\widehat{\beta}$	AVE	0.463	0.464	0.463	0.464	0.466	0.463	0.481	0.475	0.474
		RMS	0.077	0.066	0.064	0.114	0.101	0.098	0.152	0.136	0.124
		25%	0.417	0.428	0.433	0.389	0.394	0.400	0.368	0.370	0.384
		50%	0.459	0.456	0.453	0.450	0.454	0.447	0.475	0.467	0.471
		75%	0.499	0.491	0.483	0.529	0.521	0.511	0.576	0.568	0.557
	$\widehat{\sigma}_\epsilon^2$	AVE	1.311	1.314	1.314	1.132	1.130	1.129	1.088	1.088	1.090
		RMS	0.649	0.646	0.650	0.347	0.334	0.333	0.266	0.255	0.247
		25%	0.927	0.929	0.922	0.926	0.928	0.914	0.910	0.915	0.933
		50%	1.209	1.236	1.221	1.116	1.111	1.116	1.072	1.072	1.089
		75%	1.586	1.578	1.587	1.337	1.322	1.319	1.245	1.254	1.234
0.9	$\widehat{\beta}$	AVE	0.484	0.482	0.481	0.496	0.484	0.483	0.568	0.522	0.506
		RMS	0.427	0.426	0.426	0.434	0.439	0.438	0.388	0.423	0.434
		25%	0.419	0.426	0.431	0.375	0.382	0.390	0.416	0.369	0.360
		50%	0.470	0.464	0.463	0.477	0.464	0.458	0.567	0.519	0.505
		75%	0.534	0.514	0.511	0.617	0.576	0.565	0.719	0.675	0.646
	$\widehat{\sigma}_\epsilon^2$	AVE	0.840	0.837	0.836	0.695	0.678	0.678	0.717	0.678	0.667
		RMS	0.475	0.459	0.469	0.398	0.401	0.398	0.361	0.391	0.396
		25%	0.541	0.554	0.542	0.526	0.508	0.513	0.552	0.530	0.513
		50%	0.756	0.753	0.749	0.688	0.660	0.664	0.706	0.677	0.661
		75%	1.025	1.018	1.016	0.847	0.825	0.816	0.871	0.829	0.812

Table 5: Simulation III — Maximum Likelihood Estimation

β		T	40			100			200		
		N	2000	5000	10000	2000	5000	10000	2000	5000	10000
0.5	$\widehat{\beta}$	AVE	0.336	0.327	0.330	0.383	0.382	0.382	0.423	0.425	0.424
		RMS	0.264	0.270	0.271	0.234	0.234	0.232	0.186	0.190	0.181
		25%	0.153	0.137	0.131	0.207	0.217	0.219	0.308	0.308	0.319
		50%	0.296	0.302	0.297	0.387	0.391	0.396	0.439	0.443	0.439
		75%	0.503	0.482	0.504	0.545	0.540	0.545	0.548	0.557	0.549
	$\widehat{\sigma}_\eta^2$	AVE	1.174	1.145	1.134	1.064	1.071	1.068	1.068	1.050	1.046
		RMS	0.728	0.716	0.711	0.466	0.459	0.454	0.349	0.339	0.337
		25%	0.646	0.626	0.604	0.725	0.729	0.737	0.829	0.798	0.814
		50%	1.030	0.988	0.999	0.998	1.016	1.022	1.053	1.030	1.012
		75%	1.542	1.501	1.487	1.345	1.355	1.351	1.297	1.269	1.259
0.9	$\widehat{\beta}$	AVE	0.748	0.749	0.751	0.844	0.844	0.845	0.872	0.876	0.875
		RMS	0.257	0.255	0.255	0.110	0.109	0.107	0.065	0.060	0.056
		25%	0.663	0.667	0.676	0.803	0.800	0.803	0.845	0.847	0.849
		50%	0.810	0.811	0.815	0.868	0.867	0.865	0.883	0.885	0.884
		75%	0.900	0.896	0.896	0.909	0.908	0.910	0.912	0.915	0.910
	$\widehat{\sigma}_\eta^2$	AVE	1.695	1.671	1.657	1.307	1.298	1.295	1.154	1.136	1.138
		RMS	1.089	1.062	1.044	0.562	0.547	0.542	0.363	0.336	0.345
		25%	1.102	1.098	1.098	0.975	0.967	0.971	0.923	0.921	0.906
		50%	1.516	1.470	1.467	1.239	1.229	1.232	1.114	1.112	1.106
		75%	2.092	2.035	2.048	1.547	1.552	1.561	1.347	1.303	1.318

