Nonlinear and nonnormal filter using importance sampling: antithetic monte carlo integration

Tanizaki, Hisashi

(Citation)
Communications in Statistics: Simulation and Computation,28(2):463-486

(Issue Date)
1999

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90000565

NONLINEAR AND NONNORMAL FILTER USING IMPORTANCE SAMPLING: ANTITHETIC MONTE CARLO INTEGRATION

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

Key Words: *State-Space Model, Filtering, Monte Carlo Integration, Importance Sampling, Antithetic Monte Carlo, Resampling.*

ABSTRACT: In this paper, the importance sampling filter proposed by Mariano and Tanizaki (1995), Tanizaki (1996), Tanizaki and Mariano (1994) is extended using the antithetic Monte Carlo method to reduce the simulation errors. By Monte Carlo studies, the importance sampling filter with the antithetic Monte Carlo method is compared with the importance sampling filter without the antithetic Monte Carlo method. It is shown that for all the simulation studies the former is clearly superior to the latter especially when number of random draws is small.

1 INTRODUCTION

The traditional nonlinear filters such as the extended Kalman filter, the second-order nonlinear filter and the Gaussian sum filter are used from various aspects. Especially, the extended Kalman filter is obtained as follows: (i) the nonlinear measurement and transition equations are linearized by the first-order Taylor series expansion and (ii) the linearized functions are applied directly to the standard Kalman filter formula. See Gelb (1974) and Anderson and Moore (1979) for the traditional nonlinear filters.

Recently, the nonlinear filters have been developed since Kitagawa (1987) and Harvey (1989), where the nonlinear filter is derived based on the underlying density functions. The traditional nonlinear filters are obtained by approximating the nonlinear measurement and transition equations by the Taylor series expansions. Kitagawa (1987) and Kramer and Sorenson (1988) proposed the nonlinear filter approximating the density functions by numerical integration, which has the disadvantage that it is not feasible in the higher dimensional cases of the state-vector. In order to improve the disadvantage, various density-based nonlinear filters have been developed.

Tanizaki (1996), Tanizaki and Mariano (1994) and Mariano and Tanizaki (1995) derived the nonlinear filtering algorithm using Monte Carlo integration with importance sampling, which utilizes the random numbers generated from the appropriately chosen importance density to obtain the filtering mean and variance. See, for example, Geweke (1988, 1989a, 1989b, 1996, 1997), Shao (1989) and Koop (1994) for Monte Carlo integration with importance sampling. Moreover, Carlin, Polson and Stoffer (1992), Carter and Kohn (1994, 1996) and Chib and Greenberg (1996) proposed the nonlinear smoothing by Gibbs sampling. The nonlinear filter with rejection sampling is introduced by Mariano and Tanizaki (1999), Tanizaki (1996, 1999) and Tanizaki and Mariano (1998), where random draws are directly generated from the filtering densities. Gordon, Salmond and Smith (1993) and Kitagawa (1996) applied the resampling procedure to the nonlinear filter and smoother.

Thus, in the last decade, a large amount of research on filtering theory has been devoted to evaluating the filtering mean and variance by generating random draws. In this paper, we re-consider the importance sampling filter developed by Mariano and Tanizaki (1995), Tanizaki (1996) and Tanizaki and Mariano (1994). In the case where we evaluate a nonlinear function by generating random draws by a computer, we have simulation errors. The simulation errors disappear as number of the random draws increases, which is obtained from the central limit theorem. However, in practice, a small sample is usually used for the random draws from computational point of view. Therefore, we need to consider variance reduction of the Monte Carlo method. The antithetic Monte Carlo method is utilized for variance reduction (see, for example, Geweke (1988, 1989a, 1989b, 1996, 1997) for the antithetic Monte Carlo method). Thus, in this paper, an attempt is made to reduce the simulation errors of the importance sampling filter by the antithetic Monte Carlo method. The importance sampling filter with the antithetic Monte Carlo method is compared with the resampling filter, which is one of the most recently developed nonlinear filters (see Gordon, Salmond and Smith (1993) and Kitagawa (1996) for the resampling filter). The resampling filter is briefly described in Appendix A.

As a result, it is concluded from several Monte Carlo studies that (i) the

importance sampling filter with the antithetic Monte Carlo method works well especially when number of random draws is small and (ii) under the same number of random draws, the importance sampling filters perform better than the resampling filter.

2 PRELIMINARIES

2.1 OVERVIEW OF NONLINEAR AND NONNORMAL FILTER

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

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

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

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. $h_t(\cdot, \cdot)$ and $f_t(\cdot, \cdot)$ are vector functions, which are assumed to be known. 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., $Y_s = \{y_s, y_{s-1}, \cdots, y_1\}$. Define $P_y(y_t|\alpha_t)$ and $P_\alpha(\alpha_t|\alpha_{t-1})$ 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.

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

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

$$P(\alpha_t|Y_t) = \frac{P_y(y_t|\alpha_t)P(\alpha_t|Y_{t-1})}{\displaystyle\int P_y(y_t|\alpha_t)P(\alpha_t|Y_{t-1})\mathrm{d}\alpha_t}, \tag{4}$$

where the initial condition is given by:

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

The two densities $P_\alpha(\alpha_t|\alpha_{t-1})$ and $P_y(y_t|\alpha_t)$ are computed from the measurement equation (1) and the transition equation (2), respectively. Equation (3) corresponds to one-step ahead prediction, which plays a role of predicting α_t

using the past information Y_{t-1}. Equation (4) combines the present sample with the past information. Accordingly, equation (3) is called the prediction equation while equation (4) is known as the update equation. Based on the two densities, the density-based filtering algorithm is represented as the following two steps: (i) equation (3) yields $P(\alpha_t|Y_{t-1})$ given $P(\alpha_{t-1}|Y_{t-1})$ and (ii) equation (4) yields $P(\alpha_t|Y_t)$ given $P(\alpha_t|Y_{t-1})$. Thus, repeating predicting and updating for all t, the filtering densities $P(\alpha_t|Y_t)$, $t = 1, 2, \cdots, T$, can be obtained.

Filtering Mean and Variance: Our goal is to estimate the state-variable α_t. Once we have the recursions shown in equations (3) and (4), the filtering mean (i.e., $a_{t|s} \equiv \mathrm{E}(\alpha_t|Y_s)$) and the filtering variance (i.e., $\Sigma_{t|s} \equiv \mathrm{Var}(\alpha_t|Y_s)$) are easily obtained as follows:

$$a_{t|s} = \int \alpha_t P(\alpha_t|Y_s)\mathrm{d}\alpha_t, \tag{5}$$

$$\Sigma_{t|s} = \int (\alpha_t - a_{t|s})(\alpha_t - a_{t|s})' P(\alpha_t|Y_s)\mathrm{d}\alpha_t, \tag{6}$$

for $s = t-1, t$.

2.2 MONTE CARLO INTEGRATION WITH IMPORTANCE SAMPLING

Let x be a random variable associated with a density function $P(x)$. Suppose that $g(x)$ is a function of x. The expectation of $g(x)$ is represented as follows:

$$\mathrm{E}(g(x)) = \int g(x)P(x)\mathrm{d}x \equiv \mu. \tag{7}$$

Suppose that the integration in equation (7) cannot be carried out analytically in the present case. In such a case, the expectation is evaluated by Monte Carlo integration with importance sampling. We utilize another density function of x, say $P_I(x)$, which is called the importance density and is appropriately assumed by a researcher. Let x_i, $i = 1, 2, \cdots, n$, be the random numbers from the density function $P_I(x)$. Define the weight function $\omega(x)$ as a ratio of the two densities, i.e.,

$$\omega(x) = \frac{P(x)}{P_I(x)}.$$

Then, we have the following approximation of equation (7):

$$\mu \equiv \mathrm{E}(g(x)) = \int g(x)\omega(x)P_I(x)\mathrm{d}x \approx \frac{1}{n}\sum_{i=1}^{n} g(x_i)\omega(x_i) \equiv \overline{g}_n. \tag{8}$$

The expectation $\mathrm{E}(g(x))$ is evaluated as $\overline{g}_n$. The properties of $\overline{g}_n$ are known as follows (see, for example, Geweke (1988, 1989a, 1989b, 1996, 1997), Shao (1989) and Koop (1994)):

(i) $\overline{g}_n$ converges almost surely to μ, if the support of $P(x)$ is included in the support of $P_I(x)$.

