CIRJE-F-1080

Particle rolling MCMC with double block sampling: conditional SMC update approach

Naoki Awaya
Graduate School of Economics, The University of Tokyo

Yasuhiro Omori
The University of Tokyo

March 2018

Particle rolling MCMC with double block sampling: conditional SMC update approach

Naoki Awaya* and Yasuhiro Omori†

September 15, 2017
Revised March 15, 2018

Abstract

An efficient simulation-based methodology is proposed for the rolling window estimation of state space models. Using the framework of the conditional sequential Monte Carlo update in the particle Markov chain Monte Carlo estimation, weighted particles are updated to learn and forget the information of new and old observations by the forward and backward block sampling with the particle simulation smoother. These particles are also propagated by the MCMC update step. Theoretical justifications are provided for the proposed estimation methodology. The computational performance is evaluated in illustrative examples, showing that the posterior distributions of model parameters and marginal likelihoods are estimated with accuracy. Finally, as a special case, our proposed method can be used as a new sequential MCMC based on Particle Gibbs, which is shown to outperform SMC^2 that is the promising alternative method based on Particle MH in the simulation experiments.

Keywords: Block sampling; Forward and backward sampling; Importance sampling; Particle Gibbs; Particle Markov chain Monte Carlo; Particle simulation smoother; Rolling window estimation; Sequential Monte Carlo; State space model; Structural change

*Graduate School of Economics, The University of Tokyo, Tokyo, Japan.
†Faculty of Economics, The University of Tokyo, Japan. E-mail: omori@e.u-tokyo.ac.jp.

1 Introduction

State space models have been popular and widely used in the analysis of economic and financial time series. They are flexible to capture the dynamics of the complex economic structure, but it is well-known that there have been several structural changes in the long economic time series. For example, it is often said that there has been a major structural change before and after the financial crisis in 2008. Accordingly, model parameter estimates are found to change in the econometric models depending on the selection of the sample period. That is, the parameters are not necessary to be constant and may be subject to such changes during the long sample period. If the precise time of a structural change is known, we could divide the sample period into two periods, before and after the structural change. However, it is usually unknown, and the change may occur gradually from one state to another. To reflect the recent unobserved structural change in the forecasting without delay, the rolling window estimation is used where we fix the number of observations to estimate model parameters and update the dataset to improve the forecasting performance.

In nonlinear or non-Gaussian state space models, it is often the case that the likelihood is not obtained analytically and that the maximum likelihood estimation is difficult to implement. Markov Chain Monte Carlo (MCMC) method is a popular and powerful technique in such a case to estimate model parameters and state variables by generating random samples from the posterior distribution given a set of observed data for various complex state space models. The highly efficient simulation algorithms have been proposed for linear Gaussian state space models (e.g. de Jong and Shephard (1995), Durbin and Koopman (2002)) and nonlinear non-Gaussian state space models (e.g. Shephard and Pitt (1997), Watanabe and Omori (2004)). However, it is computationally expensive to implement the MCMC simulation every time when we obtain a new observation.

To overcome this difficulty, we take an alternative approach based on sequential Monte Carlo (SMC) method (e.g. Doucet et al. (2001)). In the rolling window estimation, we incorporate a new observation and discard the oldest observation sequentially, using the weighted particles to approximate the target posterior density of the model parameters and state variables. Since the weight degeneracy problem arises when we generate (or discard) one state variable given others, we propose to sample a set of state variables together as a block at one time. The idea of block sampling in SMC has been explored and found to be highly efficient (e.g. Doucet et al. (2006), Polson et al. (2008)). Our proposed algorithm is also based on the block sampling, but different from other methods in that (1) we go forward to generate a set of newest state variables (which we call a forward block sampling) in adding a new observation and also go backward to generate a set of old state variables (which we call a backward block sampling) in removing the oldest observation to shift the rolling window, and (2) we sample a new particle path by proposing a cloud of candidates given each current particle path using the conditional SMC update of particle MCMC (PMCMC) (Andrieu et al. (2010)) and update its importance weight to approximate the new target posterior density. Theoretical justifications are provided to show that our target density is obtained as a marginal density of the artificial target density in the 'double' (forward and backward) block sampling. The closely related algorithm is SMC^2 (Chopin et al. (2013), Fulop and Li (2013)) which implements particle Metropolis-Hastings algorithm of PMCMC to sample from the target posterior distribution of model parameters and state variables sequentially where new observations are added sequentially one at a time. Our approach can also be applied to such a sequential estimation as a special case of the rolling window estimation.

The rest of the paper is organized as follows. In Section 2, we introduce the rolling window estimation in state space models and show that a simple particle rolling MCMC causes the weight degeneracy phenomenon using an example with the simulated data. Section 3 introduces a new methodology named double block sampling with the particle simulation smoother to overcome this difficulty. Theoretical justifications of the proposed method are provided in Section 4. Section 5 gives illustrative examples using a linear Gaussian state space model with the simulated data and a realized stochastic volatility model with S&P500 index returns data. Finally, Section 6 concludes the paper.

2 Particle rolling MCMC in state space models

2.1 Rolling window estimation in state space model

Consider the state space model which consists of a measurement equation and a state equation with an observation vector y_t, an unobserved state vector α_t given a static parameter vector θ. For the prior distribution of θ, we let $p(\theta)$ denote its prior probability density function. Further define $\alpha_{s:t} \equiv (\alpha_s, \alpha_{s+1}, \ldots, \alpha_t)$ and $y_{s:t} \equiv (y_s, y_{s+1}, \ldots, y_t)$. We assume that the distribution of y_t given $(y_{1:t-1}, \alpha_{1:t}, \theta)$ depends only on α_t and θ and that the distribution of α_t given $(\alpha_{1:t-1}, \theta)$ depends only on α_{t-1} and θ. The corresponding probability density functions are

$$p(y_t \mid y_{1:t-1}, \alpha_{1:t}, \theta) = p(y_t \mid \alpha_t, \theta) \equiv g_\theta(y_t \mid \alpha_t), \quad t = 1, \ldots, n, \tag{1}$$

$$p(\alpha_t \mid \alpha_{1:t-1}, \theta) = p(\alpha_t \mid \alpha_{t-1}, \theta) \equiv f_\theta(\alpha_t \mid \alpha_{t-1}), \quad t = 2, \ldots, n, \tag{2}$$

$$p(\alpha_1 \mid \theta) \equiv \mu_\theta(\alpha_1), \tag{3}$$

where $\mu_\theta(\alpha_1)$ denotes the unconditional probability density function for the initial value of α_1 given θ. To take account of the correlation between y_t and α_{t+1}, we further assume

$$
\begin{aligned}
p(y_t \mid y_{1:t-1}, \alpha_{1:t+1}, \theta) &= p(y_t \mid \alpha_t, \alpha_{t+1}, \theta) \equiv g_\theta(y_t \mid \alpha_t, \alpha_{t+1}), \quad t = 1, \ldots, n, \qquad (4)\\
p(\alpha_{t+1} \mid \alpha_{1:t}, y_{1:t}, \theta) &= p(\alpha_{t+1} \mid \alpha_t, y_t, \theta) \equiv f_\theta(\alpha_{t+1} \mid \alpha_t, y_t), \quad t = 1, \ldots, n-1. \qquad (5)
\end{aligned}
$$

As we shall illustrate in our empirical example, the negative correlation between the stock return y_t today and the log volatility α_{t+1} in the following day is well-known as a leverage effect or an asymmetry in the stochastic volatility models for the stock returns data (Omori et al. (2007)).

In the rolling window analysis of time series, the number of observations (or the window size) in the sample period is fixed and it is set equal to, for example, $L+1$. We estimate the posterior distribution of θ and $\alpha_{s:t}$ given the observations $y_{s:t}$ with $t = s + L$ for $s = 1, 2 \ldots$, and its probability density function is given by

$$
\begin{aligned}
\pi(\theta, \alpha_{s:t} \mid y_{s:t}) &\propto p(\theta)\mu_\theta(\alpha_s) g_\theta(y_s \mid \alpha_s) \left\{ \prod_{j=s+1}^{t} f_\theta(\alpha_j \mid \alpha_{j-1}, y_{j-1}) g_\theta(y_j \mid \alpha_j) \right\} \qquad (6)\\
&\propto p(\theta)\mu_\theta(\alpha_s) \left\{ \prod_{j=s+1}^{t} f_\theta(\alpha_j \mid \alpha_{j-1}) g_\theta(y_{j-1} \mid \alpha_j, \alpha_{j-1}) \right\} g_\theta(y_t \mid \alpha_t). \qquad (7)
\end{aligned}
$$

The contribution of this paper is to develop an efficient simulation-based method to construct a collection of particles $(\theta^n, \alpha^n_{s:t})$ with the importance weight $W^n_{s:t}$ $(n = 1, \ldots, N)$ which gives the exact approximation of the posterior distribution with its density $\pi(\theta, \alpha_{s:t} \mid y_{s:t})$ in the framework of the sequential Monte Carlo (SMC) (e.g. Doucet et al. (2001), Andrieu et al. (2010), Chopin et al. (2013)). We also prove that the posterior probability density function is obtained as a marginal density of our artificial target density in the particle rolling MCMC simulation.

2.2 Simple particle rolling MCMC

Assume that, at time $t-1$, we have a collection of particles $(\theta^n, \alpha^n_{s-1:t-1})$ with the importance weight $W^n_{s-1:t-1}$, $(n = 1, \ldots, N)$ which is a discrete approximation of $\pi(\theta, \alpha_{s-1:t-1} \mid y_{s-1:t-1})$. These particles are updated by adding a new observation y_t and removing the oldest observation y_{s-1}. If the weights of particles are degenerated, all particles are resampled and updated using the SMC sampler with the MCMC kernel (Del Moral et al. 2006). The algorithm consists of the following two steps.

(A1) Simple particle rolling MCMC sampler

Step 1 : *Adding a new observation y_t to the information set.*

- **1a** Generate α_t^n given $(\theta^n, \alpha_{s-1:t-1}^n)$ using some proposal distribution and construct a collection of particles $(\theta^n, \alpha_{s-1:t}^n)$ with the importance weight $W_{s-1:t}^n$ $(n = 1, \ldots, N)$ to approximate the posterior distribution with the density $\pi(\theta, \alpha_{s-1:t} \mid y_{s-1:t})$.
- **1b** If some degeneracy criterion is fulfilled, resample all the particles and set $W_{s-1:t}^n = 1/N$. Further, update $(\theta^n, \alpha_{s-1:t}^n)$ using the MCMC kernel of the invariant distribution with the density $\pi(\theta, \alpha_{s-1:t} \mid y_{s-1:t})$ $(n = 1, \ldots, N)$.

Step 2 *Removing the oldest observation y_{s-1} from the information set.*

- **2a** Discard α_{s-1}^n and construct a collection of particles $(\theta^n, \alpha_{s:t}^n)$ with the updated importance weight $W_{s:t}^n$ $(n = 1, \ldots, N)$ to approximate the posterior distribution with its density $\pi(\theta, \alpha_{s:t} \mid y_{s:t})$.
- **2b** If some degeneracy criterion is fulfilled, resample all the particles and set $W_{s:t}^n = 1/N$. Further, update $(\theta^n, \alpha_{s:t}^n)$ using the MCMC kernel of the invariant distribution with the density $\pi(\theta, \alpha_{s:t} \mid y_{s:t})$ $(n = 1, \ldots, N)$.

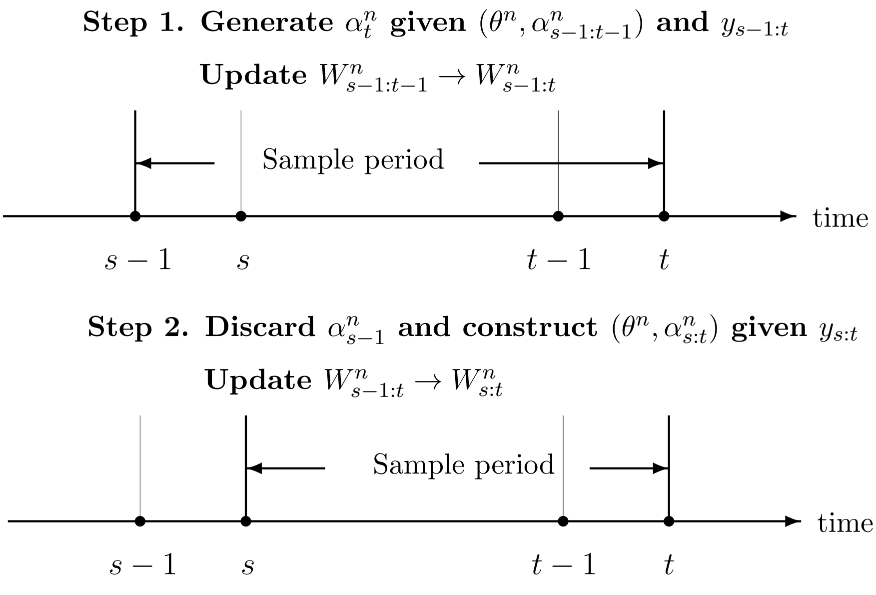

Figure 1: Rolling estimation in two steps.

Step 1.

