# Nonlinear and nonnormal filters using Monte Carlo methods

Tanizaki, Hisashi

(Citation)
Computational Statistics & Data Analysis,25(4):417-439

(Issue Date)
1997-09

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90000572

# Nonlinear and Nonnormal Filters Using Monte-Carlo Methods*

Hisashi Tanizaki
Faculty of Economics, Kobe University, Nadaku, Kobe 657, JAPAN

**ABSTRACT:** In this paper, a non-linear and non-normal filter using Monte-Carlo simulation techniques is proposed, where the density function derived from the measurement equation and the random draws of the state-vector generated from the transition equation are utilized. The proposed filter has less computational burden and easier programming than the other non-linear and non-normal filters such as the numerical integration procedure and the Monte-Carlo integration approach. Furthermore, the proposed filter is extended to prediction and smoothing algorithms. Finally, by Monte-Carlo experiments, we compare the non-linear and non-normal procedures.

**KEY WORDS:** Prediction, Filtering, Smoothing, Numerical Integration, Monte-Carlo Integration, and Monte-Carlo Simulations.

# 1 INTRODUCTION

The filtering theory developed by Kalman (1960) and Kalman and Bucy (1961) has been extended in two directions, i.e., non-linearity and non-normality.

The non-linear measurement and transition equations are linearized by a Taylor series expansion and the linearized system is directly applied to the linear recursive algorithm. Thus, the extended Kalman filter, the second-order non-linear filter and so on are derived (Wishner, Tabaczynski and Athans (1969), Sorenson and Alspach (1971), Alspach and Sorenson (1972), Gelb (1974), Anderson and Moore (1979) and Tanizaki and Mariano (1996)). However, these non-linear filters based on the Taylor series expansions give us the biased estimators (see Tanizaki (1993b), Tanizaki and Mariano (1994) and Mariano and Tanizaki (1995)). As an important application of the nonlinear filter, Fahrmeir (1992), Tanizaki (1993a) and Fahrmeir and Tutz (1994) considered the dynamic generalized linear models, where the transition equation is assumed as the first-order autoregressive process while the measurement equation connected with discrete observations is taken as a cumulative distribution function.

Furthermore, Meinhold and Singpurwalla (1989) pointed out that the Kalman filter model is not robust under the normality assumption. The state-space model with non-normal errors has been developed based on the underlying density functions, where a recursive algorithm of density functions is derived. Kitagawa (1987) and Kramer and Sorenson (1988) evaluate each density by Numerical Integration (NI). Recently, an attempt has been made to apply simulation techniques in order to evaluate a non-linear and non-normal equation (for example, Brown and Mariano (1984, 1989), McFadden (1989) and Mariano and Brown (1983, 1989)). Tanizaki (1993b), Tanizaki and Mariano (1994) and Mariano and Tanizaki (1995) applied a simulation technique to the filtering theory, utilizing Monte-Carlo Integration with Importance sampling (MI)[1] to obtain each integration included in the recursive algorithm of densities. Carlin, Polson and Stoffer (1992) and Carter and Kohn (1994, 1996) proposed a non-linear and non-normal state-space modeling by Gibbs sampling[2] , which does not yield a recursive algorithm. Moreover, Kitagawa (1996) proposed a Monte-Carlo

*The author wishes to thank the editor and two anonymous referee for careful review and valuable suggestions.

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

[2] For the Gibbs sampler, see Carlin and Polson (1991), Gelfand, Hills, Racine-Poon and Smith (1990), Gelfand and Smith (1990), Geman and Geman (1984) and Zeger and Karim (1991). For the Gibbs sampler approach, we need to specify the prior densities of the nuisance parameters.

filter, where successive prediction and filtering conditional probability density functions are approximated by many of their realizations. The disadvantage of these non-linear and non-normal procedures based on the densities is to require a great amount of computational burden and a complicated computer programming. In addition, each conditional density function has to be calculated by hand based on the measurement and transition equations in order to make a program of each filter. Improving these problems, we propose a non-linear and non-normal filter, where random draws generated from the transition equation and the conditional densities obtained from the measurement equation are utilized (Monte-Carlo Approach, i.e., MA). The proposed filters have less computational burden and can be easily implemented. Also, the non-linear and non-normal procedures are extended to prediction and smoothing algorithms in the exactly same fashion. Finally, by Monte-Carlo experiments, we compare MA with NI and MI.

# 2 OVERVIEW OF NON-LINEAR AND/OR NON-NORMAL STATE-SPACE MODEL

We consider the following general non-linear and non-normal state-space model:

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

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

where $t = 1, \cdots, T$. A $p \times 1$ vector, $y_t$, is observable while a $k \times 1$ vector, $\alpha_t$, is unobserved. $h_t(\cdot)$ and $f_t(\cdot)$ are $p \times 1$ and $k \times 1$ vector functions, which are assumed to be known. $\epsilon_t$ and $\eta_t$ are mutually independent. The distribution functions of the error terms $\epsilon_t$ and $\eta_t$ may or may not be normal. Let $P(\cdot|\cdot)$ and $Y_s$ be the conditional density function and the information set up to time $s$, i.e., $Y_s = \{y_s, y_{s-1}, \cdots, y_1\}$. Define $P_y(y_t|\alpha_t)$ as the density function based on the measurement equation (1) and $P_\alpha(\alpha_t|\alpha_{t-1})$ as the density function obtained from the transition equation (2). The initial value of the state-vector, $\alpha_0$, is assumed to be distributed as $P_\alpha(\alpha_0)$.

We have two kinds of density-based algorithms. One is of a recursive type and the other is not. The former includes the NI procedure (Kitagawa (1987) and Kramer and Sorenson (1988)) and the MI approach (Tanizaki (1993b), Tanizaki and Mariano (1994) and Mariano and Tanizaki (1995)), while the latter gives us the Gibbs sampling procedure proposed by Carlin, Polson and Stoffer (1992) and Carter and Kohn (1994, 1996). The MA approach proposed in this paper belongs to the latter (i.e., nonrecursive algorithm).

Given the above conventional non-linear and non-normal state-space model (1) and (2), for evaluation of the state vector $\alpha_t$ we have three kinds of estimates, i.e., prediction, filtering and smoothing.

The density of $L$-step ahead prediction is represented as:

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

where $P_\alpha(\alpha_{t+L}|\alpha_{t+L-1})$ can be derived from the transition equation (2). Given the filtering density $P(\alpha_t|Y_t)$ and the density $P_\alpha(\alpha_{t+L}|\alpha_{t+L-1})$, the prediction density, i.e., $P(\alpha_{t+L}|Y_t)$ for $L = 1, 2, \cdots$, is recursively obtained.

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

$$P(\alpha_t|Y_{t-1}) = \int P_\alpha(\alpha_t|\alpha_{t-1})P(\alpha_{t-1}|Y_{t-1})\mathrm{d}\alpha_{t-1}, \tag{4}$$

$$P(\alpha_t|Y_t) = \frac{P_y(y_t|\alpha_t)P(\alpha_t|Y_{t-1})}{\int P_y(y_t|\alpha_t)P(\alpha_t|Y_{t-1})\mathrm{d}\alpha_t}, \tag{5}$$

where the initial condition is given by:

$$P(\alpha_0|Y_0) = P_\alpha(\alpha_0).$$

The densities $P_\alpha(\alpha_t|\alpha_{t-1})$ and $P_y(y_t|\alpha_t)$ are computed from the measurement equation (1) and the transition equation (2), respectively. Based on the two densities, equation (4) yields $P(\alpha_t|Y_{t-1})$ given $P(\alpha_{t-1}|Y_{t-1})$ and equation (5) yields $P(\alpha_t|Y_t)$ given $P(\alpha_t|Y_{t-1})$. Thus, repeating predicting and updating for all $t$, the filtering densities $P(\alpha_t|Y_t)$, $t = 1, \cdots, T$, can be obtained.

Also, the smoothing formula is represented as:

$$P(\alpha_t|Y_T) = P(\alpha_t|Y_t) \int \frac{P(\alpha_{t+1}|Y_T)P_\alpha(\alpha_{t+1}|\alpha_t)}{P(\alpha_{t+1}|Y_t)} \mathrm{d}\alpha_{t+1}, \tag{6}$$

where $P(\alpha_t|Y_t)$ and $P(\alpha_{t+1}|Y_t)$ are computed from the above filtering algorithm represented by equations (4) and (5), while $P_\alpha(\alpha_{t+1}|\alpha_t)$ is obtained from the transition equation (2). Smoothing is represented by the backward recursion (i.e., $t = T-1, T-2, \cdots, 1$) shown in equation (6).

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

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

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

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

$$\begin{aligned} P(Y_T) &= \prod_{t=1}^{T} P(y_t|Y_{t-1}) \\ &= \prod_{t=1}^{T} \left( \int P_y(y_t|\alpha_t)P(\alpha_t|Y_{t-1})\mathrm{d}\alpha_t \right). \end{aligned} \tag{8}$$

Note that we have the following equation:

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

which corresponds to the denominator of equation (5).