(ii) If $\omega(x)$ is bounded above,

$$\sqrt{n}(\overline{g}_n - \mu) \longrightarrow N(0, \Sigma),$$

where $\Sigma \equiv \mathrm{E}[(g(x)\omega(x) - \mu)(g(x)\omega(x) - \mu)']$ and the arrow denotes convergence in distribution.

It is shown from the result (ii) that $\overline{g}_n$ in equation (8) is consistent but convergence speed is quite slow as $\sqrt{n}$.

Moreover, Geweke (1988, 1989a, 1989b, 1996, 1997) pointed out the following. It is important that the importance density $P_I(x)$ is not too different from $P(x)$ and especially important that the weight function does not become very large over the support of $P(x)$. This indicates that numerical accuracy is adversely affected by large relative values of the weight function, i.e., such large values indicate a poor approximation of $\mathrm{E}(g(x))$.

Using the Monte Carlo integration method with importance sampling, Mariano and Tanizaki (1995), Tanizaki (1996), Tanizaki and Mariano (1994) evaluated the integrations included in the nonlinear filtering algorithm represented by equations (3) and (4), which are described in the next section.

2.3 NONLINEAR FILTER USING IMPORTANCE SAMPLING

Mariano and Tanizaki (1995), Tanizaki (1996) and Tanizaki and Mariano (1994) proposed the nonlinear filtering algorithm with importance sampling, where density functions are converted by weight functions and a recursive algorithm of the weight functions is obtained. Define the weight function as:

$$\omega(\alpha_t|Y_s) = \frac{P(\alpha_t|Y_s)}{P_I(\alpha_t)},$$

for $s = t-1, t$, where the density function $P_I(\alpha_t)$ has to be appropriately specified by a researcher, which is called the importance density.

Using the weight functions, $\omega(\alpha_t|Y_s)$ for $s = t-1, t$, the density-based filtering algorithm given by equations (3) and (4) are rewritten as follows:

$$\begin{aligned}\omega(\alpha_t|Y_{t-1}) &= \int \frac{P_\alpha(\alpha_t|\alpha_{t-1})}{P_I(\alpha_t)}\omega(\alpha_{t-1}|Y_{t-1})P_I(\alpha_{t-1})\mathrm{d}\alpha_{t-1} \\ &\approx \frac{1}{n}\sum_{j=1}^{n}\frac{P_\alpha(\alpha_t|\alpha_{j,t-1})}{P_I(\alpha_t)}\omega(\alpha_{j,t-1}|Y_{t-1}), \qquad (9)\end{aligned}$$

$$\omega(\alpha_t|Y_t) = \frac{P_y(y_t|\alpha_t)\omega(\alpha_t|Y_{t-1})}{\int P_y(y_t|\alpha_t)\omega(\alpha_t|Y_{t-1})P_I(\alpha_t)\mathrm{d}\alpha_t}$$
$$\approx \frac{P_y(y_t|\alpha_t)\omega(\alpha_t|Y_{t-1})}{\frac{1}{n}\sum_{j=1}^{n} P_y(y_t|\alpha_{j,t})\omega(\alpha_{j,t}|Y_{t-1})}, \tag{10}$$

where $\alpha_{i,t}$, $i = 1, 2, \cdots, n$, denote the random draws generated from the importance density $P_I(\alpha_t)$. Note that we need to use the same importance density to evaluate the two integrals in equations (9) and (10). For $P(\alpha_t|Y_t)$ and $P(\alpha_t|Y_{t-1})$, it is not possible to obtain the explicit functional form in general. Accordingly, $\omega(\alpha_t|Y_t)$ and $\omega(\alpha_t|Y_{t-1})$ cannot be explicitly obtained. We need to evaluate the two weight functions at the same random draws in order to construct the recursive algorithm on the weight functions.

Again, in equations (9) and (10), α_t is evaluated at $\alpha_{i,t}$. Thus, a recursive filtering algorithm of the weight functions is derived as follows:

$$\omega(\alpha_{i,t}|Y_{t-1}) \approx \frac{1}{n}\sum_{j=1}^{n} \frac{P_\alpha(\alpha_{i,t}|\alpha_{j,t-1})}{P_I(\alpha_{i,t})}\omega(\alpha_{j,t-1}|Y_{t-1}), \tag{11}$$

$$\omega(\alpha_{i,t}|Y_t) \approx \frac{P_y(y_t|\alpha_{i,t})\omega(\alpha_{i,t}|Y_{t-1})}{\frac{1}{n}\sum_{j=1}^{n} P_y(y_t|\alpha_{j,t})\omega(\alpha_{j,t}|Y_{t-1})}, \tag{12}$$

for $i = 1, 2, \cdots, n$ and $t = 1, 2, \cdots, T$, where each weight function is evaluated at the random draws, $\alpha_{i,t}$ for $i = 1, 2, \cdots, n$. The initial condition of the weight function is given by:

$$\omega(\alpha_{i,1}|Y_0) \approx \frac{1}{n}\sum_{j=1}^{n} \frac{P_\alpha(\alpha_{i,1}|\alpha_{j,0})}{P_I(\alpha_{i,1})}\omega(\alpha_{j,0}|Y_0), \qquad \text{if } \alpha_0 \text{ is stochastic,}$$
$$\omega(\alpha_{i,1}|Y_0) = \frac{P_\alpha(\alpha_{i,1}|\alpha_0)}{P_I(\alpha_{i,1})}, \qquad \text{if } \alpha_0 \text{ is nonstochastic.}$$

The recursive algorithm shown in equations (11) and (12) is implemented by the following two steps: (i) given $\omega(\alpha_{j,t-1}|Y_{t-1})$ for $j = 1, 2, \cdots, n$, $\omega(\alpha_{i,t}|Y_{t-1})$ for $i = 1, 2, \cdots, n$ are obtained from equation (11), and (ii) given $\omega(\alpha_{j,t}|Y_{t-1})$ for $j = 1, 2, \cdots, n$, $\omega(\alpha_{i,t}|Y_t)$ for $i = 1, 2, \cdots, n$ are obtained from equation (12). Thus, the weight functions are recursively evaluated for all time $t = 1, 2, \cdots, T$.

Filtering Mean and Variance: Equations (5) and (6) are approximated as:

$$a_{t|s} \approx \frac{1}{n}\sum_{i=1}^{n} \alpha_{i,t}\omega(\alpha_{i,t}|Y_s), \tag{13}$$

$$\Sigma_{t|s} \approx \frac{1}{n}\sum_{i=1}^{n} (\alpha_{i,t} - a_{t|s})(\alpha_{i,t} - a_{t|s})'\omega(\alpha_{i,t}|Y_s), \tag{14}$$

for $s = t-1, t$. Thus, equations (13) and (14) represent the filtering estimates based on Monte Carlo integration with importance sampling.

Filtering Density: Sometimes, we need shape of the filtering densities. From definition of the weight function, we can evaluate the filtering density $P(\alpha_t|Y_s)$ at $\alpha_t = \alpha_{i,t}$, which is as follows:

$$P(\alpha_{i,t}|Y_s) = \omega(\alpha_{i,t}|Y_s)P_I(\alpha_{i,t}),$$

for $s = t-1, t$.

3 VARIANCE REDUCTION METHOD IN MONTE CARLO INTEGRATION

The problem of the importance sampling filter is choice of the importance density $P_I(\alpha_t)$, because precision of the filtering estimates is sensitive to the importance density.

Choice of Importance Density: In order to compute the weight functions $\omega(\alpha_t|Y_{t-1})$ and $\omega(\alpha_t|Y_t)$, it is important that the importance density $P_I(\alpha_t)$ is not too different from both $P(\alpha_t|Y_{t-1})$ and $P(\alpha_t|Y_t)$.