In Step 1a, we generate α_t^n using some proposal density $q_{t,\theta^n}(\cdot \mid \alpha_{t-1}^n, y_t)$ and compute the importance weight

$$W_{s-1:t}^n \quad \propto \quad \frac{f_{\theta^n}(\alpha_t^n \mid \alpha_{t-1}^n, y_{t-1}) g_{\theta^n}(y_t \mid \alpha_t^n)}{q_{t,\theta^n}(\alpha_t^n \mid \alpha_{t-1}^n, y_t)} W_{s-1:t-1}^n, \tag{8}$$

where we add a new observation y_t to the information set. In Step 1b, we compute some degeneracy criterion such as the effective sample size (ESS) defined by

$$ESS_{s-1:t} = \frac{1}{\sum_{n=1}^N (W_{s-1:t}^n)^2}, \tag{9}$$

and the particles are resampled if ESS $< cN$ (e.g. $c = 0.5$).

In the additional MCMC implementation, the transition kernel K satisfies

$$\int \pi(\theta, \alpha_{s-1:t} \mid y_{s-1:t}) K\left((\theta, \alpha_{s-1:t}), (\tilde{\theta}, \tilde{\alpha}_{s-1:t})\right) d\theta d\alpha_{s-1:t} = \pi(\tilde{\theta}, \tilde{\alpha}_{s-1:t})$$

where we suppress n for convenience. This is also considered as the SMC sampler (Del Moral et al. (2006)) as follows. Suppose we have particles with the importance weight $W_{s-1:t}(= 1/N)$ obtained from the target distribution π_1 at time 1, and the target density π_2 at time 2 is the same as π_1 where

$$\pi_1(x_1) \quad = \quad \pi(\theta, \alpha_{s-1:t} \mid y_{s-1:t}), \quad \pi_2(x_2) = \pi(\tilde{\theta}, \tilde{\alpha}_{s-1:t} \mid y_{s-1:t}),$$

and $x_1 = (\theta, \alpha_{s-1:t})$ and $x_2 = (\tilde{\theta}, \tilde{\alpha}_{s-1:t})$. Then the MCMC sampling is equivalent to sampling from the artificial joint target distribution with the density $\tilde{\pi}_2$ defined as

$$\tilde{\pi}_2(x_1, x_2) \quad = \quad \pi_2(x_2) L(x_2, x_1), \quad L(x_2, x_1) = \frac{\pi_1(x_1) K(x_1, x_2)}{\pi_2(x_2)}, \tag{10}$$

where the so-called (unnormalized) incremental weight $\tilde{w}_2$ is

$$\tilde{w}_2(x_1, x_2) = \frac{\pi_2(x_2) L(x_2, x_1)}{\pi_1(x_1) K(x_1, x_2)} = 1, \tag{11}$$

and hence there is no change in the weight $W_{s-1:t}$. This is because the Markov kernel leaves $\pi_2(x_2)$ invariant. We note that one can also update $\alpha_{s-1:t}$ in the MCMC kernel step by Particle Gibbs sampler (Andrieu et al. (2010)), where it leaves the artificial target distribution invariant in the augmented space and the posterior distribution $\pi_2(x_2)$ is obtained as its marginal distribution.

Step 2.

In Step 2a, the target density is $p(\tilde{\alpha}_{s-1} \mid \tilde{\alpha}_s, \tilde{\theta})\pi(\tilde{\theta}, \tilde{\alpha}_{s:t} \mid y_{s:t})$ and

$$\begin{aligned} \frac{p(\tilde{\alpha}_{s-1} \mid \tilde{\alpha}_s, \tilde{\theta})\pi(\tilde{\theta}, \tilde{\alpha}_{s:t} \mid y_{s:t})}{\pi(\tilde{\theta}, \tilde{\alpha}_{s-1:t} \mid y_{s-1:t})} &= \frac{p(\tilde{\alpha}_{s-1} \mid \tilde{\alpha}_s, \tilde{\theta})\mu_{\tilde{\theta}}(\tilde{\alpha}_s)}{\mu_{\tilde{\theta}}(\tilde{\alpha}_{s-1})f_{\tilde{\theta}}(\tilde{\alpha}_s \mid \tilde{\alpha}_{s-1})g_{\tilde{\theta}}(y_{s-1} \mid \tilde{\alpha}_{s-1}, \tilde{\alpha}_s)} \\ &= g_{\tilde{\theta}}(y_{s-1} \mid \tilde{\alpha}_{s-1}, \tilde{\alpha}_s)^{-1}. \end{aligned} \tag{12}$$

Thus we update the importance weight

$$W_{s:t} \quad \propto \quad g_{\tilde{\theta}}(y_{s-1} \mid \tilde{\alpha}_{s-1}, \tilde{\alpha}_s)^{-1} W_{s-1:t}, \tag{13}$$

for each particle where we remove the oldest observation y_{s-1} from the information set, and discard α_{s-1}^n.

In Step2b, if some degeneracy criterion is fulfilled, resample all the particles by implementing MCMC algorithm as in Step 1b.

2.3 Weight degeneracy problem

In Step 1a of the above particle rolling MCMC sampler, we often use a prior density $f_\theta(\alpha_t|\alpha_{t-1}, y_{t-1})$ as a proposal density $q_{t,\theta}$. It is known to cause the weight degeneracy problem since the prior density does not take account of the new observation y_t. It is necessary to construct a good proposal to approximate $p(\alpha_t \mid \alpha_{t-1}, y_t, \theta) \equiv f_\theta(\alpha_t \mid \alpha_{t-1}, y_{t-1})g_\theta(y_t \mid \alpha_t)/p(y_t \mid \alpha_{t-1}, y_{t-1}, \theta)$. However, as we shall see in the following example, even if we are able to use a fully adapted proposal density, $q_{t,\theta}(\alpha_t \mid \alpha_{t-1}, y_{s-1:t}) = p(\alpha_t \mid \alpha_{t-1}, y_t, \theta)$, we still have the weight degeneracy phenomenon. In Step 2a, the weight degeneracy is more serious due to the incremental weight (12).

Example 1.

Consider the following univariate linear Gaussian state space model:

$$\begin{aligned} y_t &= \alpha_t + \epsilon_t, \ \epsilon_t \sim \mathcal{N}(0, \sigma^2), \ t = 1, \ldots, 2000 \\ \alpha_{t+1} &= \mu + 0.25(\alpha_t - \mu) + \eta_t, \ \eta_t \sim \mathcal{N}(0, 2\sigma^2), \ t = 1, \ldots, 2000, \\ \alpha_1 &= \mu + \frac{\eta_0}{\sqrt{1 - 0.25^2}}, \ \eta_0 \sim \mathcal{N}(0, 2\sigma^2), \end{aligned}$$

where $\theta = (\mu, \sigma^2)'$ is a model parameter vector, and we adopt conjugate priors, $\mu \mid \sigma^2 \sim \mathcal{N}(0, 10\sigma^2)$ and $\sigma^2 \sim \mathcal{IG}(5/2, 0.05/2)$. We generate MCMC samples from the posterior distributions of θ and $\alpha_{s:t}$ given $y_{s:t}$ using the particle rolling MCMC sampler with the window size of $L + 1 = 1000$ for $s = 2, 3, \ldots, 1001$ and $t = 1001, 1002, \ldots, 2000$. Taking account of the large number of observations, the posterior distribution will not be sensitive to the choice of the prior distribution.

For the MCMC sampling in Steps 1b and 2b, we update the particles of state variables with the simulation smoother (e.g. de Jong and Shephard (1995), Durbin and Koopman (2002)) and those of parameters by sampling from their conditional posterior distribution using Gibbs sampler. For each step of Steps 1b and 2b, the MCMC sampling is implemented only once to update the particles. We set $N = 1000$ and use a fully adapted proposal density, $q_{t,\theta}(\alpha_t \mid \alpha_{t-1}, y_{s-1:t}) = p(\alpha_t \mid \alpha_{t-1}, y_t, \theta)$ in order to evaluate the weight degeneracy in the best-case scenario. To evaluate the weight degeneracy in Steps 1a and 2a, we define two ratios:

$$R_{1t} = \frac{ESS_{s-1:t}}{ESS_{s-1:t-1}}, \quad R_{2t} = \frac{ESS_{s:t}}{ESS_{s-1:t}}. \tag{14}$$

The ratio R_{1t} measures the relative magnitude of ESS in Step 1a after adding a new observation when compared with that of the previous step. If the particles are scattered around the state space as in the previous step, R_{1t} would be close to 1. If, on the other hand, we have the weight degeneracy problem, it will be close to 0. Similarly, the ratio R_{2t} measures the relative magnitude of ESS in Step 2a after removing the oldest observation when compared with that of the previous step.

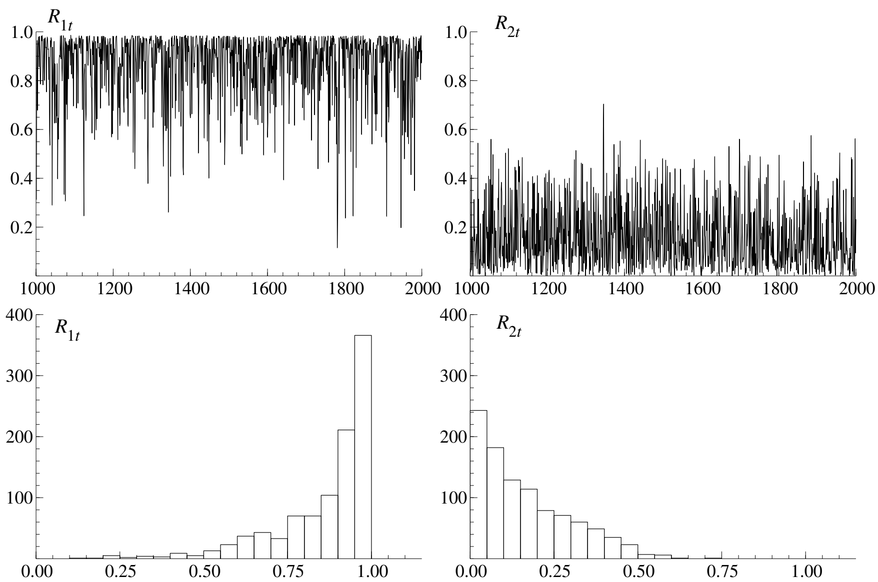

Figure 2: Time series plots and histograms of R_{1t} (left panels) and R_{2t} (right panels) for the rolling window estimation of $p(\theta, \alpha_{t-999:t} \mid y_{t-999:t}), t = 1001, \ldots, 2000$ in the linear Gaussian state space model. The fully adapted proposal density is used for $q_{t,\theta}$.

The time series plot and histogram of R_{1t} are shown in the left panels of Figure 2. The mean and standard deviation of R_{1t} are 0.86 and 0.15. Although R_{1t}'s are close to one, they are sometimes found to become very low. In such a case, the ESS falls rapidly and we need to implement the MCMC update of state variables α, which is time-consuming. Even if the proposal $q_{t,\theta}$ is a fully adapted density $p(\alpha_t \mid \alpha_{t-1}, y_t)$, we still have the weight degeneracy, which suggests we need to take a further step to improve ESS. On the other hand, the right panels of Figure 2 show the time series plot and histogram of R_{2t} where the mean and standard deviation of R_{2t} are 0.16 and 0.14. The R_{2t}'s are found to be extremely small, which indicates that the weight degeneracy problem in removing the oldest observation is more serious than in adding a new observation.

3 Particle rolling MCMC with double block sampling

To overcome this difficulty of the weight degeneracy, we propose the novel block sampling approach for $\pi(\theta, \alpha_{s:t} \mid y_{s:t})$. When we add a new observation y_t, we sample a set of $K+1$ state variables, $\alpha^n_{t-K:t}$, as a block given $(\theta^n, \alpha^n_{s-1:t-K-1})$ for $n = 1, \ldots, N$. The block sampling scheme for $\alpha^n_{t-K:t}$ is constructed using the approach of the conditional SMC update (Andrieu et al. (2010)), which is a special type of PMCMC called Particle Gibbs. We fix one particle path out of $\{\alpha^{n,m}_{t-K:t}\}_{m=1}^M$ with some prespecified ancestral lineage and other $M-1$ particle paths are randomly updated as usual given the selected path. Finally, one path is selected among $\{\alpha^{n,m}_{t-K:t}\}_{m=1}^M$.

It is well-known that sampling state variables as a block reduces the weight degeneracy drastically in the particle filtering given the parameter θ (Doucet et al. (2006)) and that it improves the sampling efficiency in the MCMC simulation (e.g. Shephard and Pitt (1997), Watanabe and Omori (2004)). We call it the forward block sampling since we construct a candidate path sequentially from α^n_{t-K} to α^n_t. Further, when removing the oldest observation y_{s-1}, we sample a set of $K+1$ state variables $\{\alpha^n_{s-1:s+K-1}\}$ as a block in a similar manner but construct a candidate path sequentially from α^n_{s+K-1} to α^n_{s-1} in the opposite direction. Thus we call it the backward block sampling. The general algorithm is described below with Figure 3, and the details are given in the following subsections.

(A2) Particle rolling MCMC sampler with double block sampling

---

Suppose that we have a collection of particles $\{\theta^n, \alpha^n_{s-1:t-1}\}$ with the importance weight $W^n_{s-1:t-1}$ ($n = 1, \ldots, N$) to approximate the posterior distribution with the density $\pi(\theta, \alpha_{s-1:t-1} | y_{s-1:t-1})$.