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

## 2.1 NUMERICAL INTEGRATION APPROACH (NI)

Kitagawa (1987) and Kramer and Sorenson (1988) proposed a non-linear and non-normal filter using NI. Consider a scalar case of the state vector. Let $\alpha_{i,t}$, $i = 0, \cdots, n$, be the nodes for numerical integration. For all $t$, $\alpha_{i,t}$ are sorted by size with respect to $i$, i.e., $\alpha_{0,t}$ is the smallest value and $\alpha_{n,t}$ the largest one. There are some methods to evaluate integration numerically; a sum of rectangles rule, a sum of trapezoids, Simpson's formula and so on. Here, for simplicity of discussion, we use the sum of rectangles method.

By NI, equation (3) is evaluated as:

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

which is called NI prediction.

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

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

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

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

Equation (6) is rewritten as:

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

which is called the NI smoothing algorithm.

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

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

According to the density approximation based on NI, each density is specified by the number of segments (i.e., $n$), location of nodes (i.e., $\alpha_{i,t}$), and the value at each node (i.e., $P_\alpha(\alpha_{i,t}|\alpha_{i,t-1})$ and $P_y(y_t|\alpha_{i,t})$). The nodes $\alpha_{i,t}$, $i = 1, \cdots, n$ and $t = 1, \cdots, T$, have to be chosen by a researcher.

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

$$P(Y_T) = \prod_{t=1}^{T} \left( \sum_{j=1}^{n} P_y(y_t|\alpha_{j,t})P(\alpha_{j,t}|Y_{t-1})(\alpha_{j,t} - \alpha_{j-1,t}) \right), \tag{9}$$

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

Since computational errors drastically grow up as $t$ increases, we need to impose as follows:

$$\sum_{i=1}^{n} P(\alpha_{i,r}|Y_s)(\alpha_{i,r} - \alpha_{i-1,r}) = 1 \tag{10}$$

for $(r, s) = (t + L, t), (t, t), (t, T)$ and $t = 1, \cdots, T$ in order to avoid accumulation of the errors.[3]

The NI procedure can evaluate each expectation quite precisely. When the state vector $\alpha_t$ is extended to multidimensional cases, however, we should keep in mind that a computational burden increases more than proportionally. Therefore the NI procedure is intractable in such a case.

## 2.2 MONTE-CARLO INTEGRATION APPROACH (MI)

Tanizaki (1993b), Tanizaki and Mariano (1994) and Mariano and Tanizaki (1995) developed a non-linear and non-normal filter with MI, where a recursive form of the density functions is converted to that of the weight functions. Define the weight function $\omega_{r|s} = P(\alpha_r|Y_s)/P_I(\alpha_r)$ for $(r, s) = (t + L, t), (t, t), (t, T)$, which is a ratio of two densities. The density function $P_I(\alpha_t)$, which is called the importance density, has to be specified by the researcher appropriately[4] . Moreover, define $\omega_{i,r|s} = P(\alpha_{i,r}|Y_s)/P_I(\alpha_{i,r})$ as the weight function evaluated at $\alpha_{i,r}$, where $\alpha_{i,r}$, $i = 1, \cdots, n$, are the random draws generated from the importance density $P_I(\alpha_r)$.

[3] Note that equation (10) comes from the following equation:

$$\int P(\alpha_r|Y_s)\mathrm{d}\alpha_r = 1.$$

[4] The importance density $P_I(\alpha_t)$ have to satisfy the following two conditions: (i) the importance density $P_I(\alpha_t)$ have to cover the two densities, i.e., $P(\alpha_t|Y_{t-1})$ and $P(\alpha_t|Y_t)$ and (ii) $P_I(\alpha_t)$ should be the density function with broader tails than $P(\alpha_t|Y_{t-1})$ and $P(\alpha_t|Y_t)$.

By MI, equation (3) is given by:

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

which is called MI prediction.[5]

The MI filter is given by the following two equations:

$$\omega_{i,t|t-1} = \frac{1}{n}\sum_{j=1}^{n} \frac{P_\alpha(\alpha_{i,t}|\alpha_{j,t-1})}{P_I(\alpha_{i,t})}\omega_{j,t-1|t-1}, \tag{12}$$

$$\omega_{i,t|t} = \frac{P_y(y_t|\alpha_{i,t})\omega_{i,t|t-1}}{\frac{1}{n}\sum_{j=1}^{n} P_y(y_t|\alpha_{j,t})\omega_{j,t|t-1}}, \tag{13}$$

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

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

$$\omega_{i,t|T} = \omega_{i,t|t}\frac{1}{n}\sum_{j=1}^{n} \frac{\omega_{j,t+1|T}P_\alpha(\alpha_{j,t+1}|\alpha_{i,t})}{\omega_{j,t+1|t}P_I(\alpha_{j,t+1})}, \tag{14}$$

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

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

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

The likelihood function (8) is rewritten as:

$$P(Y_T) = \prod_{t=1}^{T}\left(\frac{1}{n}\sum_{j=1}^{n} P_y(y_t|\alpha_{j,t})\omega_{j,t|t-1}\right). \tag{15}$$

---

When the range of the state-vector is not restricted, we may use the following mixture distribution in practice (see Sections 4.1, 4.2 and 4.4):

$$P_I(\alpha_t) = \frac{1}{2}N(a_{t|t-1}, c\Sigma_{t|t-1}) + \frac{1}{2}N(a_{t|t}, c\Sigma_{t|t}),$$

where $a_{t|t-1}$, $\Sigma_{t|t-1}$, $a_{t|t}$ and $\Sigma_{t|t}$ are the filtering and one-step ahead prediction estimates obtained from the extended Kalman filter, and $c$ is constant (we take $c = 4$ in Section 4). That is, we generate the random numbers for $\alpha_{i,t}$ from the normal density $N(a_{t|t-1}, c\Sigma_{t|t-1})$ with probability 1/2 and from the normal density $N(a_{t|t}, c\Sigma_{t|t})$ with probability 1/2. Since the importance density $P_I(\alpha_t)$ should cover $P(\alpha_t|Y_{t-1})$ and $P(\alpha_t|Y_t)$, the above mixture distribution is used. Moreover, $P_I(\alpha_t)$ should be the density function with broader tails than $P(\alpha_t|Y_{t-1})$ and $P(\alpha_t|Y_t)$ and accordingly we take $c = 4$ in Section 4. See, for example, Tanizaki (1993b), Tanizaki and Mariano (1994) and Mariano and Tanizaki (1995) for choice of the importance density.

However, when the state-vector lies on an interval, we may take a uniform distribution (see Section 4.3).

[5] For derivation of equation (11), using the weight functions $\omega_{t+L|t}$ and $\omega_{t+L-1|t}$, the prediction density (3) is rewritten as:

$$\omega_{t+L|t} = \int \frac{P_\alpha(\alpha_{t+L}|\alpha_{t+L-1})}{P_I(\alpha_{t+L})}\omega_{t+L-1|t}P_I(\alpha_{t+L-1})\mathrm{d}\alpha_{t+L-1}.$$

Generating $n$ random numbers of $\alpha_{t+L-1}$ from the importance density $P_I(\alpha_{t+L-1})$ and evaluating the weight function $\omega_{t+L|t}$ at the random numbers $\alpha_{i,t+L}$ generated from $P_I(\alpha_{t+L})$, we obtain equation (11).

In the exactly same fashion, equations (4) – (6) reduce to equations (12) – (14).

Since simulation errors drastically accumulate as $t$ increases, we need to impose as follows:

$$\frac{1}{n}\sum_{i=1}^{n}\omega_{i,r|s}=1, \tag{16}$$

for $(r,s)=(t+L,t),(t,t),(t,T)$ and $t=1,\cdots,T$.[6]

The MI procedure can be easily implemented in multidimensional cases of the state vector, which is an attractive feature over the NI approach. It is easily shown that $\widehat{g}_{r|s}$ is a consistent estimator but convergence is quite slow as $\sqrt{n}$.[7]

# 3 STATE-SPACE MODELING WITH MONTE-CARLO APPROACH (MA)

In this section, we propose a new non-linear and non-normal filter using a Monte-Carlo simulation technique. The method of Monte-Carlo simulations is utilized as an alternative solution to the non-linear and non-normal state-space model. Prediction, filtering and smoothing algorithms are introduced. For the NI and the MI approaches, choice of $\alpha_{i,t}$ is one of the critical problems, because estimation depends on the location of the nodes for NI and choice of the importance density for MI. The non-linear and non-normal procedure proposed here does not require such assumptions.