In the case where the state-variable lies on a certain interval, for example, in the case of $0 < \alpha_t < 1$, we may choose the importance density $P_I(\alpha_t)$ as the following uniform distribution between zero and one:

$$\alpha_{i,t} = u_{i,t}, \tag{15}$$

where $u_{i,t} \sim U(0,1)$ for $i = 1, 2, \cdots, n$ and $t = 1, 2, \cdots, T$.

If the range of the state-variable α_t is from $-\infty$ to $+\infty$, Mariano and Tanizaki (1995), Tanizaki (1996) and Tanizaki and Mariano (1994) suggested utilizing the following bimodal distribution for choice of the importance density:

$$P_I(\alpha_t) = \frac{1}{2}N(a^*_{t|t-1}, c\Sigma^*_{t|t-1}) + \frac{1}{2}N(a^*_{t|t}, c\Sigma^*_{t|t}),$$

which denotes the average of two normal densities, where $a^*_{t|s}$ and $\Sigma^*_{t|s}$ for $s = t-1, t$ denote the filtering means and variances obtained from the extended Kalman filter and c is a constant value which should be equal to or greater than one. Thus, we construct the importance density based on the extended Kalman filter estimates, i.e., the importance density is obtained independently of the importance sampling filter.

Now, it might be natural that use of the above bimodal distribution generates the following three questions:

(i) Why do we use the bimodal distribution consisting of the two normal densities?

(ii) Why do we use the first- and second-moments of the extended Kalman filter?

(iii) Why should c be equal to or greater than one?

The answers to the questions are as follows:

(i) The importance density $P_I(\alpha_t)$ needs to cover the two target densities $P(\alpha_t|Y_s)$ for $s = t-1, t$ over the range of α_t. Usually, the peak and range of the one-step ahead prediction density $P(\alpha_t|Y_{t-1})$ is different from those of the filtering density $P(\alpha_t|Y_t)$. In general, the range of $P(\alpha_t|Y_{t-1})$ is larger than that of $P(\alpha_t|Y_t)$. For the importance sampling filter, the two densities $P(\alpha_t|Y_s)$, $s = t-1, t$, have to be approximated by one importance density $P_I(\alpha_t)$ in order to obtain the recursive algorithm. Therefore, it might be plausible to take the bimodal distribution for the importance density.

(ii) Also, the peak and range of $P(\alpha_t|Y_s)$ for $s = t-1, t$ are not known, but mean and variance of the state-variable can be estimated by the extended Kalman filter algorithm, even if the extended Kalman filter indicates the biased estimators. It is appropriate to consider that the extended Kalman filter estimates are not too far from the true values. Therefore, the importance sampling filter would be improved by utilizing the importance density based on the extended Kalman filter.

(iii) It is known that the support of the importance sampling distribution needs to include that of the target density (see, for example, Geweke (1996)). In a framework of filtering theory, we have the two target densities, i.e., $P(\alpha_t|Y_{t-1})$ and $P(\alpha_t|Y_t)$. In the above bimodal density, $N(a^*_{t|t-1}, c\Sigma^*_{t|t-1})$ is used to cover $P(\alpha_t|Y_{t-1})$ while $N(a^*_{t|t}, c\Sigma^*_{t|t})$ is $P(\alpha_t|Y_t)$. Accordingly, the support of $N(a^*_{t|t-1}, c\Sigma^*_{t|t-1})$ should include that of the one-step ahead prediction density $P(\alpha_t|Y_{t-1})$ and the support of $P(\alpha_t|Y_t)$ should be included in the support of $N(a^*_{t|t}, c\Sigma^*_{t|t})$. Since the support of $P(\alpha_t|Y_s)$ is not known in general, it is safe that the importance density $P_I(\alpha_t)$ should be more broadly distributed than the target densities $P(\alpha_t|Y_s)$, $s = t-1, t$. Therefore, it might be appropriate that $c \geq 1$ should be chosen.

Based on the above bimodal distribution, the random draws of α_t from $P_I(\alpha_t)$ are generated by:

$$\alpha_{i,t} = a + \Sigma^{1/2} z_{i,t}, \tag{16}$$

where $z_{i,t} \sim N(0,1)$ for $i = 1, 2, \cdots, n$ and $t = 1, 2, \cdots, T$. (a, Σ) takes the following discrete random variable:

$$(a, \Sigma) = \begin{cases} (a^*_{t|t-1}, c\Sigma^*_{t|t-1}), & \text{with probability } 1/2, \\ (a^*_{t|t}, c\Sigma^*_{t|t}), & \text{with probability } 1/2. \end{cases} \tag{17}$$

In order to choose $(a^*_{t|t-1}, c\Sigma^*_{t|t-1})$ or $(a^*_{t|t}, c\Sigma^*_{t|t})$ in (17), we need to generate a uniform random number between zero and one, say u. That is, we may choose $(a^*_{t|t-1}, c\Sigma^*_{t|t-1})$ when $u \leq 0.5$ and $(a^*_{t|t}, c\Sigma^*_{t|t})$ when $u > 0.5$.

Antithetic Monte Carlo Methods: The essence of the antithetic Monte Carlo method is as follows. Suppose that we have the sample $x_1, x_2, \cdots, x_n$ from the density function $P(x)$. Consider estimating mean $\mathrm{E}(x)$, which is simply estimated as the arithmetic average of $x_1, x_2, \cdots, x_n$, i.e., $\overline{x} = (1/n)\sum_{i=1}^n x_i$. Variance of the arithmetic average, $\mathrm{Var}(\overline{x})$, indicates precision of the estimated mean.

(i) When x_i, $i = 1, 2, \cdots, n$, are mutually independently distributed, i.e., when $x_i \sim N(0, \sigma^2)$ for $i = 1, 2, \cdots, n$, we have $\mathrm{Var}(\overline{x}) = \sigma^2/n$.

(ii) To make things easier, consider the following case: $\mathrm{Cov}(x_i, x_j) = \rho\sigma^2$ when $j = i + m$ and $\mathrm{Cov}(x_i, x_j) = 0$ otherwise, where $i = 1, 2, \cdots, m$ and $n = 2m$. When x_i and x_{i+m} are negatively correlated (i.e., $\rho < 0$) and $x_i \sim N(0, \sigma^2)$ for $i = 1, 2, \cdots, m$, we obtain $\mathrm{Var}(\overline{x}) = (1 + \rho)\sigma^2/n$, which is clearly less than variance in Case (i).

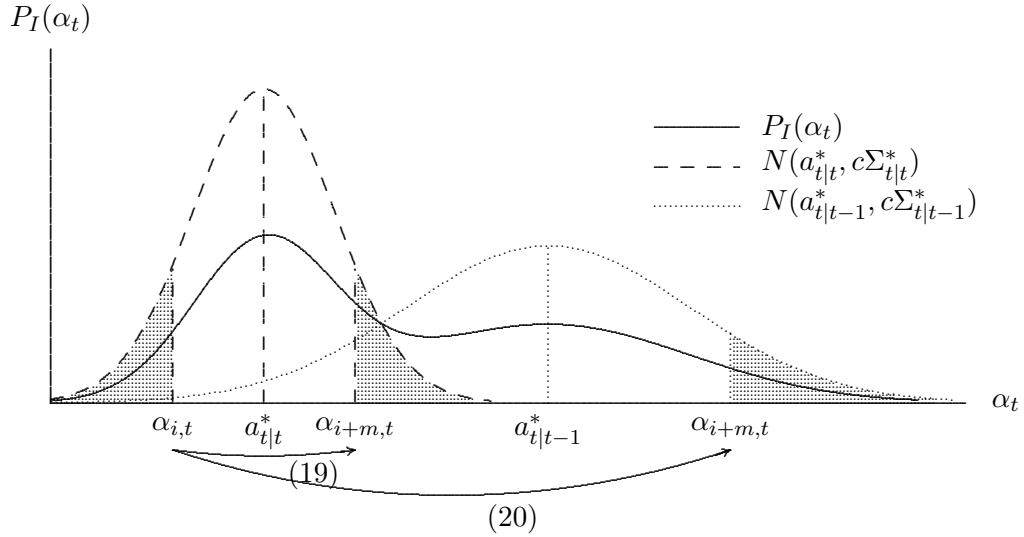
Let x_i, $i = 1, 2, \cdots, n$, be the random draws from $P(x)$. Then, variance of the estimated mean $\overline{x}$ is given by Case (i). However, in the case where x_i, $i = 1, 2, \cdots, m$, are the random draws from $P(x)$, if $\mathrm{Cov}(x_i, x_{i+m}) < 0$, the estimated mean in Case (ii) has less variance than that in Case (i). Accordingly, negatively correlated sample is more optimal than random sample in a sense of reduction of the simulation errors. Similarly, for a function $g(\cdot)$, mean of $g(x)$ is estimated as the arithmetic average of $g(x_1), g(x_2), \cdots, g(x_n)$. If $g(x_i)$ is negatively correlated with $g(x_{i+m})$, the obtained estimate of $\mathrm{E}(g(x))$ performs better. In this paper, we apply the above antithetic Monte Carlo procedure to the importance sampling filter proposed by Tanizaki (1996), Tanizaki and Mariano (1994).

