Kobe University Repository : Kernel

PDF issue: 2026-11-07

Non-linear and non-normal filter based on Monte-Carlo technique

Tanizaki, Hisashi

(Citation)
Systems Analysis Modelling Simulation,27(4):265-280

(Issue Date)
1997

(Resource Type)
journal article

(Version)
Accepted Manuscript

(URL)
https://hdl.handle.net/20.500.14094/90000571

KOBE UNIVERSITY

Non-Linear and Non-Normal Filter Based on Monte-Carlo Technique

Hisashi TANIZAKI
Kobe University, Kobe 657, JAPAN

Abstract: A non-linear and/or non-normal filter is proposed in this paper. Generating random draws of the state vector directly from the filtering density, the filtering estimate is obtained, which gives us a recursive algorithm. There, we do not evaluate any integration included in the density-based filtering algorithm such as the numerical integration procedure and the Monte-Carlo integration approach. The Monte-Carlo experiments indicate that the proposed non-linear and non-normal filter shows a good performance.

Key Words: Non-Linear, Non-Normal, Filtering, Random Draws, Rejection Sampling.

1 Introduction

A non-linear and non-normal filtering algorithm is proposed in this paper. Given random draws of the state vector which are directly generated from the filtering density, the filtering estimate is recursively obtained. There, we do not evaluate any integration included in the density-based filtering algorithm.

We have numerous density-based filtering algorithms. Kitagawa [6] and Kramer and Sorenson [8] proposed the numerical integration procedure. Tanizaki [11], Tanizaki and Mariano [12] and Mariano and Tanizaki [9] utilized the Monte-Carlo integration with importance sampling to derive non-linear and non-normal filtering algorithm. Moreover,

Carlin, Polson and Stoffer [2] suggested the Monte-Carlo integration procedure with Gibbs sampling.[1]

Thus, in this paper, a Monte-Carlo procedure of filtering algorithm using the simulation technique is proposed, where we utilize the random draws only. The procedure improves over the other non-linear filters developed in the past from the following three points, i.e., computational burden, simplicity of computer programming and no *ad hoc* assumptions.

The numerical procedure proposed in Kitagawa [6] and Kramer and Sorenson [8] has the disadvantages:

(i) Location of nodes has to be set by a researcher.

(ii) We have to derive the densities from measurement and transition equations by hand.

(iii) Computational burden increases more than proportionally as the dimension of the state variable is high.

The problems of the Monte-Carlo integration procedure with importance sampling developed by Tanizaki [11], Tanizaki and Mariano [12] and Mariano and Tanizaki [9] are:

(i) The importance density has to be appropriately chosen by a researcher.

(ii) We need to derive the densities from measurement and transition equations by hand.

Use of Monte-Carlo integration with Gibbs sampler (Carlin, Polson and Stoffer [2]) also has some problems:

(i) We need to assume the distributions of nuisance parameters.

(ii) The Gibbs sampler leads to a great amount of data storage.

(iii) This procedure also takes a lot of time computationally.

The Monte-Carlo procedure proposed in Tanizaki and Mariano [13] has the following disadvantages:

(i) The Monte-Carlo approach requires a large amount of data storage.

(ii) Precision of the state-variable estimate is poor.

[1] The non-linear and non-normal filter which Carlin, Polson and Stoffer [2] proposed is not a recursive algorithm. For the Gibbs sampler, see Geman and Geman [4].

An alternative procedure which we propose utilizes the Monte-Carlo technique. By generating random draws directly from filtering density, we evaluate the filtering estimates. For the random number generation, rejection sampling is adopted, which is a method of the random number generation from any distribution function.[2] The Monte-Carlo procedure does not need to evaluate the functional form of filtering density numerically.