Let $A_s$ be a collection of the state-vectors up to time $s$, i.e., $A_s \equiv \{\alpha_s,\alpha_{s-1},\cdots,\alpha_0\}$. Note that the joint density function of $(Y_s,A_T)$ is written as $P(Y_s,A_T)=P(Y_s|A_s)P(A_T)$, where $P(Y_s|A_s)$ and $P(A_T)$ are given by:

$$P(Y_s|A_s)=\prod_{t=1}^{s}P_y(y_t|\alpha_t),$$

$$P(A_T)=P_\alpha(\alpha_0)\prod_{t=1}^{T}P_\alpha(\alpha_t|\alpha_{t-1}).$$

$P_y(y_t|\alpha_t)$ and $P_\alpha(\alpha_t|\alpha_{t-1})$ are obtained from the measurement and transition equations. Therefore, the conditional density of $\alpha_r$ given information $Y_s$ is given by:

$$P(\alpha_r|Y_s)=\frac{\int P(Y_s|A_s)P(A_T)\mathrm{d}A^*_{T,-r}}{\int P(Y_s|A_s)P(A_T)\mathrm{d}A_T},$$

---

[6] Equation (16) comes from the following integration of the density function:

$$\int P(\alpha_r|Y_s)\mathrm{d}\alpha_r=1.$$

See Koop (1994) for a survey of integration methods.

[7] As shown in Geweke (1989a, 1989b), MI gives us a consistent estimate if the sample size $n$ increases. Note that $\widehat{g}_{r|s}=(1/n)\sum_{i=1}^{n}g(\alpha_{i,r})\omega_{i,r|s}$, where $g(\alpha_{i,r})\omega_{i,r|s}$ is the random number generated from $P_I(\alpha_r)$. Define $g_{r|s}$ and $\Sigma_{r|s}$ as:

$$g_{r|s}\equiv E\big(g(\alpha_{i,r})\omega_{i,r|s}\big)=E\big(g(\alpha_r)|Y_s\big),$$

$$\Sigma_{r|s}\equiv E\big(g(\alpha_{i,r})g(\alpha_{i,r})'\omega^2_{i,r|s}\big)=E\big(g(\alpha_r)g(\alpha_r)'\omega_{r|s}|Y_s\big).$$

By the central limit theorem, it is shown that the following asymptotic normality holds:

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

where $A^*_{T,-r} = \{\alpha_T, \cdots, \alpha_{r+1}, \alpha_{r-1}, \cdots, \alpha_0\}$, which is the set that $\alpha_r$ is excluded from $A_t$.[8] The expectation of $g(\alpha_r)$ given $Y_s$, i.e., $g_{r|s} \equiv E\big(g(\alpha_r)|Y_s\big)$, is obtained as follows:

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

which is equivalent to equation (7).

Let us define $\overline{g}_{r|s}$ as the MA estimate of $g_{r|s}$ and $A_{i,T} = \{\alpha_{i,T}, \alpha_{i,T-1}, \cdots, \alpha_{i,0}\}$, $i = 1, \cdots, n$, as the random draws of $A_T$ generated from $P(A_T)$. That is, $A_{i,s}$ is a set of random draws from the transition equation $\alpha_{i,t} = f_t(\alpha_{i,t-1}, \eta_{i,t})$, where $\eta_{i,t}$ is a random number generated from the underlying density for $t = 1, \cdots, s$ and $i = 1, \cdots, n$.

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

$$\overline{g}_{r|s} = \frac{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} g(\alpha_{i,r})P(Y_s|A_{i,s})}{\dfrac{1}{n}\displaystyle\sum_{j=1}^{n} P(Y_s|A_{j,s})}. \tag{17}$$

In this paper, we call $\overline{g}_{r|s}$ the MA prediction when $(r, s) = (t + L, t)$, the MA filtering when $(r, s) = (t, t)$ and the MA smoothing when $(r, s) = (t, T)$, respectively.

The following formula is easy and convenient to compute the estimate of $g_{r|s}$ (i.e., $\overline{g}_{r|s}$). Since $P(Y_s|A_{i,s})$ is rewritten as $P(Y_s|A_{i,s}) = \prod P_y(y_t|\alpha_{i,t})$, equation (17) is represented as:

$$\overline{g}_{r|s} = \sum_{i=1}^{n} g(\alpha_{i,r})\omega_{i,s}, \tag{18}$$

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

$$\omega_{i,s} = \frac{\displaystyle\prod_{t=1}^{s} P_y(y_t|\alpha_{i,t})}{\displaystyle\sum_{j=1}^{n}\prod_{t=1}^{s} P_y(y_t|\alpha_{j,t})} = \frac{P_y(y_s|\alpha_{i,s})\omega_{i,s-1}}{\displaystyle\sum_{j=1}^{n} P_y(y_s|\alpha_{j,s})\omega_{j,s-1}}. \tag{19}$$

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

$$\omega_{i,0} = \frac{1}{n}, \tag{20}$$

for $i = 1, \cdots, n$. Thus, $\overline{g}_{r|s}$ denotes the MA estimate of $g(\alpha_r)$ given information $Y_s$,[9] where the Monte-Carlo simulation technique is utilized based on the random draws $A_{i,s} = \{\alpha_{i,s}, \alpha_{i,s-1}, \cdots, \alpha_{i,0}\}$ generated from the transition equation. Note that $\omega_{i,s}$ represents the weight function such that $\sum_{i=1}^{n} \omega_{i,s} = 1$ for all $s$.

[8] Note that $A^*_{t,-t} \equiv A_{t-1}$.

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

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

Therefore, the prediction estimate obtained by the Monte-Carlo simulation is given by:

$$\overline{g}_{t+L|t} = \sum_{i=1}^{n} g(\alpha_{i,t+L})\omega_{i,t},$$

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

The features of the above prediction, filtering and smoothing are that the random numbers of $\alpha_t$ (i.e., $\alpha_{i,t}$, $i = 1, \cdots, n$) are generated from the transition equation (2) for all $t$, and that the algorithm requires the functional form of the density function of $y_t$ given $\alpha_t$ (i.e., $P_y(y_t|\alpha_t)$) for all $t$. We do not need *ad hoc* assumptions such as choice of the nodes for NI and that of the importance density for MI. Note that the MA procedure does not yield the recursive algorithm, which is different from NI and MI approaches and similar to the Gibbs sampler approach.

The computation procedure of the new proposal is as follows: (i) the random draws of the initial state variable $\alpha_0$ are generated from the initial density $P_\alpha(\alpha_0)$, which are denoted by $\alpha_{i,0}$, $i = 1, \cdots, n$, (ii) given $\alpha_{i,0}$, the random numbers $\alpha_{i,t}$ are obtained from the transition equation $\alpha_{i,t} = f_t(\alpha_{i,t-1}, \eta_{i,t})$ using the random draws $\eta_{i,t}$ for $i = 1, \cdots, n$ and $t = 1, \cdots, T$, (iii) given the initial weight (20), use equation (19) to obtain the weight functions for $i = 1, \cdots, n$ and $t = 1, \cdots, T$, and (iv) $\overline{g}_{r|s}$ can be computed by equation (18).

Thus, the transition equation (2) is utilized in order to generate the random numbers of $A_s$ for $s = t, T$. The density function $P_y(y_t|\alpha_t)$ comes from the measurement equation (1). The proposed estimator needs to evaluate the density $P_y(y_t|\alpha_t)$ at $\alpha_{i,t}$, while for both the NI and the MI procedures we have to compute $P_\alpha(\alpha_t|\alpha_{t-1})$ and $P_y(y_t|\alpha_t)$ by hand, which implies much easier programming of the new proposal than NI and MI approaches. Moreover, the proposed procedure is more attractive with respect to computational time than the other procedures.

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

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

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

$$P(Y_T) = \frac{1}{n}\sum_{i=1}^{n} P(Y_T|A_{i,T}), \tag{21}$$

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

Under some conditions, the theorems proved by Geweke (1989a, 1989b), which are related to the asymptotic behavior of the Bayes mean using MI, hold without any modification in the case of MA proposed above. That is, $\Sigma_{r|s}$ is defined as follows:

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

The sample variance of $\overline{\alpha}_{r|s}$ (i.e., $\overline{\Sigma}_{r|s}$) is defined as:

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

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

(i) $\overline{g}_{r|s} \xrightarrow{\text{a.s.}} g_{r|s}$,

(ii) $\sqrt{n}(\overline{g}_{r|s} - g_{r|s}) \xrightarrow{\text{d}} N(0, \Sigma_{r|s})$,
$n\overline{\Sigma}_{r|s} \xrightarrow{\text{a.s.}} \Sigma_{r|s}$,

where " $\xrightarrow{\text{a.s.}}$ " and " $\xrightarrow{\text{d}}$ " denote almost sure convergence and convergence in distribution, respectively. The above asymptotic properties indicate that $\overline{g}_{r|s}$ is consistent but convergence is slow as $\sqrt{n}$.

Finally, note that, from equation (19), it might be expected that the simulation errors increase as $t$ is large. Accordingly, when $T$ is large, the MA procedure is intractable.