Table 6: Simulation III — Bayesian Estimation

β		T	40			100			200		
		N	2000	5000	10000	2000	5000	10000	2000	5000	10000
0.5	$\widehat{\beta}$	AVE	0.387	0.387	0.387	0.392	0.393	0.393	0.406	0.412	0.408
		RMS	0.156	0.155	0.154	0.167	0.164	0.164	0.156	0.150	0.154
		25%	0.309	0.311	0.312	0.298	0.302	0.302	0.314	0.322	0.318
		50%	0.370	0.372	0.372	0.377	0.379	0.380	0.401	0.411	0.408
		75%	0.455	0.449	0.451	0.475	0.475	0.470	0.495	0.501	0.493
	$\widehat{\sigma}_\eta^2$	AVE	1.782	1.779	1.778	1.308	1.309	1.308	1.182	1.174	1.169
		RMS	1.158	1.143	1.139	0.568	0.567	0.563	0.383	0.374	0.381
		25%	1.173	1.147	1.160	0.963	0.961	0.954	0.932	0.936	0.934
		50%	1.625	1.625	1.631	1.241	1.254	1.258	1.145	1.141	1.116
		75%	2.247	2.240	2.222	1.617	1.606	1.611	1.416	1.376	1.394
0.9	$\widehat{\beta}$	AVE	0.697	0.698	0.698	0.819	0.818	0.819	0.865	0.865	0.866
		RMS	0.265	0.264	0.264	0.129	0.129	0.129	0.063	0.064	0.065
		25%	0.583	0.583	0.583	0.772	0.773	0.772	0.835	0.836	0.839
		50%	0.735	0.732	0.736	0.843	0.844	0.844	0.874	0.875	0.876
		75%	0.838	0.839	0.838	0.890	0.890	0.890	0.903	0.903	0.905
	$\widehat{\sigma}_\eta^2$	AVE	2.460	2.461	2.460	1.576	1.577	1.577	1.270	1.271	1.260
		RMS	1.781	1.780	1.773	0.762	0.762	0.761	0.424	0.418	0.425
		25%	1.747	1.740	1.770	1.206	1.210	1.206	1.044	1.049	1.002
		50%	2.261	2.271	2.261	1.508	1.518	1.516	1.226	1.251	1.232
		75%	2.949	2.948	2.963	1.879	1.884	1.872	1.472	1.451	1.467

Table 7: Simulation IV — Maximum Likelihood Estimation

	T	40			100			200		
N		2000	5000	10000	2000	5000	10000	2000	5000	10000
$\widehat{\sigma}_\eta^2$	AVE	1.785	2.032	2.119	1.620	1.753	1.888	1.327	1.571	1.631
	RMS	1.411	1.647	1.764	1.173	1.320	1.440	0.942	1.130	1.195
	25%	0.786	0.900	0.931	0.692	0.727	0.821	0.526	0.635	0.654
	50%	1.577	1.777	1.839	1.566	1.646	1.819	1.179	1.473	1.529
	75%	2.526	2.882	3.044	2.369	2.589	2.764	2.039	2.352	2.464
$\widehat{\sigma}_\epsilon^2$	AVE	0.988	0.981	0.977	0.993	0.989	0.987	0.995	0.994	0.985
	RMS	0.232	0.230	0.230	0.151	0.152	0.151	0.108	0.104	0.106
	25%	0.823	0.817	0.816	0.885	0.879	0.879	0.919	0.926	0.910
	50%	0.975	0.970	0.966	0.985	0.984	0.983	0.995	0.990	0.981
	75%	1.132	1.123	1.122	1.091	1.083	1.081	1.063	1.058	1.056