2 Overview of Non-Linear and Non-Normal Filter

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

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

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

where $t = 1, \cdots, T$. A $p \times 1$ vector, y_t, is observable while a $k \times 1$ vector, α_t, is unobserved. ϵ_t and η_t are mutually independently distributed. $h_t(\cdot)$ and $f_t(\cdot)$ are $p \times 1$ and $k \times 1$ 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 $P(\cdot|\cdot)$ 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 $P_y(y_t|\alpha_t)$ and $P_\alpha(\alpha_t|\alpha_{t-1})$ as the density functions obtained from the measurement equation (1) and the transition equation (2).

The recursive density algorithm on filtering is known as follows (for example, see Kitagawa [6] and Harvey [5]):

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

[2] The rejection sampling is as follows: Let x be a random variable from a density function $f(\cdot)$. When we want to generate random draws from $f(\cdot)$, we need to find the density $g(\cdot)$ which satisfies $f(x) \leq cg(x)$ for all x, where c is constant. For $g(\cdot)$ we should choose the distribution function such that we can easily generate random draws. Define $w(x) \equiv f(x)/cg(x)$. Note that $0 \leq w(x) \leq 1$. Let u be a uniform random number between zero and one and v be a random number from $g(\cdot)$. Then, to generate a random draw from density function $f(\cdot)$, we take the following procedures:

(i) Generate u from a uniform distribution between zero and one.

(ii) Generate v from a density $g(\cdot)$.

(iii) Take v as x if $u \leq w(v)$, and return to (i) otherwise.

The disadvantage of random number generation by rejection sampling is that it takes a long time when $w(\cdot)$ is close to zero. Note that we do not have to generate a random draw in such a case.

For rejection sampling, see Boswell, Gore, Patil and Taillie [1], Knuth [7], and O'Hagan [10].

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

where 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 (3) yields $P(\alpha_t|Y_{t-1})$ given $P(\alpha_{t-1}|Y_{t-1})$ and 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 $P(\alpha_t|Y_t)$, $t = 1, \cdots, T$, can be obtained.

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

$$\begin{aligned} P(Y_T) &= \prod_{t=1}^{T} P(y_t|Y_{t-1}) \\ &= \prod_{t=1}^{T} \left(\int P_y(y_t|\alpha_t)P(\alpha_t|Y_{t-1})\mathrm{d}\alpha_t \right). \end{aligned} \tag{5}$$

$P(y_t|Y_{t-1})$ in equation (5) corresponds to the denominator of equation (4).

The above integration in equations (3) – (5) is evaluated in two ways, i.e., numerical integration and Monte-Carlo integration with importance sampling.

Kitagawa [6] and Kramer and Sorenson [8] proposed a non-linear and non-normal filter using numerical integration. Consider a scalar case of the state vector. Numerical integration requires the nodes, which are denoted by $\alpha_{i,t}$, $i = 0, \cdots, n$. 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 rectangle rule, a sum of trapezoids, Simpson's formula and so on.[3]

[3] Here, for simplicity of discussion, taking the numerical integration method by a sum of rectangles, the filtering algorithm given by equations (3) and (4) are represented as:

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

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

for $i = 0, 1, \cdots, n$ and $t = 1, \cdots, T$. $\alpha_{i,t}$ denotes the node. Note that $\alpha_{i,t} \geq \alpha_{j,t}$ when $i \geq j$ for all t.

Tanizaki [11], Tanizaki and Mariano [12] and Mariano and Tanizaki [9] developed a non-linear and non-normal filter with Monte-Carlo integration with importance sampling, where a recursive algorithm 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 $(r,s) = (t,t-1),(t,t)$, where the density function $P_I(\alpha_t)$ has to be appropriately specified by a researcher, which is called the importance density. 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 $P_I(\alpha_r)$. By Monte-Carlo integration with importance sampling, a recursive filtering algorithm of the weight functions is derived.[4]

Carlin, Polson and Stoffer [2] proposed a solution to multivariate state-space modeling, where they allowed for the possibilities of non-normal errors and non-linear functions in the state-space model. They introduced nuisance parameters into the model. Their crucial assumptions are that distribution functions of ϵ_t and η_t depend on nuisance parameters and that the distribution functions of the nuisance parameters are assumed, which are called the prior densities in the Bayesian framework. By the conditional densities of the state-variables and the posterior densities of the nuisance parameters, the Gibbs sampler is implemented.