# 4 MONTE-CARLO EXPERIMENTS

In this section, we compare the estimators NI, MI and MA defined in the previous sections 2.1, 2.2 and 3. Sections 4.1 – 4.5. give the results from a Monte-Carlo study.

## 4.1 LINEAR AND NORMAL MODEL

Consider the following linear scalar system:

$$\text{Measurement Equation:} \quad y_t = \alpha_t + \epsilon_t,$$

$$\text{Transition Equation:} \quad \alpha_t = \delta\alpha_{t-1} + \eta_t,$$

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

$$\alpha_0 \sim N(0,1),$$

where $t = 1, \cdots, T$. We compare the following estimates: the Kalman (K) estimate $a_{r|s}$[10] , the NI estimate $\widetilde{\alpha}_{r|s}$ (Section 2.1)[11] , the MI estimate $\widehat{\alpha}_{r|s}$ (Section 2.2)[12] , and the MA estimate $\overline{\alpha}_{r|s}$ (Section 3). Note that the K estimate $a_{r|s}$ gives us the optimal solution in this example because the system in this section is a linear and normal model.

The simulation procedure is as follows:

(i) Generating standard normal random numbers of $\epsilon_t$ and $\eta_t$ for $t = 1, \cdots, T$, we obtain a set of data $y_t$ and $\alpha_t$, $t = 1, \cdots, T$, from the system, where $T = 10, 20, 30$.

(ii) We choose $n = L$ for NI and MI, and $n = L^2$ for MA in order to take an equal computational burden for all the procedures, where $L = 20, 50, 100$.

(iii) Given $Y_s$, obtain $a_{r|s}$, $\widetilde{\alpha}_{r|s}$, $\widehat{\alpha}_{r|s}$ and $\overline{\alpha}_{r|s}$ for $t = 1, \cdots, T$ and $(r, s) = (t, t-1)$, $(t, t)$, $(t, T)$.

(iv) Repeat (i)–(iii) $G$ times and compute the root mean square error (RMSE) for each estimate, which are defined as:

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

[10] We define $a_{r|s}$ as the K prediction estimate when $(r, s) = (t, t-1)$, the K filtering estimate (Kalman filter estimate) when $(r, s) = (t, t)$ and the K smoothing estimate (Kalman smoothed estimate) when $(r, s) = (t, T)$. In the case of linear and normal cases, $a_{r|s}$ is given by the conventional linear recursive algorithms, which are derived from equations (3) – (6). In the case of non-linear state-space model (Sections 4.2 – 4.5), K implies the extended Kalman filter in this paper, which is the non-linear filter based on the first-order Taylor series expansion.

[11] For NI, half of the nodes are chosen from the interval $[a_{t|t-1} - 4\Sigma_{t|t-1}^{1/2}, a_{t|t-1} + 4\Sigma_{t|t-1}^{1/2}]$ and the rest of the nodes are from $[a_{t|t} - 4\Sigma_{t|t}^{1/2}, a_{t|t} + 4\Sigma_{t|t}^{1/2}]$, where $a_{t|t}$ and $\Sigma_{t|t}$ denote the K filtering estimates of $\alpha_t$ and its variance and $a_{t|t-1}$ and $\Sigma_{t|t-1}$ denote the one-step ahead K prediction estimates of $\alpha_t$ and its variance.

[12] For MI, we generate the random numbers for $\alpha_{i,t}$ from the normal density $N(a_{t|t-1}, 4\Sigma_{t|t-1})$ with probability 1/2 and from the normal density $N(a_{t|t}, 4\Sigma_{t|t})$ with probability 1/2, which corresponds to the importance density represented by $P_I(\alpha_t) = \frac{1}{2}N(a_{t|t-1}, 4\Sigma_{t|t-1}) + \frac{1}{2}N(a_{t|t}, 4\Sigma_{t|t})$. For the importance density $P_I(\alpha_t)$, note that we need to choose the density function with broader tails than $P(\alpha_t|Y_{t-1})$ and $P(\alpha_t|Y_t)$.

Table 1: Linear and Normal Model ($\delta$ known)

| | $L$ | $T=10$ P | $T=10$ F | $T=10$ S | $T=20$ P | $T=20$ F | $T=20$ S | $T=30$ P | $T=30$ F | $T=30$ S |
|---|---|---|---|---|---|---|---|---|---|---|
| $\delta=0.5$ | | | | | | | | | | |
| K | | 1.0721 | 0.7266 | 0.7048 | 1.0628 | 0.7334 | 0.7121 | 1.0603 | 0.7304 | 0.7088 |
| MA | 20 | 1.1012 | 0.7521 | 0.7662 | 1.1555 | 0.8201 | 0.9133 | 1.2163 | 0.8818 | 1.0150 |
| | 50 | 1.0793 | 0.7320 | 0.7244 | 1.1144 | 0.7764 | 0.8285 | 1.1587 | 0.8170 | 0.9273 |
| | 100 | 1.0750 | 0.7292 | 0.7120 | 1.0887 | 0.7553 | 0.7824 | 1.1252 | 0.7844 | 0.8764 |
| NI | 20 | 1.0745 | 0.7268 | 0.7049 | 1.0652 | 0.7335 | 0.7122 | 1.0628 | 0.7305 | 0.7089 |
| | 50 | 1.0725 | 0.7266 | 0.7048 | 1.0632 | 0.7334 | 0.7121 | 1.0607 | 0.7304 | 0.7088 |
| | 100 | 1.0722 | 0.7266 | 0.7048 | 1.0629 | 0.7334 | 0.7121 | 1.0604 | 0.7304 | 0.7088 |
| MI | 20 | 1.0963 | 0.7454 | 0.7243 | 1.0869 | 0.7547 | 0.7336 | 1.0815 | 0.7521 | 0.7312 |
| | 50 | 1.0825 | 0.7360 | 0.7141 | 1.0694 | 0.7411 | 0.7204 | 1.0683 | 0.7384 | 0.7176 |
| | 100 | 1.0776 | 0.7294 | 0.7076 | 1.0659 | 0.7370 | 0.7160 | 1.0652 | 0.7350 | 0.7134 |
| $\delta=1.0$ | | | | | | | | | | |
| K | | 1.2826 | 0.7815 | 0.6824 | 1.2766 | 0.7938 | 0.6841 | 1.2716 | 0.7904 | 0.6795 |
| MA | 20 | 1.3384 | 0.8326 | 0.8082 | 1.5003 | 0.9983 | 1.1030 | 1.6599 | 1.1620 | 1.3378 |
| | 50 | 1.2977 | 0.7937 | 0.7200 | 1.3875 | 0.8879 | 0.9252 | 1.5152 | 1.0060 | 1.1368 |
| | 100 | 1.2890 | 0.7862 | 0.6987 | 1.3381 | 0.8444 | 0.8357 | 1.4386 | 0.9354 | 1.0302 |
| NI | 20 | 1.2845 | 0.7815 | 0.6825 | 1.2787 | 0.7939 | 0.6842 | 1.2736 | 0.7903 | 0.6795 |
| | 50 | 1.2829 | 0.7815 | 0.6824 | 1.2769 | 0.7938 | 0.6841 | 1.2719 | 0.7904 | 0.6795 |
| | 100 | 1.2827 | 0.7815 | 0.6824 | 1.2766 | 0.7938 | 0.6841 | 1.2716 | 0.7904 | 0.6795 |
| MI | 20 | 1.3209 | 0.8066 | 0.7091 | 1.3151 | 0.8217 | 0.7123 | 1.3070 | 0.8189 | 0.7104 |
| | 50 | 1.3005 | 0.7939 | 0.6929 | 1.2884 | 0.8037 | 0.6955 | 1.2846 | 0.8014 | 0.6918 |
| | 100 | 1.2911 | 0.7853 | 0.6857 | 1.2822 | 0.7987 | 0.6896 | 1.2800 | 0.7961 | 0.6848 |

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

The results are in Table 1. P, F and S denote one-step ahead prediction, filtering and smoothing. K indicates the best estimator of the three in this simulation study because the system is linear and normal. NI is very close to K. MI approaches K and NI as $L$ increases. From RMSE, MA is a poor estimator as $T$ increases, which implies that the MA estimates diverge from the true value as $T$ increases. As $L$ increases, precision of the MA estimates becomes better. However, convergence is very slow, especially when $T$ is large. Also, as $\delta$ approaches one, MA shows a poor performance because of a unit-root problem. Thus, convergence speed of MA is slow when $\delta$ is one and comparably fast when $\delta$ goes to zero but still slow, compared with that of NI and MI. One reason why MA is poor (or convergence speed is too slow) in this Monte-Carlo experiment is because the range of $\alpha$ lies on the interval from $-\infty$ to $+\infty$. For MA, numerous random numbers have to be generated for improvement of precision in such a case.