Step 1 : *Adding a new observation y_t to the information set.*

1a Generate a block of $\alpha^n_{t-K:t}$ given $(\theta^n, \alpha^n_{s-1:t-K-1})$ and $y_{s-1:t}$, and construct a collection of particles $(\theta^n, \alpha^n_{s-1:t})$ with the importance weight $W^n_{s-1:t}$ $(n = 1, \ldots, N)$ to approximate the posterior distribution with the density $\pi(\theta, \alpha_{s-1:t} \mid y_{s-1:t})$. The particle simulation smoother is implemented to improve the mixing property.

1b If some degeneracy criterion is fulfilled, resample all the particles and set $W^n_{s-1:t} = 1/N$. Further, update particles $(\theta^n, \alpha^n_{s-1:t})$ using the MCMC kernel of the invariant distribution with the density $\pi(\theta, \alpha_{s-1:t} \mid y_{s-1:t})$, $(n = 1, \ldots, N)$.

Step 2 *Removing the oldest observation y_{s-1} from the information set.*

2a Generate $\alpha^n_{s-1:s+K-1}$ given $(\theta^n, \alpha^n_{s+K:t})$ and $y_{s:t}$, and construct a collection of particles $(\theta^n, \alpha^n_{s:t})$ with the importance weight $W^n_{s:t}$ $(n = 1, \ldots, N)$ to approximate the posterior distribution with the density $\pi(\theta, \alpha_{s:t} \mid y_{s:t})$. Discard α^n_{s-1} and the particle simulation smoother is implemented.

2b If some degeneracy criterion is fulfilled, resample all the particles and set $W^n_{s:t} = 1/N$. Further, update particles $(\theta^n, \alpha^n_{s:t})$ using the MCMC kernel of the invariant distribution with the density $\pi(\theta, \alpha_{s:t} \mid y_{s:t})$, $(n = 1, \ldots, N)$.

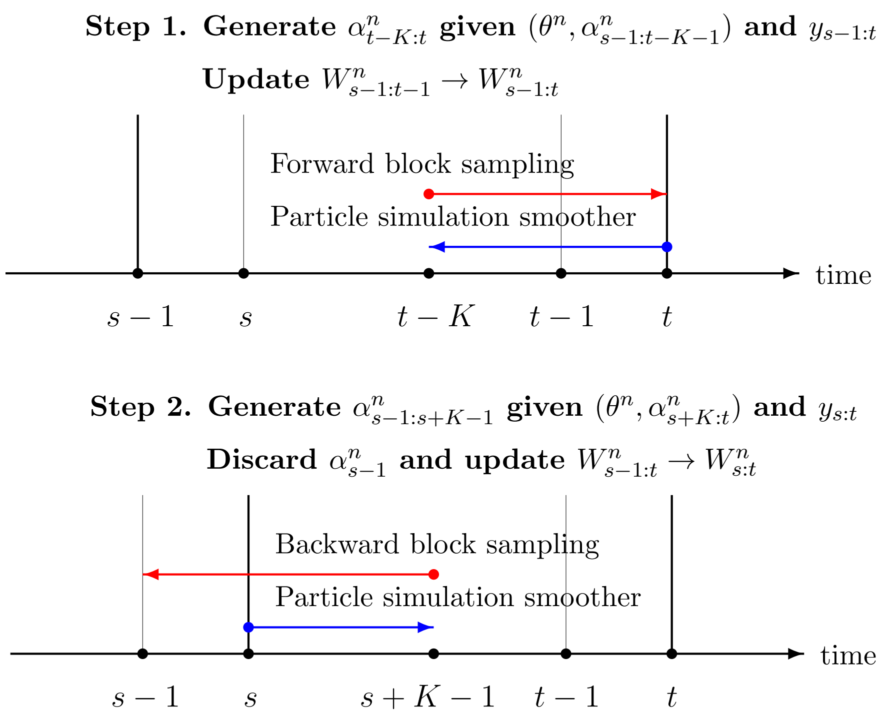

Figure 3: Double block sampling with particle simulation smoother.

3.1 Forward block sampling

Recent studies on Monte Carlo methods consider generating a cloud of values for one particle path. Andrieu et al. (2010) proposed particle Gibbs algorithms in which a lot of candidates are generated by the modified version of SMC, named conditional SMC, and determine one of the generated paths to sample from the posterior distribution of state variables. The SMC^2 or marginalized resample-move techniques in Chopin et al. (2013) and Fulop and Li (2013) is a nested SMC algorithm which generates a cloud of particles to compute the importance weight of particles approximating $p(\theta \mid y_{1:t})$ sequentially. This paper proposes a novel block sampling algorithm where we use the idea of the conditional SMC update to determine new particles and compute their importance weights to approximate the posterior density $\pi(\theta, \alpha_{s:t} \mid y_{s:t})$ in the rolling estimation with $t - s = L$ fixed.

First, we 'move' the particles $\alpha^n_{t-K:t-1}$ using the conditional SMC update for each $n = 1, \ldots, N$. That is, we generate a number of candidates $\alpha^{n,m}_{t-K:t-1}$ $(m = 1, \ldots, M)$ where the current 'lineage' $\alpha^n_{t-K:t-1}$ is fixed as one of candidates. The $\alpha^{n,m}_t$ is also generated conditional on $\alpha^{n,m}_{t-K:t-1}$. Second, one determines which lineage is appropriate as $\alpha^n_{t-k:t}$ and compute its importance weight for each $n = 1, \ldots, N$.

Before we describe the sampling algorithm, we introduce the 'parent' index variable a_j $(j = t-K, \ldots, t-1)$ and random indices k_j $(j = t-K, \ldots, t)$ and k^*_j $(j = t-K, \ldots, t)$ following the rules

$$
\begin{aligned}
a_j^{k_{j+1}} &= k_j, \quad j = t-K, \ldots, t-1, \quad k_t = 1, && (15)\\
a_j^{k^*_{j+1}} &= k^*_j, \quad j = t-K, \ldots, t-1, && (16)
\end{aligned}
$$

in order to make the expression simple. The details of Step 1a are given as follows.

Step 1a. Forward block sampling for $\alpha^n_{t-K:t}$ given $(\theta^n, \alpha^n_{s-1:t-K-1})$ and $y_{s-1:t}$

For each n, we first generate M particle paths, $\alpha^{n,1:M}_{t-K:t} \equiv (\alpha^{n,1}_{t-K:t}, \ldots, \alpha^{n,M}_{t-K:t})$, and sample one path, $\alpha^n_{t-K:t}$, from $\alpha^{n,1:M}_{t-K:t}$ as below.

(1) Sample k_j from $1:M$ with probability $1/M$ $(j = t-K, \ldots, t-1)$ and set

$$
(\alpha^{n,k_{t-K}}_{t-K}, \ldots, \alpha^{n,k_{t-1}}_{t-1}) = \alpha^n_{t-K:t-1}, \quad (a^{k_{t-K+1}}_{t-K}, \ldots, a^{k_t}_{t-1}) = (k_{t-K}, \ldots, k_{t-1}),
$$

where $\alpha^n_{t-K:t-1}$ is a current sample with the importance weight $W^n_{s-1:t-1}$.

(2) Set $\alpha^{n,a^m_{t-K-1}}_{t-K-1} = \alpha^n_{t-K-1}$ for all m according to the convention, and sample $\alpha^{n,m}_{t-K} \sim q_{t-K,\theta^n}(\cdot \mid \alpha^n_{t-K-1}, y_{t-K})$ for each $m \in \{1, \ldots, M\} \setminus \{k_{t-K}\}$. Let $j = t-K+1$.

(3) Sample $a_{j-1}^m \sim \mathcal{M}(V_{j-1,\theta^n}^{1:M})$ and $\alpha_j^{n,m} \sim q_{j,\theta^n}(\cdot \mid \alpha_{j-1}^{n,a_{j-1}^m}, y_j)$ for each $m \in \{1,\dots,M\} \setminus \{k_j\}$ where $V_{j-1,\theta^n}^{1:M} \equiv (V_{j-1,\theta^n}^1, \dots, V_{j-1,\theta^n}^M)$ and

$$V_{j,\theta^n}^m = \frac{v_{j,\theta^n}(\alpha_{j-1}^{n,a_{j-1}^m}, \alpha_j^{n,m})}{\sum_{i=1}^M v_{j,\theta^n}(\alpha_{j-1}^{n,a_{j-1}^i}, \alpha_j^{n,i})}, \tag{17}$$

$$v_{j,\theta^n}(\alpha_{j-1}^{n,a_{j-1}^m}, \alpha_j^{n,m}) = \frac{f_{\theta^n}(\alpha_j^{n,m} \mid \alpha_{j-1}^{n,a_{j-1}^m}, y_{j-1}) g_{\theta^n}(y_j \mid \alpha_j^{n,m})}{q_{j,\theta^n}(\alpha_j^{n,m} \mid \alpha_{j-1}^{n,a_{j-1}^m}, y_j)}, \quad m = 1, \dots, M. \tag{18}$$

(4) If $j < t-1$, set $j \leftarrow j+1$ and go to (3). Otherwise, sample $\alpha_t^{n,m}$ $(m = 1, \dots, M)$ and k_t^* as follows.

   (i) Sample $\alpha_t^{n,1} \sim q_{t,\theta^n}(\cdot \mid \alpha_{t-1}^{n,k_{t-1}})$.

   (ii) Sample $a_{t-1}^{n,m} \sim \mathcal{M}(V_{t-1,\theta^n}^{1:M})$ and $\alpha_t^{n,m} \sim q_{t,\theta^n}(\cdot \mid \alpha_{t-1}^{n,a_{t-1}^m}, y_t)$ for for each $m \in \{2, \dots, M\}$.

   (iii) Sample $k_t^* \sim \mathcal{M}(V_{t,\theta^n}^{1:M})$ and obtain k_j^* $(j = t-1, \dots, t-K)$ using (16).

(5) Let $\alpha_{s-1:t}^n = (\alpha_{s-1}^n, \dots, \alpha_{t-K-1}^n, \alpha_{t-K}^{n,k_{t-K}^*}, \dots, \alpha_t^{n,k_t^*})$ and compute the importance weight

$$W_{s-1:t}^n \propto \hat{p}(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}^n, \theta^n) \times W_{s-1:t-1}^n. \tag{19}$$

where

$$\hat{p}(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}^n, \theta^n) = \frac{1}{M} \sum_{m=1}^M v_{t,\theta^n}(\alpha_{t-1}^{n,a_{t-1}^{n,m}}, \alpha_t^{n,m}). \tag{20}$$

(6) Implement the particle simulation smoother to sample $(k_{t-K}^*, k_{t-K+1}^*, \dots, k_t^*)$ jointly. Generate $k_j^* \sim \mathcal{M}(\bar{V}_{j,\theta}^{1:M})$, $j = t-1, \dots, t-K$, recursively where

$$\bar{V}_{j,\theta}^m \equiv \frac{V_{j,\theta}^m f_\theta(\alpha_{j+1}^{k_{j+1}^*} \mid \alpha_j^m, y_{j+1})}{\sum_{i=1}^M V_{j,\theta}^i f_\theta(\alpha_{j+1}^{k_{j+1}^*} \mid \alpha_j^i, y_{j+1})}, \quad m = 1, \dots, M, \tag{21}$$

and set $\alpha_{s-1:t}^n = (\alpha_{s-1}^n, \dots, \alpha_{t-K-1}^n, \alpha_{t-K}^{n,k_{t-K}^*}, \dots, \alpha_t^{n,k_t^*})$.

In Step (5), we use the notation $\hat{p}(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}^n, \theta^n)$ since it is an unbiased estimator of $p(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}^n, \theta^n)$ as we shall show in Proposition 4.2.

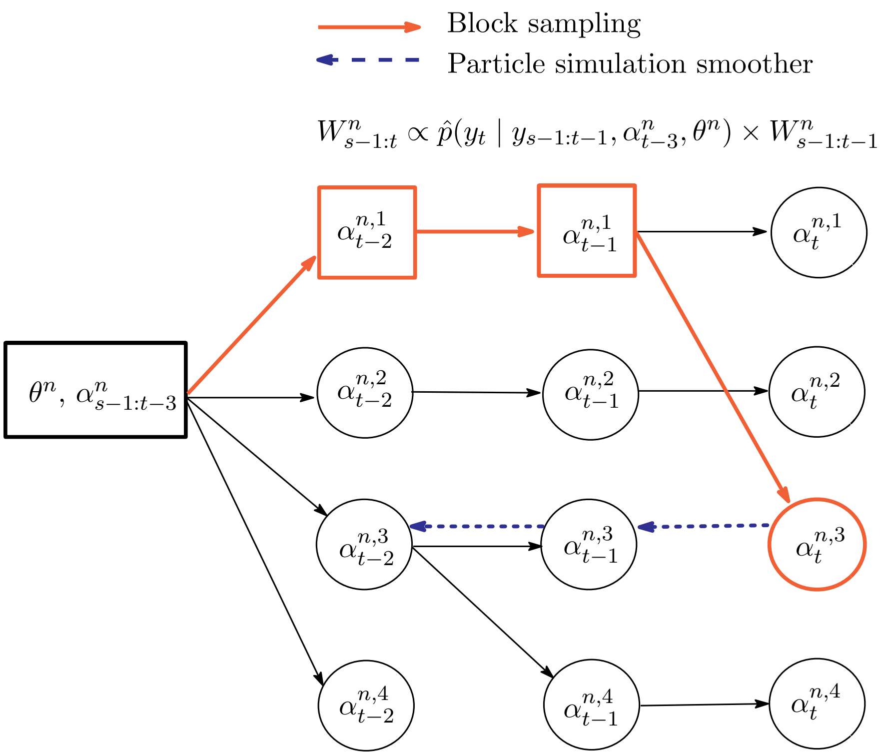

Figure 4: Forward block sampling and particle simulation smoother. $K = 2$, $M = 4$. $(\theta_n, \alpha^n_{s-1:t-1})$ is fixed and shown in the rectangle ($k_{t-2} = k_{t-1} = 1$). Other state variables in the circle are generated. $k^*_{t-2:t} = \{1, 1, 3\}$ in the block sampling, while $k^*_{t-2:t} = \{3, 3, 3\}$ in the particle simulation smoother.