3 Non-Linear and Non-Normal Filter: A Monte-Carlo Approach

An alternative procedure to non-linear and non-normal filtering problem is given in this paper. For a solution to non-linear and non-normal state-space model, in this section, we use the random draws to obtain prediction, filtering and smoothing estimates. Let $\alpha_{i,r|s}$

[4] The nonlinear filter with importance sampling are given by:

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

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

for $i = 1, \cdots, n$ and $t = 1, \cdots, T$. Each density is converted to the weight function. $\alpha_{i,t}$ is a random number generated from the imortance density $P_I(\alpha_t)$, which is appropriately assumed by a researcher.

be the i-th random draw from the density function of α_r given Y_s.

The Monte-Carlo filter is derived as follows. Suppose that the random draws $\alpha_{i,t-1|t-1}$ are available. We consider generating random numbers of α_t from the filtering density $P(\alpha_t|Y_t)$. By substituting the prediction density $P(\alpha_t|Y_{t-1})$ into equation (4), the filtering density $P(\alpha_t|Y_t)$ is represented as:

$$\begin{aligned} P(\alpha_t|Y_t) &= \frac{P_y(y_t|\alpha_t)P(\alpha_t|Y_{t-1})}{\int P_y(y_t|\alpha_t)P(\alpha_t|Y_{t-1})\mathrm{d}\alpha_t} \\ &\propto \int P_y(y_t|\alpha_t)P_\alpha(\alpha_t|\alpha_{t-1})P(\alpha_{t-1}|Y_{t-1})\mathrm{d}\alpha_{t-1} \\ &\approx \frac{1}{n}\sum_{i=1}^{n} P_y(y_t|\alpha_t)P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}) \\ &\propto \frac{1}{n}\sum_{i=1}^{n} w_1(\alpha_t;y_t)P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}), \end{aligned}$$

where $w_1(\alpha_t;y_t)$ satisfies the following conditions:

$$\begin{aligned} &w_1(\alpha_t;y_t) \propto P_y(y_t|\alpha_t), \\ &0 \leq w_1(\alpha_t;y_t) \leq 1. \end{aligned}$$

Thus, the filtering density $P(\alpha_t|Y_t)$ is approximately proportional to:

$$\frac{1}{n}\sum_{i=1}^{n} w_1(\alpha_t;y_t)P_\alpha(\alpha_t|\alpha_{i,t-1|t-1}).$$

When we cannot obtain the explicit functional form of the density $P(\alpha_t|Y_t)$, we consider directly generating random draws from the filtering density. We may use rejection sampling to obtain a random observation from $P(\alpha_t|Y_t)$. In general, the distribution which we want to sample is dominated by the $P_\alpha(\alpha_t|\alpha_{i,t-1|t-1})$ density. Therefore, for $t = 1, \cdots, T$, we choose $\alpha_{i,t-1|t-1}$ with probability $1/n$ (i.e., we choose i), sample the $P_\alpha(\alpha_t|\alpha_{i,t-1|t-1})$ random variable and accept it with probability $w_1(\alpha_t;y_t)$. Thus, the random numbers generated from $P(\alpha_t|Y_t)$ are obtained. There, the measurement equation (1) is utilized to obtain the acceptance probability $w_1(\alpha_t;y_t)$ while the transition equation (2) is used to generate random numbers of α_t.

Summarizing the above, the estimation procedure is:

(i) Pick up j with probalility $1/n$, i.e., choose $\alpha_{j,t-1|t-1}$ for $j = 1, \cdots, n$.