Table 2 shows the estimation results of unknown parameter $\delta$ (note that $\delta$ is fixed in Table 1). We obtain Table 2 as follows. (i) given $\delta$, generate $y_t$ for $t = 1, , \cdots, T$ from the measurement and transition equations, (ii) given the data $\{y_t\}$, estimate $\delta$ using the appropriate likelihood function, i.e., (9), (15), or (21), (iii) repeat (i) and (ii) 1000 times, i.e., perform 1000 simulation runs, and (iv) from 1000 estimates of $\delta$, compute arithmetic average (AVE), root mean square error (RMSE), and 25, 50 and 75 percent quantiles (25%, 50%, and 75%).

Table 2: Linear and Normal Model ($\delta$ unknown)

| | | $T=10$ | | | | | $T=20$ | | | | | $T=30$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L$ | AVE | RMSE | 25% | 50% | 75% | AVE | RMSE | 25% | 50% | 75% | AVE | RMSE | 25% | 50% | 75% |
| $\delta=0.5$ | | | | | | | | | | | | | | | | |
| K | | 0.318 | 0.439 | 0.050 | 0.370 | 0.640 | 0.382 | 0.330 | 0.180 | 0.450 | 0.620 | 0.411 | 0.273 | 0.270 | 0.470 | 0.600 |
| MA | 20 | 0.333 | 0.443 | 0.040 | 0.405 | 0.660 | 0.412 | 0.332 | 0.220 | 0.480 | 0.650 | 0.419 | 0.319 | 0.260 | 0.460 | 0.650 |
| | 50 | 0.324 | 0.443 | 0.060 | 0.400 | 0.640 | 0.405 | 0.342 | 0.220 | 0.480 | 0.650 | 0.414 | 0.296 | 0.270 | 0.465 | 0.620 |
| | 100 | 0.322 | 0.441 | 0.050 | 0.370 | 0.640 | 0.383 | 0.339 | 0.190 | 0.450 | 0.620 | 0.426 | 0.291 | 0.285 | 0.480 | 0.630 |
| NI | 20 | 0.320 | 0.438 | 0.055 | 0.370 | 0.640 | 0.384 | 0.329 | 0.180 | 0.450 | 0.620 | 0.412 | 0.273 | 0.270 | 0.465 | 0.600 |
| | 50 | 0.319 | 0.439 | 0.050 | 0.370 | 0.640 | 0.382 | 0.330 | 0.180 | 0.450 | 0.620 | 0.412 | 0.273 | 0.270 | 0.470 | 0.600 |
| | 100 | 0.319 | 0.439 | 0.050 | 0.370 | 0.640 | 0.382 | 0.330 | 0.180 | 0.450 | 0.620 | 0.412 | 0.273 | 0.270 | 0.470 | 0.600 |
| MI | 20 | 0.308 | 0.454 | 0.015 | 0.370 | 0.645 | 0.369 | 0.340 | 0.160 | 0.430 | 0.615 | 0.392 | 0.290 | 0.240 | 0.440 | 0.600 |
| | 50 | 0.317 | 0.436 | 0.050 | 0.380 | 0.630 | 0.377 | 0.334 | 0.170 | 0.440 | 0.620 | 0.408 | 0.277 | 0.260 | 0.460 | 0.605 |
| | 100 | 0.313 | 0.442 | 0.045 | 0.370 | 0.640 | 0.380 | 0.331 | 0.175 | 0.450 | 0.620 | 0.407 | 0.278 | 0.260 | 0.460 | 0.600 |
| $\delta=1.0$ | | | | | | | | | | | | | | | | |
| K | | 0.786 | 0.427 | 0.650 | 0.900 | 1.040 | 0.896 | 0.229 | 0.860 | 0.960 | 1.010 | 0.934 | 0.149 | 0.910 | 0.970 | 1.010 |
| MA | 20 | 0.884 | 0.333 | 0.885 | 0.990 | 1.040 | 0.980 | 0.126 | 0.980 | 1.000 | 1.020 | 0.995 | 0.062 | 0.990 | 1.000 | 1.020 |
| | 50 | 0.824 | 0.393 | 0.750 | 0.970 | 1.030 | 0.976 | 0.140 | 0.980 | 1.000 | 1.020 | 0.994 | 0.067 | 0.990 | 1.000 | 1.020 |
| | 100 | 0.796 | 0.420 | 0.670 | 0.960 | 1.030 | 0.967 | 0.160 | 0.980 | 1.000 | 1.020 | 0.997 | 0.055 | 0.990 | 1.000 | 1.010 |
| NI | 20 | 0.786 | 0.426 | 0.655 | 0.905 | 1.040 | 0.896 | 0.229 | 0.860 | 0.960 | 1.010 | 0.934 | 0.148 | 0.910 | 0.970 | 1.010 |
| | 50 | 0.786 | 0.426 | 0.650 | 0.900 | 1.040 | 0.896 | 0.228 | 0.860 | 0.960 | 1.010 | 0.934 | 0.149 | 0.910 | 0.970 | 1.010 |
| | 100 | 0.786 | 0.427 | 0.650 | 0.900 | 1.040 | 0.896 | 0.229 | 0.860 | 0.960 | 1.010 | 0.934 | 0.149 | 0.910 | 0.970 | 1.010 |
| MI | 20 | 0.777 | 0.446 | 0.640 | 0.900 | 1.040 | 0.894 | 0.230 | 0.850 | 0.955 | 1.020 | 0.930 | 0.157 | 0.900 | 0.970 | 1.010 |
| | 50 | 0.779 | 0.432 | 0.650 | 0.900 | 1.030 | 0.895 | 0.232 | 0.860 | 0.960 | 1.020 | 0.932 | 0.152 | 0.900 | 0.970 | 1.010 |
| | 100 | 0.783 | 0.432 | 0.650 | 0.900 | 1.030 | 0.896 | 0.229 | 0.860 | 0.960 | 1.010 | 0.933 | 0.151 | 0.905 | 0.970 | 1.010 |

The reason why MA is the worst estimator is as follows. MA is the estimator which approximates the filtering densities by generating the unconditional random draws (rather than the conditional random draws) of the state-variable $\alpha_1$ from the transition equation. In the case where the transition equation is a random walk, $\alpha_t$ does not have mean and variance. In such a situation, we do not know where the unconditional random draws from the transition equation approach as time $t$ goes to infinity. As a result, we obtained the result that MA is the worst estimator in this example.

## 4.2 NON-LINEAR MODEL I

In this section, a nonlinear function of the first-order autoregressive conditional heteroscedasticity ( i.e., ARCH(1) ) process[13] is taken. Consider the following non-linear scalar system:

$$\text{Measurement Equation:} \quad y_t = \alpha_t + \epsilon_t,$$

$$\text{Transition Equation:} \quad \alpha_t = (1-\delta+\delta\alpha_{t-1}^2)^{1/2}\eta_t,$$

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

$$\alpha_0 \sim N(0,1),$$

[13] For ARCH model, see Engle (1982).

Table 3: Non-linear Model I ($\delta$ known)

| | | $T=10$ | | | $T=20$ | | | $T=30$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $L$ | P | F | S | P | F | S | P | F | S |
| $\delta=0.5$ | | | | | | | | | | |
| K | | 1.0424 | 0.7042 | 0.7042 | 1.0215 | 0.7036 | 0.7036 | 0.9786 | 0.6989 | 0.6989 |
| MA | 20 | 1.0648 | 0.7760 | 0.8075 | 1.0841 | 0.8033 | 0.8792 | 1.0846 | 0.8022 | 0.9096 |
| | 50 | 1.0769 | 0.7446 | 0.7754 | 1.0726 | 0.7535 | 0.8185 | 1.0519 | 0.7550 | 0.8405 |
| | 100 | 1.0646 | 0.7232 | 0.7528 | 1.0458 | 0.7479 | 0.7842 | 1.0428 | 0.7346 | 0.8033 |
| NI | 20 | 1.0406 | 0.6859 | 0.6787 | 1.0199 | 0.6888 | 0.6801 | 0.9776 | 0.6858 | 0.6777 |
| | 50 | 1.0412 | 0.6855 | 0.6783 | 1.0204 | 0.6885 | 0.6798 | 0.9777 | 0.6857 | 0.6775 |
| | 100 | 1.0414 | 0.6855 | 0.6782 | 1.0205 | 0.6885 | 0.6798 | 0.9777 | 0.6857 | 0.6775 |
| MI | 20 | 1.0658 | 0.7008 | 0.6942 | 1.0362 | 0.7072 | 0.6993 | 0.9934 | 0.7046 | 0.6972 |
| | 50 | 1.0492 | 0.6927 | 0.6853 | 1.0263 | 0.6956 | 0.6871 | 0.9851 | 0.6926 | 0.6854 |
| | 100 | 1.0464 | 0.6879 | 0.6808 | 1.0217 | 0.6915 | 0.6828 | 0.9821 | 0.6896 | 0.6816 |
| $\delta=0.9$ | | | | | | | | | | |
| K | | 1.1334 | 0.7028 | 0.7028 | 1.0848 | 0.6632 | 0.6632 | 0.8041 | 0.6345 | 0.6345 |
| MA | 20 | 1.1719 | 0.9276 | 0.9675 | 1.1683 | 0.8897 | 0.9546 | 0.8770 | 0.6671 | 0.7218 |
| | 50 | 1.2048 | 0.8590 | 0.9159 | 1.1776 | 0.7446 | 0.9117 | 0.8800 | 0.6196 | 0.6714 |
| | 100 | 1.2777 | 0.7405 | 0.8499 | 1.1985 | 0.8196 | 0.8629 | 0.8782 | 0.6076 | 0.6357 |
| NI | 20 | 1.1224 | 0.6592 | 0.6465 | 1.0753 | 0.6001 | 0.5894 | 0.7983 | 0.5833 | 0.5736 |
| | 50 | 1.1269 | 0.6540 | 0.6412 | 1.0794 | 0.5957 | 0.5843 | 0.8000 | 0.5793 | 0.5689 |
| | 100 | 1.1275 | 0.6523 | 0.6394 | 1.0799 | 0.5942 | 0.5827 | 0.8002 | 0.5780 | 0.5674 |
| MI | 20 | 1.1822 | 0.6281 | 0.6166 | 1.0795 | 0.6089 | 0.5966 | 0.8302 | 0.5894 | 0.5760 |
| | 50 | 1.1758 | 0.6160 | 0.6009 | 1.1079 | 0.5904 | 0.5752 | 0.8149 | 0.5716 | 0.5572 |
| | 100 | 1.1636 | 0.6002 | 0.5859 | 1.0895 | 0.5809 | 0.5644 | 0.8127 | 0.5649 | 0.5485 |