Figure 4 illustrates an example with $K = 2$ and $M = 4$. The current sample is $(\theta^n, \alpha^n_{s-1:t-1})$ and we set $\alpha^{n,1}_{t-2} = \alpha^n_{t-2}$, $\alpha^{n,1}_{t-1} = \alpha^n_{t-1}$ with $k_{t-2} = k_{t-1} = 1$. They are fixed and shown in the rectangle. Other state variables are generated as above. In the block sampling, $M = 4$ paths are generated and $k^*_t = 3$ is selected (and hence $k^*_{t-1} = k^*_{t-2} = 1$). By implementing the particle simulation smoother, $k^*_{t-1} = k^*_{t-2} = 3$ are selected, and the updated particle is $(\theta^n, \alpha_{s-1:t})$ with the importance weight $W^n_{s-1:t}$ where $\alpha^n_{t-2} = \alpha^{n,3}_{t-2}$, $\alpha^n_{t-1} = \alpha^{n,3}_{t-1}$ and $\alpha^n_t = \alpha^{n,3}_t$.

Remark 1. The auxiliary particle filter proposed by Pitt and Shephard (1999) is well-known to improve ESS by resampling α_{t-1} and α_t simultaneously to reflect the likelihood information at time t. The forward block sampling above extends this idea of sampling $\alpha_{t-1:t}$ to sampling $\alpha_{t-K:t}$ for $K > 1$.

Remark 2. We can skip (6), the step of the particle simulation smoother, but it is expected to reduce the weight degeneracy phenomenon and produce more stable and accurate estimation results.

3.2 Backward block sampling

Suppose we have a set of particles with the importance weights $\{(\theta^n, \alpha^n_{s-1:t}), W^n_{s-1:t}\}$ ($n = 1, \ldots, N$) to approximate $\pi(\theta, \alpha_{s-1:t} \mid y_{s-1:t})$ and construct a new set of particles with the importance weights $\{(\theta^n, \alpha^n_{s:t}), W^n_{s:t}\}$ to approximate $\pi(\theta, \alpha_{s:t} \mid y_{s:t})$ by discarding the old information of y_{s-1}. In this section, we propose a backward block sampling, which samples particles sequentially as a block in the reverse order. That is, one generates a cloud of particles of $\alpha^{n,m}_{s+K-1}, \alpha^{n,m}_{s+K-2}, \ldots, \alpha^{n,m}_{s-1}$ ($m = 1, \ldots, M$) given $(\alpha^n_{s+K:t}, \theta^n)$ and $y_{s:t}$ targeting $\pi(\theta, \alpha_{s-1:t} \mid y_{s:t})$, and stochastically determines the new values of $\alpha^n_{s:t}$ (discarding α^n_{s-1}) with the updated importance weight

$$W^n_{s:t} \propto \frac{1}{\hat{p}(y_{s-1} \mid y_{s:t}, \alpha^n_{s+K}, \theta^n)} W^n_{s-1:t},$$

where we 'discount' the previous weight by the estimated old conditional likelihood $\hat{p}(y_{s-1} \mid y_{s:t}, \alpha^n_{s+K}, \theta^n)$. The incremental weight, $\hat{p}(y_{s-1} \mid y_{s:t}, \alpha^n_{s+K}, \theta^n)^{-1}$, depends only on α^n_{s+K}, while it does on both α^n_{s-1} and α^n_s in (12) for the simple particle rolling MCMC sampler. In our backward block sampling, the weights are expected to be allocated more uniformly to the particles.

This idea of 'backward filtering' algorithm is essentially the same as that of forward filtering. This is because, using the probability density $p(\alpha_{j-1} \mid \alpha_j, \theta)$, the joint prior probability density of the state variables $\alpha_{s-1:t}$ can be rewritten as

$$p(\alpha_{s-1:t} \mid \theta) = \left\{ \prod_{j=s}^{t} p(\alpha_{j-1} \mid \alpha_j, \theta) \right\} \mu_\theta(\alpha_t),$$

and thus the posterior density of θ and $\alpha_{s-1:t}$ can be expressed in the reverse order:

$$\begin{aligned}
&\pi(\theta, \alpha_{s-1:t} \mid y_{s-1:t}) \\
&\quad \propto \ p(\theta)\mu_\theta(\alpha_{s-1}) \left\{ \prod_{j=s}^{t} f_\theta(\alpha_j \mid \alpha_{j-1}) g_\theta(y_{j-1} \mid \alpha_{j-1}, \alpha_j) \right\} g_\theta(y_t \mid \alpha_t) \\
&\quad \propto \ p(\theta) \left\{ \prod_{j=s}^{t} p(\alpha_{j-1} \mid \alpha_j, \theta) g_\theta(y_{j-1} \mid \alpha_{j-1}, \alpha_j) \right\} \mu_\theta(\alpha_t) g_\theta(y_t \mid \alpha_t).
\end{aligned}$$

To implement the backward block sampling, we use an appropriate proposal density which approximates the 'posterior' density of α_{j-1} given (α_j, y_j, θ),

$$p(\alpha_{j-1} \mid \alpha_j, y_{j-1}, \theta) \propto p(\alpha_{j-1} \mid \alpha_j, \theta) g_\theta(y_{j-1} \mid \alpha_{j-1}, \alpha_j).$$

The 'prior' density $p(\alpha_{j-1} \mid \alpha_j, \theta)$ is one of candidate densities, which is obtained for linear Gaussian state equations such as AR(1) processes.

Before we describe the backward block sampling which generates a cloud of particles based on $(\alpha^n_{s+K:t}, \theta^n)$, we define the notation for the particle index as in the forward block sampling but in the reverse order. A 'parent' particle of α^m_j is chosen from $\alpha^{1:M}_{j+1}$ (not from $\alpha^{1:M}_{j-1}$) and consequently a^m_{j+1} denotes its parent's index. Hence the relationship of a^m_{j+1} and k_j is given by

$$a^{k_j}_{j+1} = k_{j+1}, \quad j = s+K-2, \dots, s-2. \tag{22}$$

The proposed backward block sampling is described as follows.

Step 2a. Backward block sampling for $\alpha^n_{s-1:s+K-1}$ given $(\theta^n, \alpha^n_{s+K:t})$ and $y_{s:t}$

For each n, we first generate M particle paths, $\alpha^{n,1:M}_{s-1:s+K-1} \equiv (\alpha^{n,1}_{s-1:s+K-1}, \dots, \alpha^{n,M}_{s-1:s+K-1})$, and sample one path, $\alpha^n_{s:t}$, from $\alpha^{n,1:M}_{s:s+K-1}$ as below.

(1) Sample indices k_j from $1:M$ with probability $1/M$ ($j = s+K-1, s+K-2, \dots, s-1$) and set

$$(\alpha^{n,k_{s-1}}_{s-1}, \dots, \alpha^{n,k_{s+K-1}}_{s+K-1}) = \alpha^n_{s-1:s+K-1}, \quad (a^{k_{s-2}}_{s-1}, \dots, a^{k_{s+K-2}}_{s+K-1}) = (k_{s-1}, \dots, k_{s+K-1}),$$

where $\alpha^n_{s-1:s+K-1}$ is a current sample with the importance weight $W^n_{s-1:t}$.

(2) Set $\alpha^{n,a^m_{s+K}}_{s+K} = \alpha^n_{s+K}$ for all m according to the convention, and sample $\alpha^{n,m}_{s+K-1} \sim q_{s+K-1,\theta^n}(\cdot \mid \alpha^n_{s+K}, y_{s+K-1})$ for each $m \in \{1, \dots, M\} \setminus \{k_{s+K-1}\}$. Let $j = s+K-2$.

(3) Sample $a^m_{j+1} \sim \mathcal{M}(V^{1:M}_{j+1,\theta^n})$ and $\alpha^{n,m}_j \sim q_{j,\theta^n}(\cdot \mid \alpha^{n,a^m_{j+1}}_{j+1}, y_j)$ for each $m \in \{1, \dots, M\} \setminus \{k_j\}$ where $V^{1:M}_{j+1,\theta^n} = (V^1_{j+1,\theta^n}, \dots, V^M_{j+1,\theta^n})$ and

$$V^m_{j,\theta^n} = \frac{v_{j,\theta^n}(\alpha^{n,m}_j, \alpha^{n,a^m_{j+1}}_{j+1})}{\sum_{i=1}^M v_{j,\theta^n}(\alpha^{n,i}_j, \alpha^{n,a^i_{j+1}}_{j+1})}, \tag{23}$$

$$v_{j,\theta^n}(\alpha^{n,m}_j, \alpha^{n,a^m_{j+1}}_{j+1}) = \frac{p(\alpha^{n,m}_j \mid \alpha^{n,a^m_{j+1}}_{j+1}, \theta) g_{\theta^n}(y_j \mid \alpha^{n,m}_j, \alpha^{n,a^m_{j+1}}_{j+1})}{q_{j,\theta^n}(\alpha^{n,m}_j \mid \alpha^{n,a^m_{j+1}}_{j+1}, y_j)}, \quad m = 1, \dots, M. \tag{24}$$

(4) If $j > s-1$, set $j \leftarrow j-1$ and go to (3). Otherwise, sample $k^*_s \sim \mathcal{M}(V^{1:M}_{s,\theta^n})$ and obtain k^*_j ($j = s+1, \dots, s+K-1$) using (22).

(5) Let $\alpha^n_{s:t} = (\alpha^{n,k^*_s}_s, \dots, \alpha^{n,k^*_{s+K-1}}_{s+K-1}, \alpha^n_{s+K}, \dots, \alpha^n_t)$ and compute its importance weight

$$W^n_{s:t} \propto \begin{cases} \frac{1}{\hat{p}(y_{s-1} \mid y_{s:t}, \alpha^n_{s+K}, \theta^n)} W^n_{s-1:t}, & \text{if } \hat{p}(y_{s-1} \mid y_{s:t}, \alpha^n_{s+K}, \theta^n) \neq 0, \\ 0, & \text{if } \hat{p}(y_{s-1} \mid y_{s:t}, \alpha^n_{s+K}, \theta^n) = 0, \end{cases} \tag{25}$$

where

$$\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}^n, \theta^n) \quad = \quad \frac{1}{M} \sum_{m=1}^{M} v_{s-1,\theta^n}(\alpha_{s-1}^{n,m}, \alpha_s^{n,a_s^m}). \tag{26}$$

(6) Implement the particle simulation smoother to sample $(k_s^*, k_{s+1}^*, \ldots, k_{s+K-1}^*)$ jointly. Generate $k_j^* \sim \mathcal{M}(\bar{V}_{j,\theta^n}^{1:M})$, $j = s+1, \ldots, s+K-1$, recursively where

$$\bar{V}_{j,\theta^n}^m = \frac{V_{j,\theta^n}^m p(\alpha_{j-1}^{k_{j-1}^*} \mid \alpha_j^m, \theta^n)}{\sum_{i=1}^M V_{j,\theta^n}^i p(\alpha_{j-1}^{k_{j-1}^*} \mid \alpha_j^i, \theta^n)}, \quad m = 1, \ldots, M. \tag{27}$$

and set $\alpha_{s:t}^n = (\alpha_s^{n,k_s^*}, \ldots, \alpha_{s+K-1}^{n,k_{s+K-1}^*}, \alpha_{s+K}^n, \ldots, \alpha_t^n)$.