(ii) Given j, generate a random draw of α_t (i.e., $\alpha_{i,t}$) from the transition equation $\alpha_{i,t} = f_t(\alpha_{j,t-1|t-1}, \eta_{i,t})$, where $\eta_{i,t}$ is the i-th random number of η_t.

(iii) (a) If $\alpha_{i,t}$ is accepted with probability $w_1(\alpha_{i,t}; y_t)$, it is taken as a random draw from the filtering density at time t, i.e., $\alpha_{i,t|t}$.

(b) If $\alpha_{i,t}$ is not accepted with probability $w_1(\alpha_{i,t}, y_t)$, go to (ii) again.

(iv) Repeat (i) – (iii) for $i = 1, \cdots, n$.

In this procedure, if the acceptance probability is small, the steps (ii) and (iii)(b) are repeated forever. In order to avoid this situation, if the acceptance probability is too small, we may go to (i), pick up another j and repeat the above procedure.

The likelihood function (5) is evaluated as:

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

Note that we have

$$\int P_y(y_t|\alpha_t) P(\alpha_t|Y_{t-1}) \mathrm{d}\alpha_t = \frac{1}{n} \sum_{i=1}^{n} P_y(y_t|\alpha_{i,t|t-1}),$$

where $\alpha_{i,t|t-1}$ is generated from the transition equation $\alpha_{i,t|t-1} = f_t(\alpha_{j,t-1|t-1}, \eta_{i,t})$. The evaluation procedure of the likelihood function is:

(i) Pick up $\alpha_{j,t-1|t-1}$ out of $\alpha_{1,t-1|t-1}, \cdots, \alpha_{n,t-1|t-1}$ with probability $1/n$.

(ii) Generate a random draw for η_t, i.e, $\eta_{i,t}$.

(iii) From the transition equation, $\alpha_{j,t-1|t-1}$ and $\eta_{i,t}$, we can obtain $\alpha_{i,t|t-1}$.

The proposed procedure gives us a general solution to the non-linear and non-normal state-space model represented by equations (1) and (2).The features of the above procedure are:

(i) We do not need the functional form of $P_y(y_t|\alpha_t)$, $P_\alpha(\alpha_t|\alpha_{t-1})$ and $P(\alpha_r|Y_s)$ for $(r, s) = (t, t-1), (t, t)$.

(ii) the random draws of α_t have to be generated from the density $P(\alpha_r|Y_s)$.

(iii) Programming is very easy, compared with the other non-linear filters by numerical integration (Kitagawa [6] and Kramer and Sorenson [8]), Monte-Carlo integration with importance sampling (Tanizaki [11], Tanizaki and Mariano [12] and Mariano and Tanizaki [9]), and Monte-Carlo integration with Gibbs sampling (Carlin, Polson and Stoffer [2]).

(iv) The proposed filter has less computational burden.[5]

We show two examples to implement the procedure, which correspond to the special cases where the state-space model is linear in error terms.

Example 1 (Linear Measurement Equation and Normal Errors): In the case where the transition equation is non-linear in the lagged state-vector α_{t-1} but the measurement equation is linear in both the state-vector α_t and the error ϵ_t, we have the following state-space model:

$$\begin{aligned}
&\text{Measurement Equation:} \quad && h_{1,t}(y_t) = Z_t\alpha_t + \epsilon_t, \\
&\text{Transition Equation:} \quad && \alpha_t = f_{1,t}(\alpha_{t-1}) + \eta_t, \\
& && \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H_t & 0 \\ 0 & Q_t \end{pmatrix} \right),
\end{aligned}$$

where $h_{1,t}(\cdot)$, Z_t, $f_{1,t}(\cdot)$, H_t and Q_t are assumed to be known. We may specify $h_{1,t}(y_t) = y_t$ and $f_{1,t}(\alpha_{t-1}) = T_t\alpha_{t-1}$.