First, in the case where α_t lies on an interval, for example, $0 < \alpha_t < 1$, we may choose $P_I(\alpha_t)$ as a uniform distribution between zero and one. In order to have antithetic relationship, we use the following sample $\alpha_{i,t}$, $i = 1, 2, \cdots, n$.

$$\alpha_{i,t} = u_{i,t}, \qquad \alpha_{i+m,t} = 1 - u_{i,t}, \tag{18}$$

where $u_{i,t} \sim U(0,1)$ for $i = 1, 2, \cdots, m$ and $n = 2m$. Clearly, it is expected that (18) is more powerful than (15) because of less variance.

FIGURE 1: Location of $\alpha_{i,t}$ and $\alpha_{i+m,t}$



Second, consider the bimodal distribution used in the case where the range of the state-variable α_t is from $-\infty$ to $+\infty$. For the variance reduction method, we can consider the following two procedures:

(i) We can take $\alpha_{i,t}$ as follows:

$$\alpha_{i,t} = a + \Sigma^{1/2} z_{i,t}, \qquad \alpha_{i+m,t} = a + \Sigma^{1/2}(-z_{i,t}), \tag{19}$$

where $z_{i,t} \sim N(0,1)$ for $i = 1, 2, \cdots, m$, $n = 2m$ and $t = 1, 2, \cdots, T$. Note that (a, Σ) is given by (17). The above procedure implies that we have both $\alpha_{i,t}$ and $\alpha_{i+m,t}$ simultaneously when $\alpha_{i,t}$ is generated from the bimodal distribution.

(ii) Also, we can consider the following procedure:

$$\alpha_{i,t} = a + \Sigma^{1/2} z_{i,t}, \qquad \alpha_{i+m,t} = \widetilde{a} + \widetilde{\Sigma}^{1/2}(-z_{i,t}), \tag{20}$$

where $(\widetilde{a}, \widetilde{\Sigma})$ takes as follows:

$$(\widetilde{a}, \widetilde{\Sigma}) = \begin{cases} (a^*_{t|t-1}, c\Sigma^*_{t|t-1}), & \text{if } (a, \Sigma) = (a^*_{t|t}, c\Sigma^*_{t|t}), \\ (a^*_{t|t}, c\Sigma^*_{t|t}), & \text{if } (a, \Sigma) = (a^*_{t|t-1}, c\Sigma^*_{t|t-1}), \end{cases}$$

where (a, Σ) is given by (17).

It is easily expected that both (19) and (20) yield less simulation errors than (16). Figure 1 displays the two transformations (i.e., (19) and (20)) from $\alpha_{i,t}$ to $\alpha_{i+m,t}$. Note that all the dotted areas are equal. When $P(\alpha_t|Y_{t-1})$ is away from $P(\alpha_t|Y_t)$, i.e., when $N(a^*_{t|t-1}, c\Sigma^*_{t|t-1})$ is away from $N(a^*_{t|t}, c\Sigma^*_{t|t})$, (20) might be better than (19), which implies that $\alpha_{i,t}$ and $\alpha_{i+m,t}$ of (20)

are more negatively correlated than those of (19). For the antithetic Monte Carlo method (19), note that $\alpha_{i,t}$ has positive correlation with $\alpha_{i+m,t}$ when $N(a^*_{t|t-1}, c\Sigma^*_{t|t-1})$ is too away from $N(a^*_{t|t}, c\Sigma^*_{t|t})$. However, in the case where $P(\alpha_t|Y_{t-1})$ is close to $P(\alpha_t|Y_t)$, (19) might be recommended rather than (20). See Appendix B for the two antithetic Monte Carlo methods (19) and (20).

In the next section, some Monte Carlo studies are performed, where the conventional Monte Carlo method given by (15) and (16) is compared with the antithetic Monte Carlo approach represented by (18) – (20).

4 MONTE CARLO STUDIES

In this section, we perform some Monte Carlo simulation studies to check whether the antithetic Monte Carlo methods work well or not in a nonlinear filtering framework. We compare the extended Kalman filter (EKF), the resampling filter (RSF) and the importance sampling filters (ISF) with and without the antithetic Monte Carlo methods. Note that the resampling filter (RSF) proposed by Kitagawa (1996) is concisely described in Appendix A.

The simulation procedure is as follows:

(i) Generating random numbers of the error terms ϵ_t and η_t, we obtain a set of data y_t and α_t, $t = 1, 2, \cdots, T$, from the system (1) and (2), where $T = 25, 50$ is taken.

(ii) Given data $Y_T = \{y_1, y_2, \cdots, y_T\}$, obtain the filtering estimates by the extended Kalman filter (EKF), the resampling filter (RSF) and the importance sampling filters (ISF (15) and ISF (18), or ISF (16), ISF (19) and ISF (20)). We choose $n = 20, 50, 100, 500$ for RSF and $n = 20, 50, 100$ for ISF's.

(iii) Repeat (i) and (ii) G times and compute the root mean square error (RMSE) for each estimate, which is defined as:

$$\text{RMSE} = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{1}{G}\sum_{g=1}^{G}(a^{(g)}_{t|t} - \alpha^{(g)}_t)^2\right)^{1/2},$$

where $a_{t|t}$ takes the extended Kalman filter estimate, the resampling filter estimate and the importance sampling filter estimates. The superscript (g) denotes the g-th simulation run and we take $G = 1000$. That is, $\alpha^{(g)}_t$ denotes the simulated state variable at time t in the g-th simulation run.

Under the above simulation procedure, we examine five Monte Carlo studies, i.e., Linear and Normal Model, ARCH Model, Stochastic Volatility Model, Nonstationary Growth Model and Logistic Model. Simulation results and some discussion are given in Section 4.1

Simulation I: Linear and Normal Model (Table 1): Consider the following linear scalar system:

$$\begin{aligned} &\text{(Measurement Equation)} \quad && y_t = \alpha_t + \epsilon_t, \\ &\text{(Transition Equation)} && \alpha_t = \delta\alpha_{t-1} + \eta_t, \end{aligned} \tag{21}$$

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

for $t = 1, 2, \cdots, T$ and $\delta = 0.5, 0.9, 1.0$.

In this simulation study, the extended Kalman filter is given by the conventional linear recursive formula (i.e., the extended Kalman filter reduces to the standard Kalman filter in this case), because the system is linear and normal. Therefore, it is expected that EKF is better than ISF's and RSF.

Simulation II: ARCH Model (Table 2): The scalar system with the nonlinear transition equation is given by:

$$\begin{aligned} &\text{(Measurement Equation)} \quad && y_t = \alpha_t + \epsilon_t, \\ &\text{(Transition Equation)} && \alpha_t = (1 - \delta + \delta\alpha_{t-1}^2)^{1/2}\eta_t, \end{aligned} \tag{22}$$

where $t = 1, 2, \cdots, T$ and $\delta = 0.5, 0.9$. The transition equation follows the first-order autoregressive conditional heteroscedasticity (ARCH(1)) process, while the measurement equation consists of the ARCH(1) term and the error. Note that the unconditional variance of α_t is assumed to be one in this Monte Carlo experiment. For α_0 and $(\epsilon_t, \eta_t)'$, we make the same assumptions as in Simulation I (in Simulations III and V, similarly, we assume the same density functions as in Simulation I for α_0 and $(\epsilon_t, \eta_t)'$). The measurement equation is linear but the transition equation is nonlinear. In this experiment, therefore, it might be expected that EKF does not show a good performance.