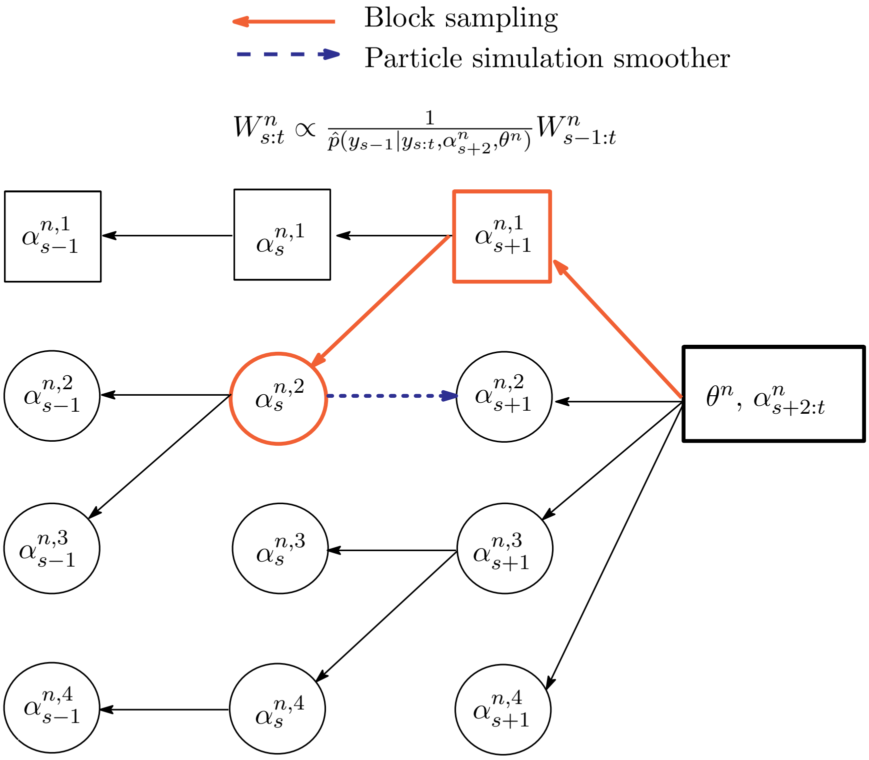

Figure 5: Backward block sampling and particle simulation smoother. $K = 2$, $M = 4$. $(\theta_n, \alpha_{s-1:t}^n)$ is fixed and shown in the rectangle ($k_{s-1:s+1} = \{1,1,1\}$). Other state variables in the circle are generated. $k_{s:s+1}^* = \{2,1\}$ in the block sampling. $k_{s:s+1}^* = \{2,2\}$ in the particle simulation smoother.

Figure 5 illustrates an example with $K = 2$ and $M = 4$. The current sample is $(\theta^n, \alpha_{s-1:t}^n)$ and we set $\alpha_{s-1}^{n,1} = \alpha_{s-1}^n$, $\alpha_s^{n,1} = \alpha_s^n$ and $\alpha_{s+1}^{n,1} = \alpha_{s+1}^n$ with $k_{s-1:s+1} = \{1,1,1\}$.

They are fixed and shown in the rectangle. Other state variables are generated as above. In the block sampling, $M = 4$ paths are generated and $k_s^* = 2$ is selected (and hence $k_{s+1}^* = 1$). By implementing the particle simulation smoother, $k_{s+1}^* = 2$ is selected, and the updated particle is $(\theta^n, \alpha_{s:t}^n)$ with the importance weight $W_{s:t}^n$ where $\alpha_s^n = \alpha_s^{n,2}$ and $\alpha_{s+1}^n = \alpha_{s+1}^{n,2}$.

In Step (4), one randomly determines the index k_s^* at time s, not at time $s-1$, to discard the old information of y_{s-1} in the new particle path. Also, in Step (5), we discount the importance weight to remove its old information by dividing by the conditional likelihood of y_{s-1}. A major difference from the incremental weight in (19) is that it is the inverse of the estimator of the conditional likelihood in (25). If $\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}^n, \theta^n) = 0$, then $v_{s-1,\theta^n}(\alpha_{s-1}^{n,m}, \alpha_s^{n,a_s^m}) = 0$ for all m, which implies no candidate proposed by q_{s-1,θ^n} falls in the support of the target density. The probability of such an event is would be very small unless the proposed density poorly approximates the target density, since the support of q_{s-1,θ^n} is chosen to include that of the target density in the important sampling scheme. We shall show that the backward block sampling is highly efficient, and its ESS is much higher than that of the simple particle rolling MCMC in illustrative examples in Section 5.

3.3 Initializing the rolling estimation

The above discussion has focused on how to update the weighted particles which approximate $\pi(\theta, \alpha_{s-1:t-1} \mid y_{s-1:t-1})$ to obtain the particles which approximate $\pi(\theta, \alpha_{s:t} \mid y_{s:t})$. It is implicitly assumed that the particles approximating $\pi(\theta, \alpha_{s-1:t-1} \mid y_{s-1:t-1})$ are at hand. In order to sample from the initial posterior distribution, it is straightforward to use MCMC methods as in the warm-up period for the practical filtering in Polson et al. (2008). However, SMC-based methods are preferred when we need to compute the marginal likelihood $p(y_{1:s-1})$ as in Section 3.4. We describe how to initialize the particle rolling MCMC below to sample $\alpha_{1:L+1}$ where $L = t - s$.

(A3) Initializing the particle rolling MCMC

Step 1 Sample (θ^n, α_1^n) from $\pi(\theta, \alpha_1 \mid y_1)$ for $n = 1, \ldots, N$.

1a Sample $\theta^n \sim p(\theta)$.

1b Sample $\alpha_1^{n,m} \sim q_{1,\theta^n}(\cdot \mid y_1)$ for each $m \in \{1, \ldots, M\}$.

1c Sample $k_1 \sim \mathcal{M}(V_{1,\theta^n}^{n,1:M})$ where

$$V_{1,\theta^n}^{n,m} = \frac{v_{1,\theta^n}(\alpha_1^{n,m})}{\sum_{i=1}^M v_{1,\theta^n}(\alpha_1^{n,i})}, \quad v_{1,\theta^n}(\alpha_1^{n,m}) = \frac{\mu_{\theta^n}(\alpha_1^{n,m}) g_{\theta^n}(y_1 \mid \alpha_1^{n,m})}{q_{1,\theta^n}(\alpha_1^{n,m} \mid y_1)}. \quad (28)$$

1d Set $\alpha_1^n = \alpha_1^{n,k_1}$ and store (θ^n, α_1^n) with its importance weight

$$W_1^n \quad \propto \quad \hat{p}(y_1 \mid \theta^n), \quad \hat{p}(y_1 \mid \theta^n) = \sum_{m=1}^{M} v_{1,\theta^n}(\alpha_1^{n,m}). \tag{29}$$

Step 2 For $j = 2, \ldots, L+1$,

2a Implement Step 1a of the forward block sampling to generate $\alpha_{1:j}^n$ and θ^n, and compute its importance weight

$$W_j^n \quad \propto \quad \hat{p}(y_j \mid y_{1:j-1}, \alpha_{j-K-1}^n, \theta^n) \times W_{j-1}^n, \tag{30}$$

where

$$\hat{p}(y_j \mid y_{1:j-1}, \alpha_{j-K-1}^n, \theta^n) = \frac{1}{M} \sum_{m=1}^{M} v_{j,\theta^n}(\alpha_{j-1}^{n,a_{j-1}^{n,m}}, \alpha_j^{n,m}). \tag{31}$$

For $j < K$, we set $K = j-1$, and all particles of $\alpha_{1:j}^n$ are resampled.

2b If some degeneracy criterion is fulfilled, resample all particles and set $W_j^n = 1/N$. Further, update particles $(\theta^n, \alpha_{1:j}^n)$ using the MCMC kernel of the invariant distribution with the density $\pi(\theta, \alpha_{1:j} \mid y_{1:j})$.

It is clear that this new algorithm can be applied to the ordinary sequential learning of $\pi(\theta, \alpha_{1:t} \mid y_{1:t})$ $(t = 1, \ldots, T)$. The algorithm proposes a new approach applying the particle MCMC scheme in Andrieu et al. (2010), especially particle Gibbs, which is different from SMC^2 applying particle MH in Chopin et al. (2013) and Fulop and Li (2013).

Remark 3. Especially when j is small and the dimension of $\alpha_{1:j}$ is smaller than that of θ, the MCMC update of θ could lead to unstable estimation results. We may need to modify the MCMC kernel or skip the update in such a case.

3.4 Estimation of the marginal likelihood

As a by-product of the proposed algorithms, we can obtain the estimate of the marginal likelihood defined as

$$p(y_{s:t}) \quad = \quad \int p(y_{s:t} \mid \alpha_{s:t}, \theta) p(\alpha_{s:t} \mid \theta) p(\theta) d\alpha_{s:t} d\theta, \tag{32}$$

so that it is used to compute Bayes factors for model comparison. Since it is expressed as

$$p(y_{s:t}) \quad = \quad \frac{p(y_t \mid y_{s-1:t-1})}{p(y_{s-1} \mid y_{s:t})} p(y_{s-1:t-1}), \tag{33}$$

we obtain the estimate $\hat{p}(y_{s:t})$ recursively by

$$\hat{p}(y_{s:t}) = \frac{\hat{p}(y_t \mid y_{s-1:t-1})}{\hat{p}(y_{s-1} \mid y_{s:t})}\hat{p}(y_{s-1:t-1}), \tag{34}$$

where

$$\hat{p}(y_t \mid y_{s-1:t-1}) = \sum_{n=1}^{N} W_{s-1:t-1}^{n}\hat{p}(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}^{n}, \theta^n), \tag{35}$$

$$\hat{p}(y_{s-1} \mid y_{s:t}) = \sum_{n=1}^{N} W_{s-1:t}^{n}\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}^{n}, \theta^n). \tag{36}$$

using (19), (20), (25) and (26). The initial estimate $\hat{p}(y_{1:L+1})$, $L = t - s$ is given by

$$\hat{p}(y_{1:L+1}) = \hat{p}(y_1)\prod_{j=2}^{L+1}\hat{p}(y_j \mid y_{1:j-1}), \tag{37}$$

where

$$\hat{p}(y_1) = \sum_{n=1}^{N}\hat{p}(y_1 \mid \theta^n), \tag{38}$$

$$\hat{p}(y_j \mid y_{1:j-1}) = \sum_{n=1}^{N} W_{j-1}^{n}\hat{p}(y_j \mid y_{1:j-1}, \alpha_{j-K-1}^{n}, \theta^n), \tag{39}$$

using (29), (30) and (31).

4 Theoretical justification of particle rolling MCMC with double block sampling

Theoretical justifications of double (forward and backward) block sampling with the conditional SMC update proposed in Section 3 are provided. We prove that our posterior density is obtained as a marginal density of the artificial target density.

4.1 Forward block sampling

The artificial target density and its marginal density. We prove that our posterior density of $(\alpha_{s-1:t}^n, \theta^n)$ given $y_{s-1:t}$ is obtained as a marginal density of the artificial target density in the forward block sampling. The superscript n will be suppressed for simplicity below.

In Step 1a-(1) of Section 3.1, the probability density function of $(\alpha_{t-K}^{k_{t-K}}, \ldots, \alpha_{t-1}^{k_{t-1}}) = \alpha_{t-K:t-1}$ and $(a_{t-K}^{k_{t-K+1}}, \ldots, a_{t-1}^{k_t})$ given (α_{t-K-1}, θ) and $y_{t-K:t-1}$ is

$$p(\alpha_{t-K:t-1}, a_{t-K}^{k_{t-K+1}}, \ldots, a_{t-1}^{k_t} \mid \alpha_{t-K-1}, y_{t-K:t-1}, \theta) = \frac{\pi(\alpha_{t-K:t-1} \mid \alpha_{t-K-1}, y_{t-K:t-1}, \theta)}{M^K}. \tag{40}$$

Let $a_j^{1:M} = (a_j^1, \ldots, a_j^M)$ and $a_j^{-k_{j+1}} \equiv a_j^{1:M} \backslash a_j^{k_{j+1}} = a_j^{1:M} \backslash k_j$ for $j = t-K, \ldots, t-1$ where we note $a_j^{k_{j+1}} = k_j$ and $k_t = 1$ in (15). Further, let $a_{t-K:t-1}^{1:M} = \{a_{t-K}^{1:M}, \ldots, a_{t-1}^{1:M}\}$, and $\alpha_j^{-k_j} = \{\alpha_j^{a_j^1}, \ldots, \alpha_j^{a_j^M}\} \setminus \alpha_j^{k_j}$. Then, in Steps 1a-(2)(3)(4), given α_{t-K-1}, $(\alpha_{t-K}^{k_{t-K}}, \ldots, \alpha_{t-1}^{k_{t-1}}) = \alpha_{t-K:t-1}$ and $(a_{t-K}^{k_{t-K+1}}, \ldots, a_{t-1}^{k_t}) = (k_{t-K}, \ldots, k_{t-1})$, the probability density function of all variables is defined as

$$
\begin{aligned}
&\psi_\theta\left(\alpha_{t-K}^{-k_{t-K}}, \ldots, \alpha_{t-1}^{-k_{t-1}}, \alpha_t^{1:M}, a_{t-K}^{-k_{t-K+1}}, \ldots, a_{t-1}^{-k_t}, k_t^* \mid \alpha_{t-K-1:t-1}, a_{t-K}^{k_{t-K+1}}, \ldots, a_{t-1}^{k_t}, y_{t-K:t}\right) \\
&\quad = \prod_{\substack{m=1 \\ m \neq k_{t-K}}}^{M} q_{t-K,\theta}(\alpha_{t-K}^m \mid \alpha_{t-K-1}, y_{t-K}) \times \prod_{j=t-K+1}^{t-1} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j-1,\theta}^{a_{j-1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_{j-1}^m}, y_j) \\
&\qquad \times q_{t,\theta}(\alpha_t^1 \mid \alpha_{t-1}^{k_{t-1}}, y_t) \times \prod_{m=2}^{M} V_{t-1,\theta}^{a_{t-1}^m} q_{t,\theta}(\alpha_t^m \mid \alpha_{t-1}^{a_{t-1}^m}, y_t) \times V_{t,\theta}^{k_t^*}. \qquad (41)
\end{aligned}
$$