From $P_y(y_t|\alpha_t)$ and $P_\alpha(\alpha_t|\alpha_{t-1})$, the posterior density $P(\alpha_t|y_t, \alpha_{t-1})$ is obtained. Picking up $\alpha_{i,t-1|t-1}$ randomly, $P(\alpha_t|Y_t)$ is represented as average of the normal densities with mean $\Sigma_t\mu_{i,t}$ and variance Σ_t, i.e.,

$$P(\alpha_t|Y_t) \approx \frac{1}{n}\sum_{i=1}^{n} N(\Sigma_t\mu_{i,t}, \Sigma_t),$$

[5] In general, when acceptance probability (i.e., $w_1(\alpha_t; y_t)$ in this paper) is close to zero, it is known that the random number generation by rejection sampling takes a lot of time computationally. An answer to the problem is as follows. When $w_1(\alpha_t; y_t)$ is small, we do not have to generate a random draw of α_t based on $\alpha_{i,t-1|t-1}$. In such a case, we may choose $\alpha_{j,t-1|t-1}$ which is different from $\alpha_{i,t-1|t-1}$ for $j \neq i$ and perform rejection sampling based on $\alpha_{j,t-1|t-1}$ in order to generate a random number of α_t. Thus, we can reduce computational burden.

where

$$\mu_{i,t} = Q_t^{-1} f_{1,t}(\alpha_{i,t-1|t-1}) + Z_t' H_t^{-1} h_{1,t}(y_t),$$

$$\Sigma_t^{-1} = Z_t' H_t^{-1} Z_t + Q_t^{-1}.$$

Note that $\alpha_{i,t-1|t-1}$ is chosen at random. Thus, the random number of α_t given Y_t is obtained from $(1/n)\sum_{i=1}^{n} N(\Sigma_t \mu_{i,t}, \Sigma_t)$. That is, picking up $\alpha_{i,t-1|t-1}$ at random, the random number of α_t given Y_t is obtained from $N(\Sigma_t \mu_{i,t}, \Sigma_t)$.

Example 2 (Non-Linear Measurement Equation and Normal Errors): In the case where the transition equation is non-linear in the lagged state-vector α_{t-1} and the measurement equation is also non-linear in the state-vector α_t, we have the following state-space model:

$$\begin{aligned}
&\text{Measurement Equation:} && h_{1,t}(y_t) = h_{2,t}(\alpha_t) + \epsilon_t, \\
&\text{Transition Equation:} && \alpha_t = f_{1,t}(\alpha_{t-1}) + \eta_t, \\
& && \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} H_t & 0 \\ 0 & Q_t \end{pmatrix} \right),
\end{aligned}$$

where $h_{1,t}(\cdot)$, $f_{1,t}(\cdot)$, H_t and Q_t are assumed to be known.

From $P_y(y_t|\alpha_t)$ and $P_\alpha(\alpha_t|\alpha_{t-1})$, the posterior density $P(\alpha_t|y_t, \alpha_{t-1})$ cannot be obtained explicitly. Therefore, we may apply rejection sampling. We pick up $\alpha_{i,t-1|t-1}$ randomly. Then, we generate a normal random draw $N\left(f_{1,t}(\alpha_{t-1}), Q_t\right)$ and accept it with probability $w_1(\alpha_t; y_t)$, where

$$w_1(\alpha_t; y_t) = \exp\left(-\frac{1}{2} (h_{1,t}(y_t) - h_{2,t}(\alpha_t))' H_t^{-1} (h_{1,t}(y_t) - h_{2,t}(\alpha_t)) \right).$$

Thus, the Monte-Carlo procedure is implemented.