Table 8: Simulation IV — Bayesian Estimation

	T	40			100			200		
N		2000	5000	10000	2000	5000	10000	2000	5000	10000
$\widehat{\sigma}_\eta^2$	AVE	2.403	2.750	2.912	1.856	2.047	2.163	1.593	1.720	1.850
	RMS	1.645	1.989	2.120	1.094	1.248	1.370	0.858	0.944	1.052
	25%	1.805	2.083	2.249	1.371	1.555	1.663	1.149	1.267	1.427
	50%	2.185	2.483	2.674	1.723	1.879	1.955	1.488	1.602	1.720
	75%	2.763	3.148	3.318	2.159	2.383	2.445	1.941	2.033	2.141
$\widehat{\sigma}_\epsilon^2$	AVE	1.095	1.090	1.088	1.034	1.031	1.029	1.006	1.011	1.004
	RMS	0.267	0.266	0.265	0.159	0.159	0.158	0.110	0.107	0.105
	25%	0.915	0.908	0.907	0.924	0.919	0.916	0.926	0.937	0.930
	50%	1.091	1.080	1.073	1.028	1.025	1.025	1.002	1.009	1.005
	75%	1.255	1.250	1.250	1.128	1.127	1.123	1.075	1.078	1.079

Table 9: Computational Time (Seconds)

Simu-	T	40			100			200		
lation	$\beta \setminus N$	2000	5000	10000	2000	5000	10000	2000	5000	10000
(a) Maximum Likelihood Estimation										
I	0.5	0.94	2.36	4.72	1.65	4.12	8.24	2.82	7.06	14.12
II	0.5	1.23	3.07	6.14	2.63	6.58	13.15	4.97	12.42	24.84
III	0.5	1.26	3.15	6.30	2.68	6.70	13.40	5.04	12.60	25.19
IV		1.51	3.77	7.54	3.29	8.23	16.47	6.27	15.67	31.35
(b) Bayesian Estimation										
I	0.5	0.90	2.31	4.52	1.60	4.01	8.00	2.76	6.90	13.82
II	0.5	0.83	2.08	4.16	1.64	4.10	8.21	2.99	7.48	14.95
III	0.5	0.85	2.13	4.26	1.69	4.22	8.45	3.09	7.72	15.43
IV		0.96	2.40	4.81	1.96	4.91	9.82	3.63	9.08	18.17

Thus, except for Simulation II, the maximum likelihood estimation procedure suggested in this paper performs better than the Bayesian estimator with the diffuse priors from the criteria of the bias and root mean square errors of the estimates. That is, the estimates from the maximum likelihood procedure are closer to the true parameter values and they have the smaller root mean square errors, compared with the Bayesian estimator.

The values in Table 9 indicate the CPU run times, where Pentium II 500MHz CPU and Watcom FORTRAN 77/32 Compiler (Version 11.0) are used for comparison between the maximum likelihood estimation and the Bayesian estimation. Each value represents the average time from the 1000 simulation runs shown in Tables 1 – 9. As discussed in Section 3.2, since the proposed maximum likelihood estimator is the two-step estimator, the maximum likelihood estimation takes more time than the Bayesian estimation, depending on the underlying state-space models. For Simulations II – IV, the Bayesian estimation is 1.6 – 1.7 times faster than the maximum likelihood estimation. However, since even the case of Simulation IV, $T = 200$ and $N = 10000$ takes 31.35 seconds, it is concluded from computational point of view that the proposed estimator does not have too much disadvantage, compared with the Bayesian estimator.

5 SUMMARY

We have shown the maximum likelihood estimation procedure when the unknown parameters are included in the nonlinear and non-Gaussian state-space models. The traditional maximization procedures such as the simple grid search and the Newton-Raphson optimization are conventionally used for maximization of the likelihood function. In the case where the number of the parameters increases, however, these procedures are not helpful. Furthermore, since the simulation-based filtering and smoothing procedures yield the sampling errors or the simulation errors, precise estimates of the parameters cannot be obtained. In this paper, therefore, we have suggested applying the Monte Carlo stochastic optimization method to maximum likelihood estimation, where the features of the suggested procedure are: (i) the likelihood function is regarded as the density function with respect to the unknown parameters, (ii) the random draws of the unknown parameters are generated by the Gibbs sampler and the Metropolis-Hastings algorithm, and (iii) based on the random draws, we construct the empirical densities of the parameters and obtain the modes which correspond to the maximum likelihood estimates. If the arithmetic averages based on the random draws is taken as the estimates, the obtained estimates reduce to the Bayesian estimates with the diffuse priors. Thus, the maximum likelihood estimation is based on the mode while the Bayesian estimation is the arithmetic average. In the Monte Carlo simulation

studies, the maximum likelihood estimation is compared with the Bayesian estimation, taking some kinds of state-space models. As a result, the maximum likelihood estimation performs better than the Bayesian estimation for almost all the cases by the bias and root mean square error criteria. Moreover, the CPU run times are compared between the suggested estimator and the Bayesian estimator. We have the result that the suggested procedure does not have too much computational disadvantage.