In Step 1a-(5), we multiply $W_{s-1:t-1}$ by $\hat{p}(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}^n, \theta^n)$ to adjust the importance weight for $W_{s-1:t}$. Thus our artificial target density is written as

$$
\begin{aligned}
&\hat{\pi}(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K:t}^{1:M}, a_{t-K:t-1}^{1:M}, k_t^* \mid y_{s-1:t}) \\
&\equiv \frac{\pi(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{k_{t-K}}, \ldots, \alpha_{t-1}^{k_{t-1}} \mid y_{s-1:t-1})}{M^K} \\
&\quad \times \psi_\theta(\alpha_{t-K}^{-k_{t-K}}, \ldots, \alpha_{t-1}^{-k_{t-1}}, \alpha_t^{1:M}, a_{t-K}^{-k_{t-K+1}}, \ldots, a_{t-1}^{-k_t}, k_t^* \mid \alpha_{t-K-1:t-1}, a_{t-K}^{k_{t-K+1}}, \ldots, a_{t-1}^{k_t}, y_{t-K:t}) \\
&\quad \times \frac{\hat{p}(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta)}{p(y_t \mid y_{s-1:t-1})} \\
&= \frac{\pi(\theta, \alpha_{s-1:t-1} \mid y_{s-1:t-1})}{M^K} \\
&\quad \times \prod_{\substack{m=1 \\ m \neq k_{t-K}}}^{M} q_{t-K,\theta}(\alpha_{t-K}^m \mid \alpha_{t-K-1}, y_{t-K}) \times \prod_{j=t-K+1}^{t-1} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j-1,\theta}^{a_{j-1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_{j-1}^m}, y_j) \\
&\quad \times \quad q_{t,\theta}(\alpha_t^1 \mid \alpha_{t-1}^{k_{t-1}}, y_t) \times \prod_{m=2}^{M} V_{t-1,\theta}^{a_{t-1}^m} q_{t,\theta}(\alpha_t^m \mid \alpha_{t-1}^{a_{t-1}^m}, y_t) \times V_{t,\theta}^{k_t^*} \\
&\quad \times \frac{\hat{p}(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta)}{p(y_t \mid y_{s-1:t-1})}. \qquad (42)
\end{aligned}
$$

Note that $p(y_t \mid y_{s-1:t-1})$ is the normalizing constant of this target density, which will be shown in Proposition 4.2. The proposed forward block sampling is justified by proving that the marginal density of $(\theta, \alpha_{s-1}, \ldots, \alpha_{t-K-1}, \alpha_{t-K}^{k_{t-K}^*}, \ldots, \alpha_t^{k_t^*})$ in the above artificial target density $\hat{\pi}$ is $\pi(\theta, \alpha_{s-1}, \ldots, \alpha_{t-K-1}, \alpha_{t-K}^{k_{t-K}^*}, \ldots, \alpha_t^{k_t^*} \mid y_{s-1:t})$.

We first establish the following lemma which describes a property of the local conditional SMC.

Lemma 4.1. *For any t and t_0 $(t-K \le t_0 \le t)$,*

$$\frac{\pi(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{k_{t-K}}, \ldots, \alpha_{t_0}^{k_{t_0}} \mid y_{s-1:t_0})}{M^{t_0-(t-K)+1}}$$
$$\times \prod_{\substack{m=1 \\ m \neq k_{t-K}}}^{M} q_{t-K,\theta}(\alpha_{t-K}^m \mid \alpha_{t-K-1}, y_{t-K}) \times \prod_{j=t-K+1}^{t_0} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j-1,\theta}^{a_{j-1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_{j-1}^m}, y_j) \tag{43}$$
$$= \pi(\theta, \alpha_{s-1:t-K-1} \mid y_{s-1:t-K-1}) \times \prod_{m=1}^{M} q_{t-K,\theta}(\alpha_{t-K}^m \mid \alpha_{t-K-1}, y_{t-K})$$
$$\times \prod_{j=t-K+1}^{t_0} \prod_{m=1}^{M} V_{j-1,\theta}^{a_{j-1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_{j-1}^m}, y_j) \times V_{t_0,\theta}^{k_{t_0}} \times \prod_{j=t-K}^{t_0} \frac{\hat{p}(y_j \mid y_{s-1:j-1}, \alpha_{t-K-1}, \theta)}{p(y_j \mid y_{s-1:j-1})}, \tag{44}$$

where

$$\hat{p}(y_j \mid y_{s-1:j-1}, \alpha_{t-K-1}, \theta) = \frac{1}{M} \sum_{m=1}^{M} v_{j,\theta}(\alpha_{j-1}^{a_{j-1}^m}, \alpha_j^m), \quad j = t-K, \ldots, t_0, \tag{45}$$

with $\alpha_{t-K-1}^{a_{t-K-1}^m} = \alpha_{t-K-1}$ *and* $a_{j-1}^{k_j} = k_{j-1}$, $j = t-K+1, \ldots, t_0$.

Proof. See Appendix A.1.

The probability density (43) corresponds to the target density π_t^* of SMC^2 in Chopin et al. (2013) which includes the random particle index. For the particle filtering, the forward block sampling considers the density of $\alpha_{t-K:t-1}^{1:M}$ conditional on $(\theta, \alpha_{s-1:t-K-1})$, while SMC^2 considers that of $\alpha_{1:t}^{1:M}$ conditional on θ. Further, the former updates the importance weight for $(\theta, \alpha_{s-1:t})$ and the latter updates that for θ sequentially. Using Lemma 4.1, we obtain the following Proposition.

Proposition 4.1. *The artificial target density $\hat{\pi}$ for the forward block sampling can be written as*

$$\hat{\pi}(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K:t}^{1:M}, a_{t-K:t-1}^{1:M}, k_t^* \mid y_{s-1:t})$$
$$= \frac{\pi(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{k_{t-K}^*}, \ldots, \alpha_t^{k_t^*} \mid y_{s-1:t})}{M^{K+1}} \times \prod_{\substack{m=1 \\ m \neq k_{t-K}^*}}^{M} q_{t-K,\theta}(\alpha_{t-K}^m \mid \alpha_{t-K-1}, y_{t-K})$$
$$\times \prod_{j=t-K+1}^{t} \prod_{\substack{m=1 \\ m \neq k_j^*}}^{M} V_{j-1,\theta}^{a_{j-1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_{j-1}^m}, y_j), \tag{46}$$

and the marginal density of $(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{k_{t-K}^*}, \ldots, \alpha_t^{k_t^*})$ *is* $\pi(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{k_{t-K}^*}, \ldots, \alpha_t^{k_t^*} \mid y_{s-1:t})$.

Proof. See Appendix A.2.

Proposition 4.1 implies that we can obtain a posterior random sample $(\theta, \alpha_{s-1:t})$ given $y_{s-1:t}$ (with the importance weight $W_{s-1:t}$) by sampling from the artificial target distribution $\hat{\pi}$. This justifies our proposed forward block sampling scheme.

Remark 4. We note that k_j's do not appear in (46). In practice, k_j's can be determined arbitrary, e.g. $k_j = 1$ $(j = t-K, \ldots, t-1)$.

Properties of the incremental weight. We consider the mean and variance of the (unnormalized) incremental weight, $\hat{p}(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta)$. Proposition 4.2 shows that it is unbiased.

Proposition 4.2. *If*

$$(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{k_{t-K}}, \ldots, \alpha_{t-1}^{k_{t-1}}, k_{t-K:t-1}) \sim \frac{\pi(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{k_{t-K}}, \ldots, \alpha_{t-1}^{k_{t-1}} \mid y_{s-1:t-1})}{M^K}$$

and

$$(\alpha_{t-K}^{-k_{t-K}}, \ldots, \alpha_t^{-k_{t-1}}, \alpha_t^{1:M}, a_{t-K}^{-k_{t-K+1}}, \ldots, a_{t-1}^{-k_t}, k_t^*) \quad \sim \quad \psi_\theta$$

where ψ_θ is given in (41), then

$$\begin{aligned} E[\hat{p}(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta)|y_{s-1:t}] &= E[p(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta)|y_{s-1:t}, \alpha_{t-K-1}, \theta] \\ &= p(y_t \mid y_{s-1:t-1}). \end{aligned}$$

Proof. See Appendix A.3.

This shows the incremental weight $\hat{p}(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta)$ is an unbiased estimator of the conditional likelihood $p(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta)$ given (α_{t-K-1}, θ). It is also an unbiased estimator of the marginal likelihood $p(y_t \mid y_{s-1:t-1})$ unconditionally, which implies that $p(y_t \mid y_{s-1:t-1})$ is a normalizing constant for the artificial target density $\hat{\pi}$.

Further, from the law of total variance, we have the decomposition of the variance as follows.

$$\begin{aligned} &\mathrm{Var}[\hat{p}(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta) \mid y_{s-1:t}] \\ &= \mathrm{Var}[p(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta) \mid y_{s-1:t}] \\ &\qquad + E\left[\mathrm{Var}[\hat{p}(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta) \mid y_{s-1:t}, \alpha_{s-1:t-K-1}, \theta]\right]. \end{aligned}$$

The variance of the incremental weight consists of two components: variance of the conditional likelihood and (expected) variance which is introduced by using M particles to approximate the conditional likelihood. This decomposition suggests what factor influences the ESS of the particles. As for the first component, for any positive integers, K_1, K_2, with $K_1 < K_2$, the following inequality holds:

$$\mathrm{Var}[p(y_t \mid y_{s-1:t-1}, \alpha_{t-K_1-1}, \theta)] \geq \mathrm{Var}[p(y_t \mid y_{s-1:t-1}, \alpha_{t-K_2-1}, \theta)],$$

which is a straightforward result from the law of total variance for $p(y_t \mid y_{s-1:t-1}, \alpha_{t-K_1-1}, \theta)$ using

$$E\left[p(y_t \mid y_{s-1:t-1}, \alpha_{t-K_1-1}, \theta) \mid \alpha_{s-1:t-K_2-1}, \theta\right] = p(y_t \mid y_{s-1:t-1}, \alpha_{t-K_2-1}, \theta). \tag{47}$$

On the other hand, the second component is expected to be controlled by changing the number of particles M. In Section 5, we investigate how K and M affect the variance of incremental weights in practice and show that large K and M actually reduce the variance in each step of sampling.

Particle simulation smoother. In Whiteley et al. (2010) and the discussion of Whiteley following Andrieu et al. (2010), the additional step is introduced to explore all possible ancestral lineages. This is expected to circumvent the weight degeneracy phenomenon and to improve the mixing property of Particle Gibbs, which is also found to be effective in the numerical experiment in Chopin and Singh (2015). We also incorporate such a particle simulation smoother into the double block sampling based on the following lemma.

Proposition 4.3. *The joint conditional density of* $(k^*_{t-K}, \ldots, k^*_t)$ *is given by*

$$\begin{aligned}
&\hat{\pi}(k^*_{t-K}, \ldots, k^*_t | \theta, \alpha_{s-1:t-K-1}, \alpha^{1:M}_{t-K:t}, a^{1:M}_{t-K:t-1}, y_{s-1:t}) \\
&\quad = \hat{\pi}(k^*_t | \theta, \alpha_{s-1:t-K-1}, \alpha^{1:M}_{t-K:t}, a^{1:M}_{t-K:t-1}, y_{s-1:t}) \\
&\quad \times \prod_{t_0=t-1}^{t-K} \hat{\pi}(k^*_{t_0} | \theta, \alpha_{s-1:t-K-1}, \alpha^{1:M}_{t-K:t_0}, a^{1:M}_{t-K:t_0-1}, \alpha^{k^*_{t_0+1}}_{t_0+1}, \ldots, \alpha^{k^*_t}_t, k^*_{t_0+1:t}, y_{s-1:t}),
\end{aligned} \tag{48}$$

where

$$\begin{aligned}
&\hat{\pi}(k^*_{t_0} | \theta, \alpha_{s-1:t-K-1}, \alpha^{1:M}_{t-K:t_0}, a^{1:M}_{t-K:t_0-1}, \alpha^{k^*_{t_0+1}}_{t_0+1}, \ldots, \alpha^{k^*_t}_t, k^*_{t_0+1:t}, y_{s-1:t}) \\
&\quad = \bar{V}^{k^*_{t_0}}_{t_0,\theta}, \quad \bar{V}^m_{j,\theta} \equiv \frac{V^m_{j,\theta} f_\theta(\alpha^{k^*_{j+1}}_{j+1} \mid \alpha^m_j, y_{j+1})}{\sum_{i=1}^M V^i_{j,\theta} f_\theta(\alpha^{k^*_{j+1}}_{j+1} \mid \alpha^i_j, y_{j+1})}.
\end{aligned} \tag{49}$$

Proof. See Appendix A.4.

Suppose we have $(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K:t}^{1:M}, a_{t-K:t-1}^{1:M}, k_t^*) \sim \hat{\pi}$ where $\hat{\pi}$ is defined in (42). In Step 1a-(4), the lineage $k_{t-K:t}^*$ is automatically determined when k_t^* is chosen. The particle simulation smoother breaks this relationship and again samples $k_{t-K:t}^*$ jointly by generating $k_j^* \sim \mathcal{M}(\bar{V}_{j,\theta}^{1:M})$, $j = t-1, \ldots, t-K$, recursively.