Table 1: Linear and Normal Model: Prediction and Filtering

	δ		n	$T=20$		$T=40$	
				P	F	P	F
BIAS	0.5	K		0.0061	0.0045	0.0052	0.0028
		M	500	0.0062	0.0045	0.0053	0.0030
		M	1000	0.0062	0.0049	0.0054	0.0030
	1.0	K		0.0105	0.0058	0.0069	0.0031
		M	500	0.0131	0.0082	0.0081	0.0044
		M	1000	0.0138	0.0088	0.0084	0.0046
RMSE	0.5	K		1.0628	0.7334	1.0698	0.7273
		M	500	1.0647	0.7357	1.0718	0.7301
		M	1000	1.0640	0.7353	1.0710	0.7298
	1.0	K		1.2766	0.7938	1.2786	0.7845
		M	500	1.3000	0.8267	1.3023	0.8179
		M	1000	1.2985	0.8259	1.3012	0.8177

4 Numerical Examples: Monte-Carlo Experiments

4.1 Linear and Normal Model (Tables 1 and 2)

Consider the following linear scalar system:

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

where $t = 1, \cdots, T$. We compare the following estimates: the Kalman filtering estimate[6] , and the Monte-Carlo filtering estimate. Note that the Kalman filtering estimate gives us

[6] In the case of linear and normal cases, Kalman filter is given by the conventional linear recursive algorithms. In the case of non-linear state-space model, Kalman estimate implies the extended Kalman filter in this paper, which is the non-linear filter based on the first-order Taylor series expansion.

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 ϵ_t and η_t for $t = 1, \cdots, T$, we obtain a set of data y_t and α_t, $t = 1, \cdots, T$, from the system, where $T = 20, 40$.

(ii) We choose $n = 500, 1000$ for the Monte-Carlo approach.

(iii) Given Y_s, obtain Kalman filtering estimate and Monte-Carlo filtering estimate.

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

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

where $\alpha_{r|s}$ takes Kalman filtering estimate and Monte-Carlo filtering estimate. 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.

It is expected in this experiment that the Kalman filtering estimate is better than the Monte-Carlo filter because the system is linear and normal.

In Table 1, given δ, K and M are compared for P and F, where K and M denote the Kalman filter estimate[7] and the proposed Monte-Carlo procedure, and P and F represent one-step ahead prediction (i.e., $(r, s) = (t, t-1)$) and filtering (i.e., $(r, s) = (t, t)$).

K is very close to M for all the cases although K is slightly better than M. For M, as n increases, precision of estimates is slightly improved.

We consider precision of estimates of unknown parameter δ. The simulation procedure is: (i) generate the data y_t, given δ, (ii) estimate δ, given y_t for $t = 1, \cdots, T$, and (iii) repeat (i) and (ii) G times to obtain G estimates of δ. Thus, Table 2 is based on G estimates of δ, where $G = 1000$. AVE, SER, RMSE, 25%, 50% and 75% denote the arithmetic average, the standard error, the root mean square error, the first quantile, the second quantile (median) and the third quantile, respectively, which are obtained from 1000 estimates of δ.

[7] In the cases of non-linear system (Sections 4.2 – 4.4), K represents the extended Kalman filter.

Table 2: Linear and Normal Model: Estimation of Unknown Parameter

T	δ		n	AVE	SER	RMSE	25%	50%	75%
20	0.5	K		0.382	0.308	0.330	0.180	0.450	0.620
		M	500	0.392	0.326	0.344	0.170	0.470	0.640
		M	1000	0.394	0.320	0.337	0.180	0.450	0.640
	1.0	K		0.896	0.204	0.229	0.860	0.960	1.010
		M	500	0.914	0.202	0.220	0.880	0.970	1.030
		M	1000	0.913	0.205	0.223	0.880	0.975	1.030
40	0.5	K		0.444	0.224	0.231	0.310	0.500	0.610
		M	500	0.458	0.237	0.241	0.320	0.510	0.640
		M	1000	0.458	0.235	0.238	0.310	0.510	0.635
	1.0	K		0.959	0.079	0.089	0.930	0.980	1.010
		M	500	0.970	0.074	0.080	0.950	0.990	1.010
		M	1000	0.970	0.074	0.080	0.950	0.990	1.010

We have the results from Table 2 that M is better than K because M is closer to the true value of δ than K. Both Tables 1 and 2 indicate that for M the case $n = 500$ is very close to the case $n = 1000$. In the proceeding experiments, accordingly, we do not estimate the case $n = 1000$.