where $0 \leq \delta < 1$ and $t = 1, \cdots, T$. We take $\delta = 0.5, 0.9$. The transition equation follows the ARCH(1) process[14] , while the measurement equation consists of the ARCH(1) term and the error. The measurement equation is linear but the transition equation is non-linear. In this experiment, therefore, it might be expected that K does not show a good performance. The nodes for NI and the importance density for MI are in the same way as Section 4.1.

In Table 3, we compare precision of the estimated state variables for K, NI, MI and MA. NI performs better than any other estimators. Especially, as it is expected, NI and MI are better than K. It is natural that K is not good, because for K the transition equation is linearized with respect to $\alpha_{t-1}$ and $\eta_t$. MA is the worst estimator from RMSE. When $T$ and $L$ are large, MA becomes better.

Table 4 shows the estimation results of unknown parameter $\delta$. There, the estimates of $\delta$ are compared for K, NI, MI and MA. Each likelihood function is maximized by a simple grid search, where we take $\delta = 0.01, 0.02, \cdots, 0.99$. It might be concluded that MA is the best estimator because AVE is close to the true parameter value and RMSE is the smallest.

[14] In this Monte-Carlo experiment, the unconditional variance of $\alpha_t$ is assumed to be one.

Table 4: Non-linear Model I ($\delta$ unknown)

| | | $T=10$ | | | | | $T=20$ | | | | | $T=30$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $L$ | AVE | RMSE | 25% | 50% | 75% | AVE | RMSE | 25% | 50% | 75% | AVE | RMSE | 25% | 50% | 75% |
| $\delta=0.5$ | | | | | | | | | | | | | | | | |
| K | | 0.428 | 0.403 | 0.010 | 0.360 | 0.860 | 0.379 | 0.368 | 0.010 | 0.320 | 0.680 | 0.346 | 0.348 | 0.010 | 0.290 | 0.590 |
| MA | 20 | 0.504 | 0.418 | 0.010 | 0.525 | 0.990 | 0.612 | 0.386 | 0.260 | 0.760 | 0.940 | 0.698 | 0.380 | 0.550 | 0.840 | 0.940 |
| | 50 | 0.507 | 0.420 | 0.010 | 0.530 | 0.990 | 0.564 | 0.375 | 0.155 | 0.670 | 0.910 | 0.612 | 0.359 | 0.350 | 0.720 | 0.905 |
| | 100 | 0.485 | 0.428 | 0.010 | 0.435 | 0.990 | 0.516 | 0.385 | 0.050 | 0.570 | 0.910 | 0.582 | 0.356 | 0.290 | 0.685 | 0.890 |
| NI | 20 | 0.328 | 0.418 | 0.010 | 0.050 | 0.675 | 0.368 | 0.378 | 0.010 | 0.290 | 0.710 | 0.420 | 0.354 | 0.010 | 0.490 | 0.740 |
| | 50 | 0.344 | 0.420 | 0.010 | 0.090 | 0.730 | 0.384 | 0.382 | 0.010 | 0.320 | 0.750 | 0.416 | 0.362 | 0.010 | 0.460 | 0.750 |
| | 100 | 0.357 | 0.420 | 0.010 | 0.110 | 0.760 | 0.389 | 0.384 | 0.010 | 0.320 | 0.760 | 0.416 | 0.364 | 0.010 | 0.455 | 0.760 |
| MI | 20 | 0.458 | 0.425 | 0.010 | 0.390 | 0.960 | 0.485 | 0.403 | 0.010 | 0.515 | 0.920 | 0.463 | 0.387 | 0.010 | 0.460 | 0.855 |
| | 50 | 0.416 | 0.429 | 0.010 | 0.240 | 0.930 | 0.456 | 0.400 | 0.010 | 0.420 | 0.870 | 0.452 | 0.386 | 0.010 | 0.420 | 0.850 |
| | 100 | 0.429 | 0.426 | 0.010 | 0.265 | 0.945 | 0.458 | 0.404 | 0.010 | 0.420 | 0.895 | 0.448 | 0.384 | 0.010 | 0.450 | 0.840 |
| $\delta=0.9$ | | | | | | | | | | | | | | | | |
| K | | 0.618 | 0.478 | 0.260 | 0.775 | 0.990 | 0.629 | 0.445 | 0.350 | 0.740 | 0.970 | 0.638 | 0.409 | 0.430 | 0.710 | 0.920 |
| MA | 20 | 0.755 | 0.381 | 0.610 | 0.970 | 0.990 | 0.844 | 0.257 | 0.830 | 0.960 | 0.990 | 0.890 | 0.183 | 0.880 | 0.960 | 0.990 |
| | 50 | 0.713 | 0.428 | 0.475 | 0.960 | 0.990 | 0.824 | 0.286 | 0.800 | 0.960 | 0.990 | 0.873 | 0.200 | 0.850 | 0.950 | 0.990 |
| | 100 | 0.700 | 0.444 | 0.355 | 0.965 | 0.990 | 0.792 | 0.330 | 0.750 | 0.950 | 0.990 | 0.852 | 0.232 | 0.830 | 0.950 | 0.990 |
| NI | 20 | 0.468 | 0.612 | 0.010 | 0.510 | 0.990 | 0.471 | 0.595 | 0.010 | 0.655 | 0.860 | 0.527 | 0.537 | 0.010 | 0.720 | 0.850 |
| | 50 | 0.498 | 0.591 | 0.010 | 0.605 | 0.990 | 0.512 | 0.568 | 0.010 | 0.710 | 0.890 | 0.561 | 0.516 | 0.010 | 0.750 | 0.880 |
| | 100 | 0.513 | 0.581 | 0.010 | 0.650 | 0.990 | 0.527 | 0.558 | 0.010 | 0.725 | 0.900 | 0.571 | 0.510 | 0.010 | 0.760 | 0.890 |
| MI | 20 | 0.628 | 0.504 | 0.010 | 0.890 | 0.990 | 0.718 | 0.413 | 0.540 | 0.920 | 0.990 | 0.743 | 0.377 | 0.680 | 0.910 | 0.990 |
| | 50 | 0.599 | 0.526 | 0.010 | 0.850 | 0.990 | 0.679 | 0.452 | 0.335 | 0.900 | 0.990 | 0.730 | 0.386 | 0.645 | 0.900 | 0.990 |
| | 100 | 0.611 | 0.515 | 0.010 | 0.860 | 0.990 | 0.689 | 0.442 | 0.400 | 0.910 | 0.990 | 0.730 | 0.385 | 0.655 | 0.890 | 0.990 |

## 4.3 NON-LINEAR MODEL II

Consider the following non-linear scalar system:

$$\begin{aligned}
&\text{Measurement Equation:} \quad y_t = \frac{\exp(\alpha_t)}{\exp(\alpha_t)+\exp(\epsilon_t)}, \\
&\text{Transition Equation:} \quad \alpha_t = \frac{\exp(\alpha_{t-1})}{\exp(\alpha_{t-1})+\exp(\eta_t)}, \\
&\qquad\qquad \begin{pmatrix}\epsilon_t \\ \eta_t\end{pmatrix} \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\right), \\
&\qquad\qquad \alpha_0 \sim N(0,1),
\end{aligned} \tag{22}$$

where $t = 1, \cdots, T$. In this experiment, both the measurement and the transition equations are non-linear[15] . For K, equation (22) is approximated by the first-order Taylor series expansion. For NI, the nodes $\alpha_{i,t}$,

[15] 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 K, however, equation (22), rather than the above equation, is linearized in this experiment.