APPENDICES

Appendix 1: Gibbs Sampler

When the two conditional densities $p(x|y)$ and $p(y|x)$ are available, we consider generating random draws of (x, y) from the joint density $p(x, y)$. The Gibbs sampler is performed as follows: (i) pick up an appropriate value for the initial value of y (i.e., y_0), (ii) generate a random draw of x (i.e., x_i) from $p(x|y_{i-1})$, (iii) generate a random draw of y (i.e., y_i) from $p(y|x_i)$, and (iv) repeat (ii) and (iii) for $i = 1, 2, \cdots, N$.

From the basic result of the Gibbs sampler, (x_N, y_N) converges to a random draw from $p(x, y)$ for sufficiently large N. We have considered the bivariate case, but it is easily extended to the multivariate cases. See Geman and Geman (1984), Tanner and Wong (1987), Gelfand and Smith (1990), Arnold (1993) and Geweke (1996, 1997) for the Gibbs sampler.

Appendix 2: Metropolis-Hastings Algorithm

Consider generating a random draw of z from $p(z)$. The Metropolis-Hastings algorithm utilizes the proposal density $p_*(z|x)$, which has to be appropriately specified by a researcher. Define the acceptance probability $\omega(x, z)$ as:

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

The Metropolis-Hastings algorithm is the random number generation method such that we can generate random draws from any density function, which can be implemented as follows: (i) take an initial value of x as 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) set $x_i = z$ if $u \leq \omega(x_{i-1}, z)$ and set $x_i = x_{i-1}$ otherwise, and (iv) repeat (ii) and (iii) for $i = 1, 2, \cdots, N$.

We can take x_N as a random draw from $p(x)$ for sufficiently large N. For choice of the proposal density $p_*(z|x)$, the proposal density should not have too large variance and too small variance, compared with the target density (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 $\omega(x, z)$. Smith and Roberts (1993), Tierney (1994), Chib and Greenberg (1995, 1996) and Geweke (1996) discussed the Metropolis-Hastings algorithm.

Appendix 3: Nonparametric Density Estimation

The nonparametric density estimation is briefly described as follows. Let x_i, $i = 1, 2, \cdots, N$, be the random observations associated with the density $p(x)$. When $p(x)$ is not known but x_i, $i = 1, 2, \cdots, N$, are available, $p(x)$ is estimated as follows.

$$\widehat{p}(x) = \frac{1}{Nh} \sum_{i=1}^{N} K\left(\frac{x - x_i}{h}\right),$$

where h is called the window width or the smoothing parameter. $K(\cdot)$ denotes the kernel function which satisfies: $\int K(x)\mathrm{d}x = 1$ and in many cases $\int xK(x)\mathrm{d}x = 0$ and $\int x^2K(x)\mathrm{d}x = 1$. Representatively, the standard normal density is chosen for $K(\cdot)$. Under some conditions, the estimate of h is given by: $\widehat{h} = 0.9AN^{-1/5}$, where $A = \min(\sigma, R/1.34)$. Note that σ and R represent the standard deviation and the interquartile range from the sample $\{x_1, x_2, \cdots, x_N\}$. See Silverman (1986, pp.45 – 48) for choice of the smoothing parameter.

As for asymptotic properties, we need to assume as follows:

$$\int |K(x)|\mathrm{d}x < \infty,$$
$$\int K(x)\mathrm{d}x = 1,$$
$$|xK(x)| \longrightarrow 0 \quad \text{as } |x| \longrightarrow \infty,$$
$$\widehat{h} \longrightarrow 0 \quad \text{and} \quad N\widehat{h} \longrightarrow \infty \quad \text{as } N \longrightarrow \infty.$$

All the assumptions with respect to $K(\cdot)$ and $\widehat{h}$ are satisfied when we take a standard normal density for $K(\cdot)$ and $\widehat{h} = 0.9AN^{-1/5}$ for $\widehat{h}$. Under the above assumptions, we have the following results:

$$\widehat{p}(x) \longrightarrow p(x) \text{ in probability as } N \longrightarrow \infty.$$