Simulation III: Stochastic Volatility Model (Table 3): Consider the following nonlinear system:

$$\begin{aligned} &\text{(Measurement Equation)} \quad && y_t = \exp\left(\frac{1}{2}\alpha_t\right)\epsilon_t, \\ &\text{(Transition Equation)} && \alpha_t = \delta\alpha_{t-1} + \eta_t, \end{aligned} \tag{23}$$

where $0 \leq \delta < 1$ and $t = 1, 2, \cdots, T$. $\delta = 0.5, 0.9$ is taken. The transition equation follows the first-order autoregressive process, while the measurement equation denotes the nonlinear function of the state variable and the error. Note that in this simulation study the antithetic Monte Carlo method (19) is equivalent to (20) because one-step ahead prediction estimates are equal to the

filtering estimates in the extended Kalman filter algorithm, i.e., $a^*_{t|t} = a^*_{t|t-1}$ and $\Sigma^*_{t|t} = \Sigma^*_{t|t-1}$.

It is known that the measurement equation in the system (23) gives us a poor approximation. Instead, taking the logarithm on both sides, the measurement equation in the system (23) is rewritten as:

$$\begin{aligned} &\text{(Measurement Equation)} \quad \log(y_t^2) = \alpha_t + \log(\epsilon_t^2), \\ &\text{(Transition Equation)} \quad \alpha_t = \delta\alpha_{t-1} + \eta_t, \end{aligned} \tag{24}$$

where $\log(\epsilon_t^2)$ is known to be distributed as mean -1.27 and variance $\pi^2/2$ (for example, see Ghysels, Harvey and Renault (1996) and Sandmann and Koopman (1998) for mean and variance of the stochastic volatility model). Therefore, we may use the measurement equation in (24), rather than that in (23). Note that the state-space model (24) reduces to the linear and non-Gaussian system. Accordingly, the Kalman filter estimate is an optimal estimator in the sense that it minimizes the mean square error within the class of all linear estimators, which implies that the Kalman filter under the normality assumption is the minimum mean square estimator and that the Kalman filter without the normality assumption is known as the minimum mean square linear estimator (see Harvey (1989)).

In this simulation study, we examine the above two kinds of the system, i.e., (23) and (24). See Ghysels, Harvey and Renault (1996), Shepherd (1996) and Sandmann and Koopman (1998) for the stochastic volatility model.

Simulation IV: Nonstationary Growth Model (Table 4): Consider an example of the following univariate nonstationary growth model:

$$\begin{aligned} &\text{(Measurement Equation)} \quad y_t = \frac{\alpha_t^2}{20} + \epsilon_t, \\ &\text{(Transition Equation)} \quad \alpha_t = \frac{1}{2}\alpha_{t-1} + \frac{25\alpha_{t-1}}{1+\alpha_{t-1}^2} \\ &\qquad\qquad\qquad\qquad +8\cos(1.2(t-1)) + \eta_t, \end{aligned} \tag{25}$$

$$\alpha_0 \sim N(0,1), \quad \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 10 \end{pmatrix} \right),$$

where $t = 1, 2, \cdots, T$. The system (25) are taken in Kitagawa (1987, 1996) and Carlin, Polson and Stoffer (1992). It is called the nonstationary growth model in Carlin, Polson and Stoffer (1992).

Simulation V: Logistic Model (Table 5): Next, we take the following logistic system:

$$\begin{aligned} &\text{(Measurement Equation)} \qquad y_t = \frac{\exp(\alpha_t)}{\exp(\alpha_t) + \exp(\epsilon_t)}, \\ &\text{(Transition Equation)} \qquad \alpha_t = \frac{\exp(\alpha_{t-1})}{\exp(\alpha_{t-1}) + \exp(\eta_t)}, \end{aligned} \tag{26}$$

where $t = 1, 2, \cdots, T$. In this experiment, both the measurement and the transition equations are logistic, which implies that the system is nonlinear. α_t lies on the interval between zero and one. Therefore, we apply (15) and (18) for the random draws generated from the importance density.

Note as follows. The measurement equation can be transformed into a linear function in the state variable, which is shown as:

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

For the Kalman filtering estimate, however, equation (26), rather than the above equation, is linearized in this experiment.

4.1 RESULTS AND DISCUSSION

The results are in Tables 1 – 5, where all the values denote the RMSE's, defined above. $n = 20, 50, 100, 500$ and $c = 1, 2, 4, 9, 16, 25$ are taken. In each table, the superscript $*$ denotes the minimum RMSE of $c = 1, 2, 4, 9, 16, 25$ for each of ISF's. Therefore, ISF's with $*$ perform better than ISF's without $*$ in the sense of RMSE. In the simulation studies, ISF's with $*$ are compared with EKF and RSF. Note that EKF does not depend on both c and n while RSF is independent of c.

In Linear and Normal Model (Table 1), $c = 2$ is the best choice for $\delta = 0.5, 0.9, 1.0$. As c is large, RMSE's become larger, which implies that too widely distributed importance density leads to a poor approximation of the filtering density. The two antithetic Monte Carlo variance reduction methods, i.e., ISF (19) and ISF (20), perform better than the conventional Monte Carlo method, i.e., ISF (16), especially when n is small. The three ISF's are close to each other as n increases and also they are very close to EKF. ISF (19) is not too different from ISF (20) but the former is slightly better than the latter. Under the same number of random draws, RSF is the worst estimator. ISF (19) with $n = 100$ is similar to RSF with $n = 500$. Therefore, RSF needs more random draws than ISF to keep the same precision.

In Table 2, ARCH(1) model is taken as an example. When $\delta^2 < 1/3$, the fourth-moment of α_t does not exists, which implies that α_t is widely distributed with fat tails. Therefore, $c = 25$ gives us the smallest RMSE when $\delta = 0.9$,

while $c = 4$ is the smallest RMSE when $\delta = 0.5$. In the case of $\delta = 0.9$, $c = 25$ and $n = 100$, ISF's perform much better than EKF.

For the stochastic volatility model in Table 3, we estimate two types of the state-space models, i.e., (23) and (24). Note for the system (23) in Table 3 that the one-step ahead prediction estimates are exactly equal to the filtering estimates in the extended Kalman filter. Therefore, there is no difference between ISF (19) and ISF (20). However, when we use the system (24), ISF (19) is different from ISF (20). Since ISF is based on the extended Kalman filter estimates, ISF with (23) is not equivalent to ISF with (24). For RSF, note that there is no difference between (23) and (24) because RSF does not depend on the extended Kalman filter. From the table, the system (24) is preferred to the system (23), where RMSE's from the system (24) are smaller than those from the system (23). In this example, the case of $c = 2$ is recommended for the systems (23) and (24).

In Table 4, an example of the nonstationary growth model is taken. $c = 25$ might be recommended in this simulation study. In almost all the cases, RMSE's of ISF (20) are smaller than those of ISF (19). We can take the system (25) as the case where $P(\alpha_t|Y_{t-1})$ is too far from $P(\alpha_t|Y_t)$. Therefore, in this example, we should take ISF (20) for the antithetic Monte Carlo method. It is surprising that RSF with $n = 20$ is close to ISF's with $n = 100$. In the case of strong nonlinearity, RSF is the best estimator even when n is small.

In the logistic model of Table 5, (18) is taken as the antithetic Monte Carlo method. In the other simulation studies the bimodal distribution is used for the importance density, but in this example the uniform distribution is taken as the importance density. We have the same conclusion as the previous simulation studies, i.e., the antithetic Monte Carlo method is good especially when n is small and ISF's are better than RSF for the same number of random draws.