Table 5: Non-linear Model II

| | | $T=10$ | | | $T=20$ | | | $T=30$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $L$ | P | F | S | P | F | S | P | F | S |
| K | | 0.2042 | 0.2138 | 0.2149 | 0.2030 | 0.2146 | 0.2168 | 0.2021 | 0.2134 | 0.2155 |
| | 20 | 0.2024 | 0.1984 | 0.1982 | 0.2010 | 0.1974 | 0.1974 | 0.2000 | 0.1963 | 0.1963 |
| MA | 50 | 0.2021 | 0.1981 | 0.1978 | 0.2008 | 0.1971 | 0.1971 | 0.1997 | 0.1959 | 0.1959 |
| | 100 | 0.2020 | 0.1980 | 0.1978 | 0.2008 | 0.1971 | 0.1970 | 0.1997 | 0.1959 | 0.1958 |
| | 20 | 0.2022 | 0.1982 | 0.1979 | 0.2007 | 0.1971 | 0.1970 | 0.1997 | 0.1959 | 0.1958 |
| NI | 50 | 0.2022 | 0.1982 | 0.1979 | 0.2007 | 0.1971 | 0.1970 | 0.1997 | 0.1959 | 0.1958 |
| | 100 | 0.2022 | 0.1982 | 0.1979 | 0.2007 | 0.1971 | 0.1970 | 0.1997 | 0.1959 | 0.1958 |
| | 20 | 0.2078 | 0.2037 | 0.2036 | 0.2060 | 0.2023 | 0.2024 | 0.2054 | 0.2015 | 0.2016 |
| MI | 50 | 0.2040 | 0.2000 | 0.1998 | 0.2025 | 0.1988 | 0.1987 | 0.2018 | 0.1979 | 0.1979 |
| | 100 | 0.2034 | 0.1994 | 0.1992 | 0.2017 | 0.1981 | 0.1979 | 0.2008 | 0.1970 | 0.1969 |

$i = 1, \cdots, L$, are set to be $\alpha_{i,t} = (i - 0.5)/L$. For MI, the random draws $\alpha_{i,t}$ are generated from a uniform distribution between zero and one.

The results are in Table 5. K is the worst estimator. Thus, K is clearly biased. NI and MA are better than K and MI. MA is a good estimator from RMSE.

## 4.4 NON-LINEAR MODEL III

Consider another example of the following univariate nonstationary growth model[16] :

$$
\begin{aligned}
&\text{Measurement Equation:} \quad y_t = \frac{\alpha_t^2}{20} + \epsilon_t, \\
&\text{Transition Equation:} \quad \alpha_t = \frac{1}{2}\alpha_{t-1} + \frac{25\alpha_{t-1}}{1+\alpha_{t-1}^2} + 8\cos(1.2(t-1)) + \eta_t, \\
&\qquad\qquad \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 10 \end{pmatrix} \right), \\
&\qquad\qquad \alpha_0 \sim N(0,1),
\end{aligned}
\tag{23}
$$

where $t = 1, \cdots, T$. We take almost the same as in Sections 4.1 and 4.2.

The results are in Table 6. MA is tremendously good from RMSE criterion. The system (23) is nonstationary. In almost the cases, the range of $\alpha$ lies on the interval between $-30$ and 30, judging from the values of the simulated $\alpha_t$ based on the transition equation in the system (23) (see Kitagawa (1987)). Accordingly, since the range of the state variable is restricted as in the previous section, we can consider that MA performs better.

Table 6: Non-linear Model III

| | | $T=10$ | | | $T=20$ | | | $T=30$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $L$ | P | F | S | P | F | S | P | F | S |
| K | | 13.89 | 17.50 | 19.12 | 14.68 | 18.49 | 19.09 | 15.09 | 19.47 | 19.66 |
| | 20 | 8.96 | 5.83 | 4.54 | 9.72 | 6.63 | 5.82 | 10.09 | 7.28 | 6.72 |
| MA | 50 | 8.46 | 5.29 | 3.90 | 9.33 | 6.06 | 4.95 | 9.71 | 6.62 | 5.81 |
| | 100 | 8.10 | 5.15 | 3.51 | 9.07 | 5.79 | 4.53 | 9.45 | 6.23 | 5.24 |
| | 20 | 12.49 | 14.48 | 14.16 | 13.36 | 15.28 | 15.36 | 13.78 | 15.67 | 15.88 |
| NI | 50 | 12.13 | 13.32 | 13.85 | 13.12 | 14.60 | 14.98 | 13.53 | 15.15 | 15.56 |
| | 100 | 12.03 | 13.09 | 13.66 | 13.01 | 14.45 | 14.84 | 13.47 | 15.05 | 15.46 |
| | 20 | 12.87 | 14.69 | 14.76 | 13.51 | 15.28 | 15.68 | 13.82 | 15.58 | 16.10 |
| MI | 50 | 11.97 | 13.36 | 13.60 | 12.73 | 14.39 | 14.55 | 13.17 | 14.80 | 15.15 |
| | 100 | 11.27 | 12.27 | 12.57 | 12.18 | 13.48 | 13.58 | 12.52 | 13.95 | 14.17 |

### 4.5 NON-NORMAL MODEL

Consider the following non-normal scalar system:

$$\text{Measurement Equation:} \quad y_t = \alpha_t + \epsilon_t,$$

$$\text{Transition Equation:} \quad \alpha_t = \delta\alpha_{t-1} + \eta_t,$$

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

$$\alpha_0 \sim TN(0,1),$$

where $t = 1, \cdots, T$. $TN$ denotes the truncated normal distribution from both sides, i.e, $|\epsilon_t + \alpha_t| < 1.0$, $|\eta_t + \delta\alpha_{t-1}| < 1.0$ and $|\alpha_0| < 1.0$. In this experiment, both the measurement and transition equations are non-normal. The simulation procedure is the almost same as Sections 4.1, 4.2 and 4.4, except for the assumption that $\alpha_{i,t}$ is less than one in absolute value for $i = 1, \cdots, n$ and $t = 0, 1, \cdots, T$. For NI, we take the nodes $\alpha_{i,t}$, $i = 1, \cdots, L$, such that $\alpha_{i,t} - \alpha_{i-1,t}$ has an equal interval for all $i$. For MI, we take the random draws $\alpha_{i,t}$, $i = 1, \cdots, L$, from the truncated bivariate normal density, i.e.,

$$P_I(\alpha_t) = \frac{1}{2}TN(a_{t|t-1}, c\Sigma_{t|t-1}) + \frac{1}{2}TN(a_{t|t}, c\Sigma_{t|t}),$$

where $a_{t|t-1}$, $\Sigma_{t|t-1}$, $a_{t|t}$ and $\Sigma_{t|t}$ are given by the extended Kalman filter. As in the previous simulation studies, $c = 4$ is taken.

The results are in Table 7. $\delta = 0.5,\ 1.0$ is examined. MA is very close to NI, while K indicates the worst estimator. In this Monte-Carlo study, the range of $\alpha$ lies on the interval from $-1$ to 1. Therefore, MA performs better than K and MI.

## 5 SUMMARY

In this paper, a Monte-Carlo procedure using simulation techniques is proposed, which is the estimator that utilizes the density function of $y_t$ given $\alpha_t$ and the random draws from the transition equation. The proposed procedure improves the other non-linear filters developed in the past from the following three points: (i) computational time, (ii) simplicity of computer programming and (iii) no *ad hoc* assumptions.

[16] The state-space model (23) is taken in Kitagawa (1987) and Carlin, Polson and Stoffer (1992). In this section, we examine the same functions for the measurement and transition equations.