4.2 Backward block sampling

The artificial target density and its marginal density. This subsection proves that our posterior density of $(\alpha_{s:t}^n, \theta^n)$ given $y_{s:t}$ is obtained as a marginal density of the artificial target density in the backward block sampling. The superscript n will be suppressed for simplicity below.

In Step 2a-(1), the probability density function of $(\alpha_{s-1}^{k_{s-1}}, \ldots, \alpha_{s+K-1}^{k_{s+K-1}}) = \alpha_{s-1:s+K-1}$ and $(a_{s-1}^{k_{s-2}}, \ldots, a_{s+K-1}^{k_{s+K-2}})$ given (α_{s+K}, θ) and $y_{s-1:t}$ is

$$p(\alpha_{s-1:s+K-1}, a_{s-1}^{k_{s-2}}, \ldots, a_{s+K-1}^{k_{s+K-2}} \mid \alpha_{s+K}, \theta, y_{s-1:t}) = \frac{\pi(\alpha_{s-1:s+K-1} \mid \alpha_{s+K}, y_{s-1:t}, \theta)}{M^{K+1}}. \tag{50}$$

In Steps 2a-(2)(3)(4), given α_{s+K}, $(\alpha_{s-1}^{k_{s-1}}, \ldots, \alpha_{s+K-1}^{k_{s+K-1}}) = \alpha_{s-1:s+K-1}$, $(a_{s-1}^{k_{s-2}}, \ldots, a_{s+K-1}^{k_{s+K-2}}) = (k_{s-1}, \ldots, k_{s+K-1})$ and $y_{s-1:s+K-1}$, the probability density function of all variables is defined as

$$\begin{aligned}
&\bar{\psi}_\theta(\alpha_{s-1}^{-k_{s-1}}, \ldots, \alpha_{s+K-1}^{-k_{s+K-1}}, a_s^{-k_{s-1}}, \ldots, a_{s+K-1}^{-k_{s+K-2}}, k_s^* \mid \alpha_{s-1:s+K}, a_{s-1}^{k_{s-2}}, \ldots, a_{s+K-1}^{k_{s+K-2}}, y_{s-1:s+K-1}) \\
&= \prod_{\substack{m=1 \\ m \neq k_{s+K-1}}}^{M} q_{s+K-1,\theta}(\alpha_{s+K-1}^m \mid \alpha_{s+K}, y_{s+K-1}) \times \prod_{j=s-1}^{s+K-2} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j+1}^{a_{j+1}^m}, y_j) \times V_{s,\theta}^{k_s^*}.
\end{aligned} \tag{51}$$

In Step2a-(5), we divide $W_{s-1:t}$ by $\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}^n, \theta^n)$ to adjust the importance weight for $W_{s:t}$. Similarly to the discussion in Section 4.1, we consider an extended space with the artificial target density written as

$$\begin{aligned}
&\check{\pi}(\theta, \alpha_{s-1:s+K-1}^{1:M}, \alpha_{s+K:t}, a_{s:s+K-1}^{1:M}, k_{s-1}, k_s^* \mid y_{s-1:t}) \\
&\equiv \frac{\pi(\theta, \alpha_{s-1:t} \mid y_{s-1:t})}{M^{K+1}} \\
&\times \prod_{\substack{m=1 \\ m \neq k_{s+K-1}}}^{M} q_{s+K-1,\theta}(\alpha_{s+K-1}^m \mid \alpha_{s+K}, y_{s+K-1}) \times \prod_{j=s-1}^{s+K-2} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j+1}^{a_{j+1}^m}, y_j) \\
&\times V_{s,\theta}^{k_s^*} \times \frac{p(y_{s-1} \mid y_{s:t})}{\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)},
\end{aligned} \tag{52}$$

where $p(y_{s-1} \mid y_{s:t})^{-1}$ is the normalizing constant of this target density as shown in Proposition 4.5.

Below we prove Lemma 4.2 and Proposition 4.4 for the backward block sampling, which correspond to Lemma 4.1 and Proposition 4.1 for the forward block sampling.

Lemma 4.2. *For any t, s_0 ,and s $(s-1 \le s_0 \le s+K-1)$,*

$$\begin{aligned}
&\frac{\pi(\theta, \alpha_{s_0}^{k_{s_0}}, \ldots, \alpha_{s+K-1}^{k_{s+K-1}}, \alpha_{s+K:t} \mid y_{s_0:t})}{M^{(s+K-1)-s_0+1}} \\
&\quad \times \prod_{\substack{m=1 \\ m \neq k_{s+K-1}}}^{M} q_{s+K-1,\theta}(\alpha_{s+K-1}^m \mid \alpha_{s+K}, y_{s+K-1}) \times \prod_{j=s_0}^{s+K-2} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j+1}^{a_{j+1}^m}, y_j) \\
&= \pi(\theta, \alpha_{s+K:t} \mid y_{s+K:t}) \times \prod_{m=1}^{M} q_{s+K-1,\theta}(\alpha_{s+K-1}^m \mid \alpha_{s+K}, y_{s+K-1}) \\
&\quad \times \prod_{j=s_0}^{s+K-2} \prod_{m=1}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j+1}^{a_{j+1}^m}, y_j) \times V_{s_0,\theta}^{k_{s_0}} \times \prod_{j=s_0}^{s+K-1} \frac{\hat{p}(y_j \mid y_{j+1:t}, \alpha_{s+K}, \theta)}{p(y_j \mid y_{j+1:t})}
\end{aligned}$$

with $\alpha_{s+K}^{a_{s+K}^m} = \alpha_{s+K}$ and $a_{j+1}^{k_j} = k_{j+1}$ $(s_0 \le j \le s+K-2)$, where

$$\hat{p}(y_j \mid y_{j+1:t}, \alpha_{s+K}, \theta) \quad = \quad \frac{1}{M} \sum_{m=1}^{M} v_{j,\theta}(\alpha_j^m, \alpha_{j+1}^{a_{j+1}^m}). \tag{53}$$

Proof. See Appendix A.5.

Proposition 4.4. *The artificial target density $\breve{\pi}$ for the backward block sampling can be rewritten as*

$$\begin{aligned}
&\breve{\pi}(\theta, \alpha_{s-1:s+K-1}^{1:M}, \alpha_{s+K:t}, a_{s:s+K-1}^{1:M}, k_{s-1}, k_s^* \mid y_{s-1:t}) \\
&\quad = \frac{\pi(\theta, \alpha_s^{k_s^*}, \ldots, \alpha_{s+K-1}^{k_{s+K-1}^*}, \alpha_{s+K:t} \mid y_{s:t})}{M^K} \times \prod_{\substack{m=1 \\ m \neq k_{s+K-1}^*}}^{M} q_{s+K-1,\theta}(\alpha_{s+K-1}^m \mid \alpha_{s+K}, y_{s+K-1:t}) \\
&\quad \times \prod_{j=s}^{s+K-2} \prod_{\substack{m=1 \\ m \neq k_j^*}}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j+1}^{a_{j+1}^m}, y_j) \times \prod_{m=1}^{M} V_{s,\theta}^{a_s^m} q_{s-1,\theta}(\alpha_{s-1}^m \mid \alpha_s^{a_s^m}, y_{s-1}) \times V_{s-1,\theta}^{k_{s-1}},
\end{aligned} \tag{54}$$

and the marginal density of $(\theta, \alpha_s^{k_s^*}, \ldots, \alpha_{s+K-1}^{k_{s+K-1}^*}, \alpha_{s+K:t})$ is $\pi(\theta, \alpha_s^{k_s^*}, \ldots, \alpha_{s+K-1}^{k_{s+K-1}^*}, \alpha_{s+K:t} \mid y_{s:t})$.

Proof. See Appendix A.6.

Although the probability density (54) in Proposition 4.4 has a bit different form from that of (46) in Proposition 4.1, its marginal probability density is found to be the target posterior density $\pi(\theta, \alpha_{s:t} \mid y_{s:t})$.

Properties of the incremental weight. Similar results to Proposition 4.2 hold for the backward block sampling, and are summarized in Proposition 4.5.

Proposition 4.5. *If*

$$(\theta, \alpha_{s-1}^{k_{s-1}}, \ldots, \alpha_{s+K-1}^{k_{s+K-1}}, \alpha_{s+K:t}, k_{s-1:s+K-1}) \sim \frac{\pi(\theta, \alpha_{s-1}^{k_{s-1}}, \ldots, \alpha_{s+K-1}^{k_{s+K-1}}, \alpha_{s+K:t} \mid y_{s-1:t})}{M^{K+1}}$$

and

$$(\alpha_{s-1}^{-k_{s-1}}, \ldots, \alpha_{s+K-1}^{-k_{s+K-1}}, a_s^{-k_{s-1}}, \ldots, a_{s+K-1}^{-k_{s+K-2}}, k_s^*) \quad \sim \quad \bar{\psi}_\theta,$$

where $\bar{\psi}_\theta$ *is given in (51), then*

$$\begin{aligned} E[\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)^{-1}] &= E[\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)^{-1} \mid y_{s:t}, \alpha_{s+K}, \theta] \\ &= p(y_{s-1} \mid y_{s:t}, \theta)^{-1}. \end{aligned}$$

Proof. See Appendix A.7.

The decomposition of the variance of the incremental weight $\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)^{-1}$ follows from the law of total variance as in Section 4.1 :

$$\begin{aligned} &\mathrm{Var}[\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)^{-1}] \\ &= \mathrm{Var}[p(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)^{-1}] + E[\mathrm{Var}[\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)^{-1} \mid \alpha_{s+K:t}, \theta]]. \end{aligned}$$

Particle simulation smoother. Similarly to the particle simulation smoother for the forward block sampling, it is possible to implement the simulation smoother for the backward block sampling. The proof is omitted since it similar to that of Lemma 4.3.

5 Illustrative examples

5.1 Linear Gaussian state space model

Example 1 (continued).
We revisit the example of the linear Gaussian state space model discussed in Section 2.3 and the rolling estimation is conducted with the window size of $L+1 = 1000$ for $t = 1, \ldots, 2000$ ($T = 2000$) and $N = 1000$ using the particle rolling MCMC sampler with and without the double block sampling. We choose $K = 1, 2, 3, 5$ and 10 to investigate the effect of the block size. Since it is possible to use a fully adapted proposal density in the linear Gaussian state space model, we consider the double block sampling with (1) a fully adapted proposal density and (2) a proposal density based on the marginalized approach discussed in Section 3. Also we use (3) the simple particle rolling MCMC sampler as a benchmark. In summary, we consider

(1) Double block sampling with a fully adapted proposal density.

In the forward block sampling, generate $\alpha^n_{t-K:t} \sim p(\alpha_{t-K:t} \mid \alpha_{t-K-1}, y_{t-K-1:t}, \theta)$ with its importance weight $W^n_{s-1:t} \propto p(y_t \mid y_{s-1:t-1}, \alpha^n_{t-K-1}, \theta^n) \times W^n_{s-1:t-1}$. In the backward block sampling, generate $\alpha^n_{s-1:s+K-1} \sim p(\alpha_{s-1} \mid \alpha_s, \theta)p(\alpha_{s:s+K-1} \mid \alpha_{s+K}, y_{s-1:s+K-1}, \theta)$ with its importance weight $W^n_{s:t} \propto p(y_{s-1} \mid y_{s:t}, \alpha^n_{s+K}, \theta^n)^{-1} \times W^n_{s-1:t}$.

(2) Double block sampling with $M = 100, 300$ and 500.

(3) Simple particle rolling MCMC sampler (without the block sampling).

Table 1: The number of resampling steps in block sampling and simple sampling.

Estimation period	K	(1) Fully adapted	(2) M 100	300	500	(3) Simple sampling
	1	30	54	37	32	184
Initial estimation	2	10	39	21	15	
($t = 1, \ldots, 1000$)	3	8	38	19	15	
	5	6	37	18	15	
	10	8	38	18	14	
	1	48	104	71	61	1027
Rolling estimation	2	8	74	33	23	
($t = 1001, \ldots, 2000$)	3	7	74	31	22	
	5	6	72	32	22	
	10	5	69	31	23	

Table 1 shows the number of resampling steps triggered in the initialization period ($t = 1, \ldots, 1000$) and the rolling estimation period ($t = 1001, \ldots, 2000$). For the simple sampling, resampling steps are triggered 184 times in the initial estimation stage and 1027 times in the rolling estimation stage. Compared with this benchmark, the weight degeneracy is drastically eased by the block sampling in (1) and (2). Using $K = 2$, the numbers of resampling steps are less than 1% and 10% (10% and 22%) for (1) and (2) in the rolling estimation (the initial estimation) period. Also, the effect of the block sampling seems to be maximized at $K = 2$, and the number of resampling steps of (2) decreases to that of (1) as M increases. Overall, we found that the double block sampling is most efficient when $K = 2$ and $M = 100$ in this example.