From the tables, we can conclude as follows.

(i) For both RSF and ISF, RMSE is small as n increases.

(ii) Except for Simulation IV (Table 4), RSF is worse than ISF's in the case of the same number of random draws. However, with no doubt, RSF becomes better than any other estimators for large n.

(iii) ISF's are better than EKF in the nonlinear systems (see Tables 2 – 5). Even in the linear and normal system, ISF's are very close to EKF (see Table 1).

(iv) ISF's with the variance reduction methods perform better than ISF without them. Especially, ISF's with the variance reduction methods are significantly better when n is small. As n increases, the ISF's with and without the variance reduction methods are close to each other.

(v) For ISF's, $c = 2, 4$ is better in almost all the cases. However, we should choose a large value for c in the case where $P_\alpha(\alpha_t|\alpha_s)$ is widely distributed. See $\delta = 0.9$ of Table 2 and Table 4.

TABLE 1: Simulation I: Linear and Normal Model (21)

Estimation		T	25			50		
Method	n	$c \setminus \delta$	0.5	0.9	1.0	0.5	0.9	1.0
EKF			0.7321	0.7783	0.7917	0.7347	0.7808	0.7943
ISF (16)	20	1	0.7507	0.8006	0.8154	0.7533	0.8035	0.8185
		2	0.7499*	0.7996*	0.8143*	0.7522*	0.8022*	0.8170*
		4	0.7536	0.8051	0.8204	0.7558	0.8069	0.8221
		9	0.7633	0.8180	0.8343	0.7654	0.8199	0.8364
	50	1	0.7404	0.7880*	0.8019*	0.7422	0.7894	0.8034
		2	0.7402*	0.7880*	0.8019*	0.7412*	0.7885*	0.8024*
		4	0.7414	0.7895	0.8036	0.7423	0.7902	0.8044
		9	0.7447	0.7934	0.8077	0.7451	0.7937	0.8081
	100	1	0.7352	0.7821	0.7957	0.7384	0.7850	0.7988
		2	0.7348*	0.7818*	0.7955*	0.7380*	0.7847*	0.7984*
		4	0.7352	0.7824	0.7962	0.7387	0.7855	0.7994
		9	0.7364	0.7838	0.7976	0.7401	0.7872	0.8011
ISF (19)	20	1	0.7409	0.7902	0.8048	0.7433	0.7928	0.8077
		2	0.7379*	0.7872*	0.8018*	0.7416*	0.7905*	0.8052*
		4	0.7393	0.7895	0.8046	0.7430	0.7931	0.8082
		9	0.7466	0.8014	0.8184	0.7493	0.8040	0.8208
	50	1	0.7354	0.7828	0.7967	0.7381	0.7853	0.7992
		2	0.7345*	0.7815*	0.7952*	0.7371*	0.7842*	0.7981*
		4	0.7348	0.7821	0.7959	0.7375	0.7850	0.7990
		9	0.7364	0.7844	0.7985	0.7385	0.7868	0.8010
	100	1	0.7342	0.7810	0.7947	0.7363	0.7829	0.7966
		2	0.7340*	0.7806*	0.7942*	0.7360*	0.7825*	0.7961*
		4	0.7340*	0.7807	0.7944	0.7362	0.7830	0.7968
		9	0.7346	0.7815	0.7951	0.7367	0.7836	0.7975
ISF (20)	20	1	0.7420	0.7919	0.8067	0.7441	0.7940	0.8089
		2	0.7400*	0.7909*	0.8060*	0.7429*	0.7928*	0.8077*
		4	0.7430	0.7960	0.8118	0.7458	0.7979	0.8135
		9	0.7539	0.8118	0.8291	0.7560	0.8130	0.8302
	50	1	0.7357	0.7835	0.7976	0.7393	0.7869	0.8009
		2	0.7353*	0.7833*	0.7973*	0.7383*	0.7856*	0.7996*
		4	0.7363	0.7854	0.7997	0.7386	0.7865	0.8006
		9	0.7396	0.7903	0.8051	0.7403	0.7896	0.8042
	100	1	0.7342*	0.7810*	0.7947*	0.7365	0.7831	0.7969
		2	0.7342*	0.7814	0.7951	0.7362*	0.7829*	0.7967*
		4	0.7347	0.7826	0.7965	0.7367	0.7839	0.7978
		9	0.7372	0.7860	0.7999	0.7380	0.7855	0.7994
RSF	20		0.7888	0.8573	0.8837	0.7897	0.8572	0.8825
	50		0.7542	0.8110	0.8270	0.7549	0.8096	0.8271
	100		0.7433	0.7922	0.8072	0.7453	0.7954	0.8090
	500		0.7339	0.7809	0.7944	0.7369	0.7837	0.7970

TABLE 2: Simulation II: ARCH Model (22)

Estimation		T	25		50	
Method	n	$c \setminus \delta$	0.5	0.9	0.5	0.9
EKF			0.7038	0.6654	0.7050	0.6356
ISF (16)	20	2	0.7069*	0.6310	0.7103	0.6058
		4	0.7070	0.6113	0.7099*	0.5881
		9	0.7137	0.5925	0.7157	0.5732
		16	0.7238	0.5842	0.7243	0.5676*
		25	0.7357	0.5817*	0.7354	0.5677
	50	2	0.6957	0.6118	0.6997	0.5931
		4	0.6952*	0.5886	0.6986*	0.5747
		9	0.6975	0.5696	0.7002	0.5590
		16	0.7005	0.5617	0.7026	0.5535
		25	0.7034	0.5583*	0.7054	0.5523*
	100	2	0.6899	0.6019	0.6959	0.5855
		4	0.6893*	0.5800	0.6953*	0.5689
		9	0.6904	0.5607	0.6963	0.5554
		16	0.6915	0.5524	0.6973	0.5499
		25	0.6925	0.5497*	0.6984	0.5483*
ISF (19)	20	2	0.6992	0.6214	0.7014	0.5998
		4	0.6976*	0.6027	0.7008*	0.5841
		9	0.7032	0.5842	0.7064	0.5698
		16	0.7144	0.5787*	0.7173	0.5662*
		25	0.7324	0.5812	0.7356	0.5705
	50	2	0.6911	0.6084	0.6960	0.5888
		4	0.6902*	0.5870	0.6948*	0.5719
		9	0.6917	0.5673	0.6953	0.5576
		16	0.6934	0.5582	0.6967	0.5522
		25	0.6952	0.5544*	0.6986	0.5507*
	100	2	0.6900	0.6017	0.6941	0.5825
		4	0.6895*	0.5800	0.6935*	0.5663
		9	0.6899	0.5612	0.6939	0.5529
		16	0.6907	0.5539	0.6946	0.5490
		25	0.6915	0.5512*	0.6951	0.5481*
ISF (20)	20	2	0.7015	0.6205	0.7027*	0.5995
		4	0.7007*	0.6025	0.7029	0.5837
		9	0.7084	0.5852	0.7110	0.5703
		16	0.7226	0.5804*	0.7243	0.5674*
		25	0.7437	0.5834	0.7452	0.5721
	50	2	0.6926*	0.6067	0.6975	0.5885
		4	0.6926*	0.5853	0.6963*	0.5713
		9	0.6954	0.5669	0.6973	0.5577
		16	0.6984	0.5597	0.6991	0.5527
		25	0.7017	0.5573*	0.7016	0.5514*
	100	2	0.6909*	0.6000	0.6946	0.5814
		4	0.6912	0.5787	0.6945*	0.5649
		9	0.6934	0.5610	0.6957	0.5525
		16	0.6953	0.5550	0.6967	0.5491
		25	0.6970	0.5535*	0.6976	0.5486*
RSF	20		0.7547	0.7527	0.7555	0.6786
	50		0.7154	0.5995	0.7180	0.5813
	100		0.7028	0.5697	0.7047	0.5610
	500		0.6895	0.5421	0.6943	0.5422

TABLE 3: Simulation III: Stochastic Volatility Models (23) and (24)