See Silverman (1986, pp.70 – 72) for consistency. In addition, we have the following two convergence rates:

$$\sup_x \left| \widehat{p}(x) - \mathrm{E}(\widehat{p}(x)) \right| = O(\widehat{h}^{-1} N^{-1/2} (\log \log N)^{1/2})$$
$$\sup_x \left| \mathrm{E}(\widehat{p}(x)) - p(x) \right| = O(\widehat{h}^{-1}).$$

See Prakasa Rao (1983, p.50) for the convergence rates. We can rewrite as follows:

$$\begin{aligned} |\widehat{p}(x) - p(x)| &= |\widehat{p}(x) - \mathrm{E}(\widehat{p}(x)) + \mathrm{E}(\widehat{p}(x)) - p(x)| \\ &\leq |\widehat{p}(x) - \mathrm{E}(\widehat{p}(x))| + |\mathrm{E}(\widehat{p}(x)) - p(x)| \end{aligned}$$

Since the supremum of the second term in the right hand side is the order of $O(\widehat{h}^{-1})$, the convergence rate from $\widehat{p}(x)$ to $p(x)$ is at most $O(\widehat{h}^{-1})$, i.e., $O(N^{1/5})$, which indicates that convergence speed is very slow.

REFERENCES

Arnold, S.F., 1993, " Gibbs Sampling, " in *Handbook of Statistics*, Vol.9, edited by Rao, C.R., pp.599 – 625, North-Holland.

Bohachevsky, I.O., Johnson, M.E. and Stein, M.L., 1986, " Generalized Simulated Annealing for Function Optimization, " *Technometrics*, Vol.28, No.3, pp.209 – 217.

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.

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.

Brooks, S.P. and Morgan, B.J.T., 1995, " Optimization Using Simulated Annealing, " *The Statistician*, Vol.44, No.2, pp.241 – 257.

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.

Chan, K.S. and Ledolter, J., 1995, " Monte Carlo EM Estimation for Time Series Models Involving Counts, " *Journal of the American Statistical Association*, Vol.90, No.429, pp.242 – 252.

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.

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., 1991, " Efficient Simulation from the Multivariate Normal and Student-t Distributions Subject to Linear Constraints, " in *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, edited by E.M., Keramidas, pp.571 – 578.

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.

Goffe, W.L., Ferrier, G.D. and Rogers, J., 1994, " Global Optimization of Statistical Functions with Simulated Annealing, " *Journal of Econometrics*, Vol.60, No.1&2, pp.65 – 99.

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.

Härdle, W., 1990, *Applied Nonparametric Regression*, Cambridge University Press.

Izenman, A.J., 1991, " Recent Developments in Nonparametric Density Estimation, " *Journal of the American Statistical Association*, Vol.86, No.413, pp.205-224.

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., 1998, " A Self-Organizing State-Space Model, " *Journal of the American Statistical Association*, Vol.93, No.443, pp.1203 – 1215.

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

Kirkpatrick, S., Gelatt, C.D., Jr. and Vecchi, M.P., 1983, " Optimization by Simulated Annealing, " *Science*, Vol.220, No.4598, pp.671 – 680.

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

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

Prakasa Rao, B.L.S., 1983, *Nonparametric Functional Estimation*, Academic Press.

Silverman, B.W., 1986, *Density Estimation for Statistics and Data Analysis* (Monographs on Statistics and Applied Probability 26), Chapman & Hall.

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

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

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

Tanizaki, H., 2000, " Nonlinear and Non-Gaussian State-Space Modeling with Monte Carlo Techniques: A Survey and Comparative Study, " in *Handbook of Statistics (Stochastic Processes: Modeling and Simulation)*, edited by Rao, C.R. and Shanbhag, D.N., North-Holland, forthcoming.

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., John Wiley & Sons, pp.245 – 261, 1995).

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., 1993, *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, Springer-Verlag.

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.

Ullah, A., 1988, " Non-Parametric Estimation of Econometric Functionals, " *Canadian Journal of Economics*, Vol.21, No.3, pp.625 – 658.

Wei, G.C.G. and Tanner, M.A., 1990, " A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms, " *Journal of the American Statistical Association*, Vol.85, No.411, pp.699 – 704.