4.2 Non-Linear Model I (Tables 3 and 4)

Consider the following non-linear scalar system:

$$\begin{aligned}
&\text{Measurement Equation:} && y_t = \alpha_t + \epsilon_t, \\
&\text{Transition Equation:} && \alpha_t = (1 - \delta + \delta\alpha_{t-1}^2)^{1/2}\eta_t, \\
& && \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right), \\
& && \alpha_0 \sim N(0, 1),
\end{aligned}$$

Table 3: Non-Linear Model I: Prediction and Filtering

	δ		n	$T=20$		$T=40$	
				P	F	P	F
BIAS	0.5	K		0.0021	0.0021	0.0018	0.0029
		M	500	0.0022	0.0018	0.0019	0.0025
	0.9	K		0.0011	0.0014	−0.0041	0.0021
		M	500	0.0011	0.0012	−0.0042	0.0020
RMSE	0.5	K		1.0215	0.7036	1.0242	0.7043
		M	500	1.0230	0.6939	1.0259	0.6945
	0.9	K		1.0848	0.6632	1.2029	0.6565
		M	500	1.0878	0.5865	1.2009	0.5863

where $0 \leq \delta < 1$ and $t = 1, \cdots, T$. We take $\delta = 0.5, 0.9$ The transition equation follows the first-order autoregressive conditional heteroscedasticity (ARCH(1)) process[8] , 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 Kalman filter does not show a good performance.

For precision of filtering estimates, M is much better than K (see Table 3). The estimates of δ obtained by M are much better than K with respect to AVE (see Table 4).

4.3 Non-Linear Model II (Table 5)

Consider the following of non-linear scalar system:

$$\begin{aligned}
&\text{Measurement Equation:} \quad y_t = \frac{\exp(\alpha_t)}{\exp(\alpha_t) + \exp(\epsilon_t)}, \\
&\text{Transition Equation:} \quad \alpha_t = \frac{\exp(\alpha_{t-1})}{\exp(\alpha_{t-1}) + \exp(\eta_t)}, \\
&\qquad \begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right), \\
&\qquad \alpha_0 \sim N(0, 1),
\end{aligned} \tag{6}$$

[8] Note that the ARCH model was proposed by Engle [3]. In this Monte-Carlo experiment, the unconditional variance of α_t is assumed to be one.

Table 4: Non-Linear Model I: Estimation of Unknown Parameter

T	δ		n	AVE	SER	RMSE	25%	50%	75%
20	0.5	K		0.376	0.350	0.372	0.000	0.320	0.680
		M	500	0.492	0.398	0.398	0.000	0.530	0.910
	0.9	K		0.628	0.355	0.447	0.350	0.740	0.970
		M	500	0.756	0.346	0.375	0.670	0.940	0.990
40	0.5	K		0.324	0.284	0.334	0.000	0.310	0.530
		M	500	0.456	0.354	0.357	0.060	0.470	0.795
	0.9	K		0.636	0.293	0.395	0.460	0.680	0.890
		M	500	0.800	0.263	0.282	0.730	0.910	0.990

Table 5: Non-Linear Model II: Prediction and Filtering

		n	$T = 20$		$T = 40$	
			P	F	P	F
BIAS	K		0.0252	0.0064	0.0256	0.0072
	M	500	−0.0026	−0.0026	−0.0016	−0.0015
RMSE	K		0.2030	0.2146	0.2041	0.2135
	M	500	0.2011	0.1975	0.2019	0.1976