State-Space Model	Estimation Method	n	c	T \ δ	25 0.5	25 0.9	50 0.5	50 0.9
(23)	EKF				1.1613	2.1569	1.1542	2.2222
	ISF (16)	20	1		0.9722	1.2265	0.9678	1.2335
			2		0.9613*	1.1953*	0.9586*	1.2006*
			4		0.9634	1.2036	0.9621	1.2190
		50	1		0.9496	1.1528	0.9463	1.1577
			2		0.9442*	1.1326*	0.9407*	1.1378*
			4		0.9455	1.1383	0.9412	1.1392
		100	1		0.9400	1.1320	0.9369	1.1308
			2		0.9370*	1.1217	0.9344*	1.1214*
			4		0.9372	1.1214*	0.9355	1.1222
	ISF (19)	20	1		0.9580	1.2147	0.9561	1.2204
			2		0.9498*	1.1854*	0.9492*	1.1950*
			4		0.9551	1.2095	0.9542	1.2269
		50	1		0.9418	1.1499	0.9415	1.1543
			2		0.9390*	1.1333*	0.9373*	1.1355*
			4		0.9406	1.1374	0.9383	1.1372
		100	1		0.9388	1.1277	0.9362	1.1309
			2		0.9361*	1.1168*	0.9335*	1.1201*
			4		0.9365	1.1205	0.9335*	1.1214
(24)	EKF				1.0071	1.2507	0.9965	1.2399
	ISF (16)	20	1		0.9616	1.1578	0.9588	1.1568
			2		0.9569*	1.1504*	0.9543*	1.1509*
			4		0.9610	1.1573	0.9586	1.1603
		50	1		0.9464	1.1292	0.9420	1.1286
			2		0.9440*	1.1264*	0.9391*	1.1232*
			4		0.9451	1.1300	0.9402	1.1244
		100	1		0.9375	1.1163	0.9355	1.1168
			2		0.9359*	1.1135*	0.9343*	1.1146*
			4		0.9364	1.1156	0.9354	1.1168
	ISF (19)	20	1		0.9466	1.1412	0.9445	1.1443
			2		0.9429*	1.1343*	0.9419*	1.1389*
			4		0.9459	1.1399	0.9452	1.1477
		50	1		0.9372	1.1208	0.9357	1.1224
			2		0.9368*	1.1183*	0.9342*	1.1194*
			4		0.9377	1.1213	0.9346	1.1204
		100	1		0.9359	1.1139	0.9326	1.1156
			2		0.9352*	1.1128*	0.9317*	1.1124*
			4		0.9354	1.1146	0.9319	1.1127
	ISF (20)	20	1		0.9466	1.1414	0.9445	1.1443
			2		0.9432*	1.1355*	0.9420*	1.1396*
			4		0.9466	1.1441	0.9457	1.1503
		50	1		0.9372	1.1211	0.9357	1.1224
			2		0.9369*	1.1192*	0.9343*	1.1200*
			4		0.9382	1.1248	0.9351	1.1226
		100	1		0.9360	1.1143	0.9326	1.1156
			2		0.9353*	1.1138*	0.9318*	1.1130*
			4		0.9359	1.1186	0.9323	1.1150
	RSF	20			0.9885	1.2191	0.9913	1.2307
		50			0.9581	1.1530	0.9546	1.1568
		100			0.9450	1.1298	0.9423	1.1310
		500			0.9348	1.1100	0.9325	1.1115

TABLE 4: Simulation IV: Nonstationary Growth Model (25)

Estimation Method	n	c \ T	25	50
EKF			22.0200	20.5308
ISF (16)	20	9	14.0078	13.8635
		16	13.5467	13.2938
		25	13.2408*	13.0008*
	50	9	11.9598	12.0086
		16	10.6172	10.5155
		25	9.7394*	9.8271*
	100	9	10.1727	10.1441
		16	8.2233	8.2563
		25	7.3472*	7.4668*
ISF (19)	20	9	14.5901	13.9493
		16	13.8567	13.2146
		25	13.6782*	12.8041*
	50	9	12.0995	12.1316
		16	10.8665	10.6546
		25	9.9000*	10.0330*
	100	9	10.2235	10.1331
		16	8.4028	8.3700
		25	7.6416*	7.4050*
ISF (20)	20	9	14.0817	13.5841
		16	13.3809	13.1203
		25	13.3197*	12.6800*
	50	9	11.8390	11.8826
		16	10.5076	10.4571
		25	9.6522*	9.5233*
	100	9	10.0773	10.1798
		16	8.4122	8.2378
		25	7.2359*	7.3390*
RSF	20		7.3768	7.5378
	50		5.8327	5.9042
	100		5.1234	5.1886
	500		4.6093	4.7345

TABLE 5: Simulation V: Logistic Model (26)

Estimation Method	n \ T	25	50
EKF		0.2147	0.2150
ISF (15)	20	0.2029	0.2032
	50	0.1991	0.1999
	100	0.1986	0.1989
ISF (18)	20	0.2002	0.2014
	50	0.1987	0.1993
	100	0.1979	0.1986
RSF	20	0.2078	0.2077
	50	0.2008	0.2018
	100	0.1994	0.1999
	500	0.1977	0.1983

5 SUMMARY

In this paper, we have extended the importance sampling filter proposed by Mariano and Tanizaki (1995), Tanizaki (1996), Tanizaki and Mariano (1994) using the antithetic Monte Carlo methods to reduce the simulation errors. In a nonlinear filtering framework, they suggested utilizing the bimodal distribution function for the importance density function, where the first- and second-moments of the extended Kalman filter are used. According to the simulation techniques such as Monte Carlo integration, the simulation errors decrease as number of random draws (i.e., n) increases, but it is well known that convergence speed is quite slow as $\sqrt{n}$. Therefore, in this paper, we have applied the antithetic Monte Carlo methods to the importance sampling filter.

In Section 4, we performed several simulation studies to examine whether the importance sampling filters with the antithetic Monte Carlo methods work well or not. The importance sampling filters with and without the antithetic Monte Carlo methods are compared with the extended Kalman filter and the resampling filter. The results obtained from the simulation studies are as follows. The importance sampling filters with the variance reduction methods show a good performance especially when number of random draws (i.e., n) is small. As n increases, ISF (16), ISF (19) and ISF (20) (or ISF (15) and ISF (18)) are close to each other. $c = 2, 3$ is better in almost all the cases, while a large value of c should be chosen in the case where $P_\alpha(\alpha_t|\alpha_{t-1})$ is widely distributed (see $\delta = 0.9$ of Table 2 and Table 4). Moreover, under the same number of random draws, the importance sampling filters perform better than the resampling filter in Simulations I – III and V. However, when the system includes strong nonlinearity as in Simulation IV, the resampling filter becomes the best estimator.

APPENDICES

APPENDIX A: RESAMPLING FILTER

The resampling procedure proposed by Gordon, Salmond and Smith (1993) and Kitagawa (1996) are described as follows.

Equation (3) is approximately represented as:

$$P(\alpha_t|Y_{t-1}) \approx \frac{1}{n}\sum_{j=1}^{n} P(\alpha_t|\alpha_{j,t-1|t-1}). \tag{27}$$

Therefore, $\alpha_{i,t|t-1}$ is generated as follows. Pick up $\alpha_{j,t-1|t-1}$ randomly (i.e., pick up j with equal probability) and generate a random number of η_t (i.e., $\eta_{i,t}$), and we have a random draw of α_t (i.e., $\alpha_{i,t|t-1}$) from the transition equation:

$\alpha_{i,t|t-1} = f_t(\alpha_{j,t-1|t-1}, \eta_{i,t})$. Thus, when $\alpha_{j,t-1|t-1}$ for $j = 1, 2, \cdots, n$ are available, $\alpha_{i,t|t-1}$ for $i = 1, 2, \cdots, n$ can be obtained from the transition equation.