Table 7: Non-Normal Model ($\delta$ known)

| | | $T=10$ | | | $T=20$ | | | $T=30$ | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $L$ | P | F | S | P | F | S | P | F | S |
| $\delta=0.5$ | | | | | | | | | | |
| K | | 0.5569 | 0.5405 | 0.5373 | 0.5585 | 0.5412 | 0.5377 | 0.5594 | 0.5427 | 0.5394 |
| MA | 20 | 0.5442 | 0.5236 | 0.5213 | 0.5444 | 0.5250 | 0.5229 | 0.5476 | 0.5278 | 0.5274 |
| | 50 | 0.5432 | 0.5229 | 0.5206 | 0.5431 | 0.5236 | 0.5202 | 0.5454 | 0.5258 | 0.5232 |
| | 100 | 0.5433 | 0.5229 | 0.5206 | 0.5428 | 0.5234 | 0.5198 | 0.5451 | 0.5255 | 0.5224 |
| NI | 20 | 0.5432 | 0.5210 | 0.5207 | 0.5426 | 0.5205 | 0.5202 | 0.5449 | 0.5229 | 0.5226 |
| | 50 | 0.5432 | 0.5207 | 0.5203 | 0.5426 | 0.5202 | 0.5198 | 0.5449 | 0.5226 | 0.5222 |
| | 100 | 0.5432 | 0.5207 | 0.5203 | 0.5426 | 0.5202 | 0.5198 | 0.5449 | 0.5225 | 0.5221 |
| MI | 20 | 0.5524 | 0.5332 | 0.5354 | 0.5536 | 0.5359 | 0.5390 | 0.5560 | 0.5384 | 0.5414 |
| | 50 | 0.5471 | 0.5289 | 0.5314 | 0.5473 | 0.5302 | 0.5333 | 0.5492 | 0.5318 | 0.5347 |
| | 100 | 0.5441 | 0.5262 | 0.5288 | 0.5445 | 0.5276 | 0.5309 | 0.5462 | 0.5290 | 0.5321 |
| $\delta=1.0$ | | | | | | | | | | |
| K | | 0.6262 | 0.5671 | 0.5438 | 0.6336 | 0.5680 | 0.5428 | 0.6362 | 0.5683 | 0.5424 |
| MA | 20 | 0.5470 | 0.5280 | 0.5248 | 0.5496 | 0.5299 | 0.5274 | 0.5507 | 0.5309 | 0.5299 |
| | 50 | 0.5462 | 0.5272 | 0.5238 | 0.5480 | 0.5287 | 0.5248 | 0.5492 | 0.5295 | 0.5260 |
| | 100 | 0.5458 | 0.5269 | 0.5234 | 0.5480 | 0.5288 | 0.5247 | 0.5488 | 0.5291 | 0.5254 |
| NI | 20 | 0.5456 | 0.5244 | 0.5237 | 0.5477 | 0.5261 | 0.5251 | 0.5485 | 0.5268 | 0.5258 |
| | 50 | 0.5455 | 0.5241 | 0.5233 | 0.5477 | 0.5258 | 0.5246 | 0.5485 | 0.5265 | 0.5253 |
| | 100 | 0.5455 | 0.5240 | 0.5232 | 0.5476 | 0.5257 | 0.5245 | 0.5485 | 0.5265 | 0.5252 |
| MI | 20 | 0.5571 | 0.5418 | 0.5489 | 0.5601 | 0.5449 | 0.5522 | 0.5612 | 0.5456 | 0.5526 |
| | 50 | 0.5496 | 0.5356 | 0.5426 | 0.5537 | 0.5392 | 0.5472 | 0.5532 | 0.5382 | 0.5458 |
| | 100 | 0.5478 | 0.5339 | 0.5414 | 0.5503 | 0.5356 | 0.5436 | 0.5511 | 0.5361 | 0.5442 |

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

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

For the MA estimator, we have to compute $P_y(y_t|\alpha_t)$ by hand but not $P_\alpha(\alpha_t|\alpha_{t-1})$, and we do not need *ad hoc* assumptions such as choice of the nodes for the numerical integration procedure, choice of the importance density for the Monte-Carlo integration with importance sampling and choice of the density function of nuisance parameters for the Monte-Carlo integration with Gibbs sampling. The measurement equation is utilized for deriving the density $P_y(y_t|\alpha_t)$ while the transition equation is used to generate the random numbers of the state variable $\alpha_t$. The features of the MA are that the random draws of the state variable $\alpha_t$ (i.e., $\alpha_{i,t}$, $i=1,\cdots,n$) are generated from the transition equation (2) for all time $t$, and that the filtering algorithm requires the functional form of the density function of $y_t$ given $\alpha_t$, i.e., $P_y(y_t|\alpha_t)$, which has to be computed by hand. Thus, the proposed procedure improves with respect to computational time, simplicity of computer programming and no *ad hoc* assumptions.

In Table 8, we summarize the results of the Monte-Carlo experiments examined in Sections 4.1 – 4.5. There, $\bigcirc$, $\triangle$ and $\times$ denote excellent, fare and poor performances, respectively. We have the following results

Table 8: Summary of Monte-Carlo Experiments (Sections 4.1 – 4.5)

| Monte-Carlo Studies | | K | MA | NI | MI |
|---|---|---|---|---|---|
| Linear and Normal Model | (Table 1) | ◯ | × | ◯ | △ |
| Non-Linear Model I | (Table 3) | △ | × | ◯ | ◯ |
| Non-Linear Model II | (Table 5) | × | ◯ | ◯ | △ |
| Non-Linear Model III | (Table 6) | × | ◯ | △ | △ |
| Non-Normal Model | (Table 7) | × | ◯ | ◯ | △ |

based on the Monte-Carlo studies. If the range of the state variable is restricted, the MA performs quite better (see Tables 5 – 7). Otherwise, the MA does not work well, i.e., convergence speed of MA is too slow, compared with that of NI or MI (see Tables 1 and 3). Accordingly, in the case where the state variable is bounded from both sides, MA might be recommended.

# REFERENCES


Alspach, D.L. and H.W. Sorenson (1972), " Nonlinear Bayesian Estimation Using Gaussian Sum Approximations, " *IEEE Transactions on Automatic Control*, Vol.AC-17, No.4, pp.439 – 448.

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

Brown, B.W. and R.S. Mariano (1984), " Residual-based Procedures for Prediction and Estimation in a Nonlinear Simultaneous System, " *Econometrica*, Vol.52, No.2, pp.321 – 343.

Brown, B.W. and R.S. Mariano (1989), " Measures of Deterministic Prediction Bias in Nonlinear Models, " *International Economic Review*, Vol.30, No.3, pp.667 – 684.

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

Carlin, B.P., N.G. Polson and D.S. Stoffer (1992), " A Monte Carlo Approach to Nonnormal and Nonlinear State Space Modeling, " *Journal of the American Statistical Association*, Vol.87, pp.493 – 500.

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

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

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

Fahrmeir, L. (1992), " Posterior Mode Estimation by Extended Kalman Filtering for Multivariate Dynamic Generalized Linear Models, " *Journal of the American Statistical Association*, Vol.87, No.418, pp.501 – 509.

Fahrmeir, L. and G. Tutz (1994), *Multivariate Statistical Modeling Based on Generalized Linear Models*, Springer-Verlag.

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

Gelfand, A.E., S.E. Hills, H.A. Racine-Poon, and A.F.M. Smith (1990), " Illustration of Bayesian Inference in Normal Data Models Using Gibbs Sampling, " *Journal of the American Statistical Association*, Vol.85, No.412, pp.972 – 985.

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

Geman, S. and D. Geman (1984), " Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, " *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. Pami-6, No.6, pp.721 – 741.

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

Geweke, J. (1989a), " Modeling with Normal Polynomial Expansions, " in *Economic Complexity: Chaos, Sunspots, Bubbles and Nonlinearity*, edited by Barnett *et al.*, Cambridge University Press.

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

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

Kalman, R.E. (1960), " A New Approach to Linear Filtering and Prediction Problems, " *Journal of Basic Engineering, Transactions ASME*, Ser.D, 82, pp.35 – 45.

Kalman, R.E. and R.S. Bucy (1961), " New Results in Linear Filtering and Prediction Theory, " *Journal of Basic Engineering, Transactions ASME*, Ser.D, 83, pp.95 – 108.

Kitagawa, G. (1987), " Non-Gaussian State-Space Modeling of Nonstationary Time Series, " *Journal of the American Statistical Association*, Vol.82, pp.1032 – 1063 (with discussion).

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

McFadden, D. (1989), " A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration, " *Econometrica*, Vol.57, No.5, pp.995 – 1026.

Mariano, R.S. and B.W. Brown (1983), " Asymptotic Behavior of Predictors in a Nonlinear Simultaneous System, " *International Economic Review*, Vol.24, No.3, pp.523 – 536.

Mariano, R.S. and B.W. Brown (1989), " Stochastic Simulation, Prediction and Validation of Nonlinear Models, " in *Economics in Theory and Practice: An Eclectic Approach*, edited by L.R. Klein and J. Marquez, pp.17 – 36, Kluwer Academic Publishers.

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

Meinhold, R.J. and N.D. Singpurwalla (1989), " Robustification of Kalman Filter Models, " *Journal of the American Statistical Association*, Vol.84, pp.479 – 486.

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

Sorenson, H.W. and D.L. Alspach (1971), " Recursive Bayesian Estimation Using Gaussian Sums, " *Automatica*, Vol.7, pp.465 – 479.

Tanizaki, H. (1993a), " Kalman Filter Model with Qualitative Dependent Variable, " *The Review of Economics and Statistics*, Vol.75, No.4, pp.747 – 752.

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

Tanizaki, H. and R.S. Mariano (1994), " Prediction, Filtering and Smoothing in Nonlinear and Nonnormal Cases Using Monte-Carlo Integration, " *Journal of Applied Econometrics*, Vol.9, No.2, pp.163 – 179.

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

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

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