Table 6: Non-Linear Model III: Prediction and Filtering

System			n	$T=20$		$T=40$	
				P	F	P	F
(i)	BIAS	K		0.1383	−0.2925	0.6127	0.2539
		M	500	−0.0162	−0.0150	0.2368	0.2247
	RMSE	K		11.6519	11.7561	11.5832	11.7027
		M	500	8.7961	8.3796	8.6877	8.2717
(ii)	BIAS	K		1.0053	1.5399	0.9152	1.1983
		M	500	0.0188	0.0281	0.2027	0.1550
	RMSE	K		14.3026	20.9436	14.8889	21.4744
		M	500	7.8308	5.3713	7.7329	5.3881

where $t = 1, \cdots, T$. In this experiment, both the measurement and the transition equations are logistic and accordingly non-linear[9] . The results are in Table 5.

4.4 Non-Linear Model III (Table 6)

Consider another example of the following two univariate nonstationary growth models, i.e., (i) equations (7) and (9) and (ii) equations (8) and (9)[10] :

$$\text{Measurement Equation:} \quad y_t = \frac{\alpha_t}{\sqrt{20}} + \epsilon_t, \tag{7}$$

$$\text{Measurement Equation:} \quad y_t = \frac{\alpha_t^2}{20} + \epsilon_t, \tag{8}$$

[9] 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 (6), rather than the above equation, is linearized in this experiment.

[10] The measurement equation (8) is taken in Kitagawa (1987) and Carlin, Polson and Stoffer (1992).

$$\text{Transition Equation:} \quad \begin{aligned} &\alpha_t = \frac{1}{2}\alpha_{t-1} + \frac{25\alpha_{t-1}}{1+\alpha_{t-1}^2} + 8\cos(1.2(t-1)) + \eta_t, \\ &\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 10 \end{pmatrix} \right), \\ &\alpha_0 \sim N(0,1), \end{aligned} \tag{9}$$

where $t = 1, \cdots, T$. In this example, the above non-linear system is taken. We take the exactly same simulation procedure as Sections 4.1 and 4.2. The results of the two systems are shown in Table 6.

The results obtained in Tables 5 and 6 are exactly same as those in Tables 1 and 3. That is, M performs much better than K, especially in the non-linear cases.

5 Summary

A non-linear and non-normal filtering algorithm is proposed in this paper. Given random draws of the state vector which are directly generated from the filtering density, the filtering estimate is recursively obtained. There, we do not evaluate any integration included in the density-based filtering algorithm.

Thus, in this paper, a Monte-Carlo procedure of filtering algorithm using the simulation technique is proposed, where we utilize the random draws only. The procedure improves over the other non-linear filters developed in the past from the following three points, i.e., computational time, simplicity of computer programming and no *ad hoc* assumptions.

References

[1] M.T. Boswell, S.D. Gore, G.P. Patil and C. Taillie, The Art of Computer Generation of Random Variables, in: C.R. Rao, Eds., *Handbook of Statistics*, Vol.9 (North-Holland, 1993) pp.661–721.

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

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

[4] S. Geman and D. Geman, Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.Pami-6 (1984) pp.721–741.

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

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

[7] D.E. Knuth, *The Art of Computer Programming*, Vol.2, (Addison-Wesley,1981).

[8] S.C. Kramer and H.W. Sorenson, Recursive Bayesian Estimation Using Piece-wise Constant Approximations, *Automatica*, Vol.24 (1988) pp.789–801.

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

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

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

[12] H. Tanizaki and R.S. Mariano, Prediction, Filtering and Smoothing in Nonlinear and Nonnormal Cases Using Monte-Carlo Integration, *Journal of Applied Econometrics*, Vol.9 (1994) pp.163–179.

[13] H. Tanizaki and R.S. Mariano, A Comparison of Three Non-linear and Non-normal Filters, Unpublished Manuscript.