For filtering, we consider generating the random draws of α_t (i.e., $\alpha_{i,t|t}$ for $i = 1, 2, \cdots, n$) from the filtering density $P(\alpha_t|Y_t)$, given $\alpha_{i,t|t-1}$ for $i = 1, 2, \cdots, n$. Using the random draws $\alpha_{i,t|t-1}$ for $i = 1, 2, \cdots, n$, equation (4) is approximately rewritten as follows:

$$P(\alpha_{i,t|t-1}|Y_t) \approx \frac{P_y(y_t|\alpha_{i,t|t-1})}{\sum_{j=1}^{n} P_y(y_t|\alpha_{j,t|t-1})}. \tag{28}$$

Note that $P(\alpha_{i,t|t-1}|Y_{t-1}) \approx 1/n$ is used to derive equation (28). That is, let x_i, $i = 1, 2, \cdots, n$, be the random draws from the density $P(x)$. Then, for all i and j, the probability which we have x_i should be equal to the probability which x_j occurs. In the case where we have n random draws of x, therefore, $P(x_i) \approx 1/n$ is obtained.

Equation (28) is interpreted as follows. The probability which α_t takes $\alpha_{j,t|t-1}$ is approximately given by $P(\alpha_{j,t|t-1}|Y_t)$. Accordingly, $\alpha_{i,t|t}$ is chosen from $\alpha_{j,t|t-1}$, $j = 1, 2, \cdots, n$, with probability $P(\alpha_{j,t|t-1}|Y_t)$. That is, the i-th random draw of α_t from $P(\alpha_t|Y_t)$ (i.e., $\alpha_{i,t|t}$) is resampled as:

$$\alpha_{i,t|t} = \begin{cases} \alpha_{1,t|t-1}, & \text{with probability } P(\alpha_{1,t|t-1}|Y_t), \\ \alpha_{2,t|t-1}, & \text{with probability } P(\alpha_{2,t|t-1}|Y_t), \\ \quad \vdots & \quad \vdots \\ \alpha_{n,t|t-1}, & \text{with probability } P(\alpha_{n,t|t-1}|Y_t). \end{cases}$$

Thus, in order to obtain random draws from the filtering density, first we have to compute $P(\alpha_{j,t|t-1}|Y_t)$ for all $j = 1, 2, \cdots, n$ using equation (28) and next obtain $\alpha_{i,t|t}$ for $i = 1, 2, \cdots, n$ by resampling $\alpha_{j,t|t-1}$ with probability $P(\alpha_{j,t|t-1}|Y_t)$, $j = 1, 2, \cdots, n$. In practice, a uniform random draw between zero and one (say, u) is generated and $\alpha_{j,t|t-1}$ is taken as $\alpha_{i,t|t}$ when $\omega_{j-1} \leq u < \omega_j$, where $\omega_j \equiv \sum_{m=1}^{j} P(\alpha_{m,t|t-1}|Y_t)$ and $\omega_0 \equiv 0$.

Based on the random draws $\alpha_{i,t|s}$ for $i = 1, 2, \cdots, n$, the filtering mean and variance are easily obtained as:

$$a_{t|s} = \frac{1}{n}\sum_{i=1}^{n} \alpha_{i,t|s}, \qquad \Sigma_{t|s} = \frac{1}{n}\sum_{i=1}^{n} (\alpha_{i,t|s} - a_{t|s})(\alpha_{i,t|s} - a_{t|s})',$$

where $s = t - 1, t$.

APPENDIX B: COMPARISON OF THE ANTITHETIC MONTE CARLO METHODS

In Section 3, we have introduced the two types of the antithetic Monte Carlo methods, i.e., (19) and (20). In this appendix, we examine which method gives us the better approximation.

TABLE 6: Correlation Coefficients between x_i and x_{i+m}

	$\mu \backslash \sigma$	1.0	1.5	2.0	2.5	3.0	4.0	5.0
(19)	0.0	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000	−1.000
	0.5	−0.883	−0.926	−0.951	−0.966	−0.975	−0.985	−0.990
	1.5	−0.283	−0.486	−0.630	−0.731	−0.799	−0.875	−0.916
	3.0	0.389	0.166	−0.055	−0.232	−0.383	−0.583	−0.706
	5.0	0.723	0.589	0.431	0.268	0.111	−0.146	−0.348
(20)	0.0	−1.000	−0.923	−0.800	−0.690	−0.600	−0.471	−0.385
	0.5	−1.000	−0.926	−0.805	−0.695	−0.605	−0.475	−0.388
	1.5	−1.000	−0.943	−0.837	−0.731	−0.639	−0.506	−0.412
	3.0	−1.000	−0.968	−0.895	−0.807	−0.725	−0.581	−0.478
	5.0	−1.000	−0.984	−0.943	−0.886	−0.822	−0.696	−0.585

Take an example of the density function: $P(x) = \frac{1}{2}N(0,1) + \frac{1}{2}N(\mu, \sigma^2)$, which corresponds to: $\alpha_t = x$, $a^*_{t|t} = 0$, $c\Sigma^*_{t|t} = 1$, $a^*_{t|t-1} = \mu$ and $c\Sigma^*_{t|t-1} = \sigma^2$ in the importance density $P_I(\alpha_t)$ in Section 3.

Generate random draws from $P(x)$, which are denoted by x_i, $i = 1, 2, \cdots, n$. Consider the two types of the antithetic Monte Carlo methods, which are shown in (19) and (20). Compute the correlation coefficient between x_i and x_{i+m}, $i = 1, 2, \cdots, m$, where $n = 2m$. We take $n = 100000$, $\mu = 0.0, 0.5, 1.5, 3.0, 5.0$ and $\sigma = 1.0, 1.5, 2.0, 2.5, 3.0, 4.0, 5.0$.

The results are shown in Table 6. For (19), x_i is positively correlated with x_{i+m} when $N(0,1)$ is too far from $N(\mu, \sigma^2)$. However, when $N(0,1)$ is close to $N(\mu, \sigma^2)$, (19) has strong negative correlation compared with (20). The antithetic method (20) always yields negative correlations between x_i and x_{i+m}, but the correlation becomes small as μ approaches zero. When σ^2 goes to one, (18) has strong negative correlation in spite of μ.

Therefore, it might be concluded that (19) should be chosen when σ^2 is large and/or μ is close to zero but (20) might be recommended when σ^2 is close to one and/or μ is close to large.

The obtained results can be applied to the importance density $P_I(\alpha_t)$ as follows: (19) is better than (20) when $a_{t|t-1}$ is close to $a_{t|t}$ and/or $c\Sigma_{t|t-1}$ is too different from $c\Sigma_{t|t}$. (20) should be chosen when $a_{t|t-1}$ is away from $a_{t|t}$ and/or $c\Sigma_{t|t-1}$ is close to $c\Sigma_{t|t}$. In practice, since $P(\alpha_t|Y_{t-1})$ is more widely distributed than $P(\alpha_t|Y_t)$ because of $Y_{t-1} \subset Y_t$, we should take (19) if $a_{t|t-1}$ is close to $a_{t|t}$ and choose (20) if $a_{t|t-1}$ is far from $a_{t|t}$.

ACKNOWLEDGEMENTS

The author would like to acknowledge two anonymous referees for careful review and useful suggestion.

REFERENCES

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

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, pp.493 – 500.

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

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

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

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

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

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

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

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

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

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

Gordon, N.J., Salmond, D.J. and Smith, A.F.M., 1993, "Novel Approach to Nonlinear / Non-Gaussian Bayesian State Estimation," *IEE Proceedings-F*, Vol.140, pp.107 – 113.

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

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.

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

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

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

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

Sandmann, G. and Koopman, J., 1998, "Estimation of Stochastic Volatility Models via Monte Carlo Maximum Likelihood," *Journal of Econometrics*, Vol.87, No.2, pp.271 – 301.

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

Shepherd, N., 1996, "Statistical Aspects of ARCH and Stochastic Volatility," in *Time Series Models: In Econometrics, Finance and Other Fields*, edited by Cox, D.R., Hinkely, D.V. and Barndorff-Nielsen, O.E., Chapman and Hall, London.

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

Tanizaki, H. , 1999, "Note on Nonlinear and Nonnormal Filter Using Rejection Sampling," *IEEE Transactions on Automatic Control*, 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 Stochastic Simulations," *Journal of Econometrics*, Vol.83, No.1, 2, pp.263 – 290.