Figure 6 shows histograms of R_{1t} and R_{2t} ($t = 1001, \ldots, 2000$) for the simple sampler using dotted lines, which are reproduced from Figure 2. The ratio R_{1t} (R_{2t}) measures the relative magnitude of ESS in Step 1a after adding a new observation (removing the oldest observation) when compared with that of the previous step at time t. The histograms of R_{1t} and R_{2t} for the sampler with the block sampling with $K = 2$ and $M = 100$ are shown using solid lines. The R_{1t}'s for the block sampling are larger and less dispersed than those for the simple sampler suggesting that the forward block sampling is more efficient. Also, the R_{2t}'s for the block sampling are much larger and much less dispersed than those for the simple sampler, which implies that the backward block sampling is highly efficient. The scatter plot of R_{1t} and R_{2t} is shown at the bottom of Figure 6 for two sampling methods. It shows that our proposed block sampling is highly efficient at both Steps 1 and 2 of each rolling step.

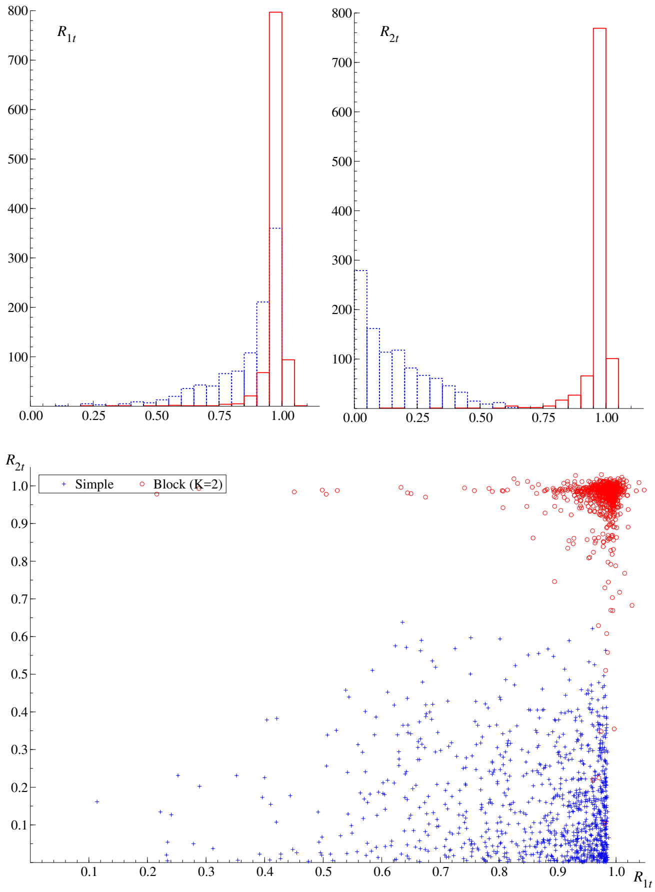

Figure 6: The histograms of R_{1t} (top left) and R_{2t} (top right) ($t = 1001, \ldots, 2000$) for the simple sampler (dotted blue) and the sampler with block sampling with $K = 2$ and $M = 100$ (solid red). The scatter plot of R_{1t} and R_{2t} (bottom).

Table 2 shows the summary statistics for the relative magnitudes of ESS in each step, R_{1t} and R_{2t}. In Step 1a, the average of R_t's for the block sampling is a bit larger than that for the simple sampling, but the standard deviation for the former is less than half for the latter. Moreover, in Step 2a, the average of R_t's for the block sampling is six times larger than that for the simple sampling, while the standard deviation for the former is about half. Thus the double block sampling drastically alleviate the weight degeneracy compared with the simple sampling method.

Table 2: Summary statistics of R_{1t} and R_{2t} for the simple sampling and the block sampling ($M = 100, K = 2$) for $t = 1001, \ldots, 2000$

	Method	Mean	Median	Std. dev.
R_{1t}	Simple	0.862	0.924	0.145
	Block	0.975	0.988	0.057
R_{2t}	Simple	**0.161**	**0.127**	**0.139**
	Block	**0.970**	**0.988**	**0.068**

Finally, to check the accuracy of the proposed rolling window estimation (with $K = 2$ and $M = 100$), we compare the estimation results with their corresponding analytical solutions. The particles are 'refreshed' in the MCMC update step so that the approximation errors do not accumulate over time. The algorithm seems to correctly capture means and 95% credible intervals of the target posterior distribution as shown in Figure 7.

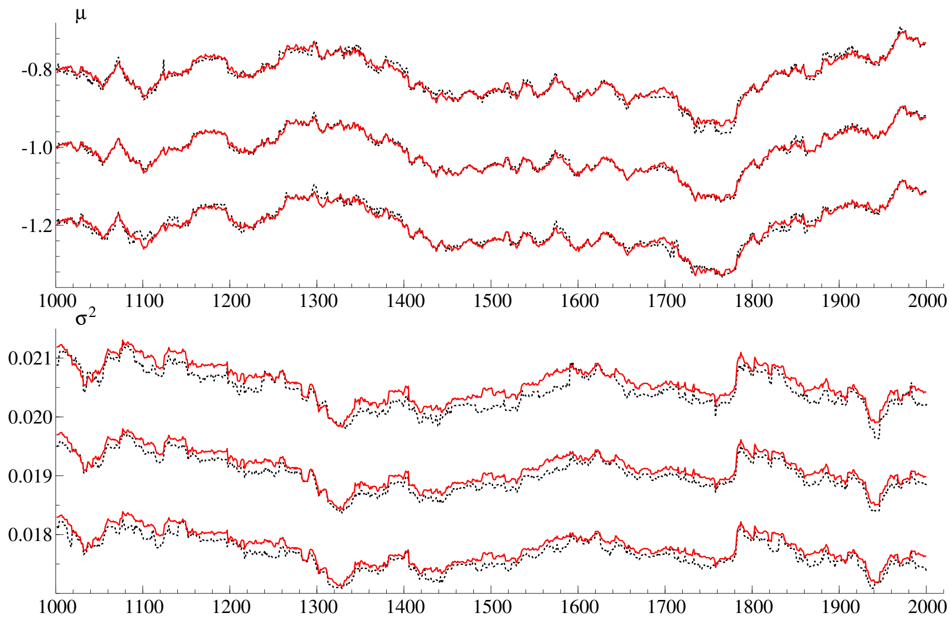

Figure 7: True posterior means and 95% credible intervals (dotted black) with their estimates (solid red) for μ and σ^2.

Further, Figure 8 shows true log marginal likelihoods and their estimates with errors for $t = 1001, \dots, 2000$. The estimation errors are very small overall, implying that the proposed algorithm estimates the marginal likelihood $p(y_{t-999:t})$ accurately.

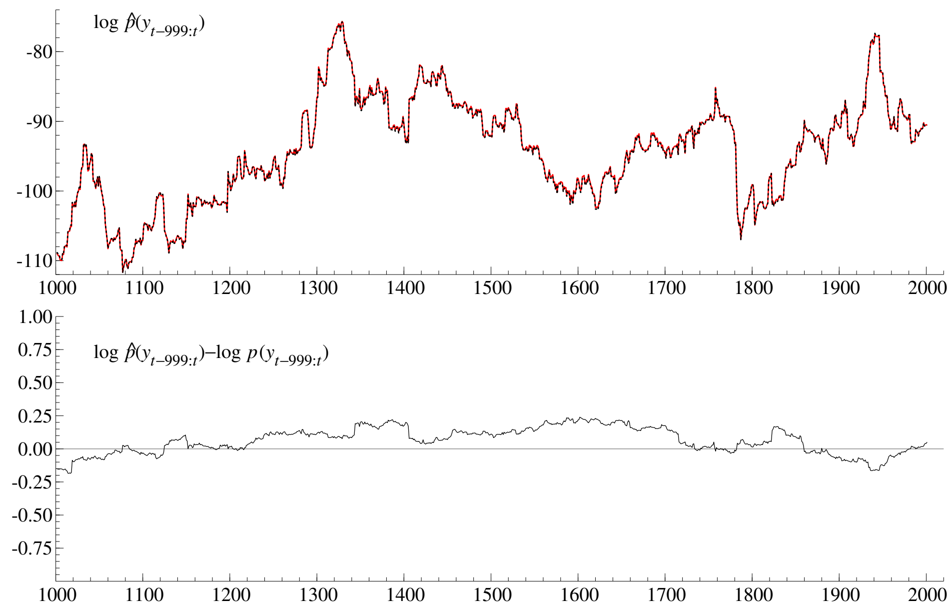

Figure 8: Top: true log marginal likelihoods $\log p(y_{t-999:t})$ (dotted black) and their estimates (solid red). Bottom: estimation errors $\log \hat{p}(y_{t-999:t}) - \log p(y_{t-999:t})$ for $t = 1001, \dots, 2000$.

5.2 Realized stochastic volatility model with leverage

Example 2. We apply our proposed methods to the rolling estimation for the financial time series. Consider a stochastic volatility (SV) model with an additional measurement equation for the realized volatility (RV), called realized stochastic volatility (RSV) model (e.g. Takahashi et al. (2009)). Let $y_{1,t}$ and $y_{2,t}$ denote the daily log return and the logarithm of the realized volatility (variance) at time t and let α_t denote the latent log volatility which is assumed to follow AR(1) process. The RSV model is defined as

$$
\begin{aligned}
y_{1,t} &= \exp(\alpha_t/2)\epsilon_t, \ \epsilon_t \sim \mathcal{N}(0,1), \ t = 1, \dots, T \\
y_{2,t} &= \alpha_t + \xi + u_t, \ u_t \sim \mathcal{N}(0, \sigma_u^2), \ t = 1, \dots, T \\
\alpha_{t+1} &= \mu + \phi(\alpha_t - \mu) + \eta_t, \ \eta_t \sim \mathcal{N}(0, \sigma_\eta^2), \ t = 1, \dots, T, \\
\alpha_1 &= \mu + \frac{1}{\sqrt{1-\phi^2}}\eta_0, \ \eta_0 \sim \mathcal{N}(0, \sigma_\eta^2),
\end{aligned}
$$

where

$$\begin{pmatrix} \epsilon_t \\ u_t \\ \eta_t \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & \rho\sigma_\eta \\ 0 & \sigma_u^2 & 0 \\ \rho\sigma_\eta & 0 & \sigma_\eta^2 \end{bmatrix} \right). \tag{55}$$

The correlation between ϵ_t and η_t is introduced to express the leverage effect. It is often found to be negative in empirical studies, which implies that the decrease in the today's log return is followed by the increase in the log volatility next day. All static parameters in the model are assumed to be unknown so that $\theta = (\mu, \phi, \sigma_\eta^2, \xi, \sigma_u^2, \rho)'$.

For y_{1t} and y_{2t}, we use the daily log returns and log realized volatilities of Standard and Poor's (S&P) 500 index, which are obtained from Oxford-Man Institute Realized Library [1] created by Heber et al. (2009) (see Shephard and Sheppard (2010) for details). The initial estimation period is from January 1, 2000 ($t = 1$) to December 31, 2007 ($t = 1988$) with $L + 1 = 1988$. The rolling estimation started from sampling from the posterior distribution using this initial sample period and moved the window until December 30, 2016 ($T = 4248$). Thus the first estimation period is before the financial crisis caused by the bankruptcy of Lehman Brothers and the last estimation period is after the crisis. We basically set $K = 10$, $M = 300$ and $N = 1000$, and further, in Steps 1b and 2b, we always implement 10 MCMC iterations for the comparison below unless otherwise stated.

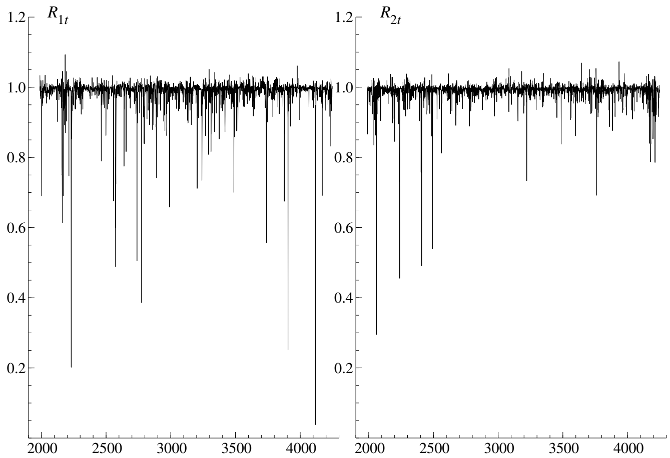

Figure 9: Traceplot of R_{1t} (left) and R_{2t} (right), ($t = 1988, \dots, 4248$) in RSV model.

First, the time series plots of the ratios of ESS in the forward and backward block sampling, are shown in Figure 9 for the period from December 31, 2007 to December 30, 2016 ($t = 1988, \dots, 4248$). Both of R_{1t} and R_{2t} are close to one most of the time throughout

[1]The data is downloaded at http://realized.oxford-man.ox.ac.uk/data/download

the sample period, which implies the particles are scattered around the state space, and the particle rolling MCMC with double block sampling is highly efficient. Also, the summary statistics of R_{1t} and R_{2t} are shown in Table 3 where we use $K = 5, 10$ and 15. As K increases, R_{1t} and R_{2t} become larger and less dispersed, but the difference becomes smaller for $K = 10$ and $K = 15$.

	K	Mean	Median	Std. dev.
R_{1t}	5	0.981	0.995	0.058
	10	0.985	0.996	0.053
	15	0.986	0.997	0.055
R_{2t}	5	0.983	0.993	0.044
	10	0.988	0.994	0.036
	15	0.988	0.994	0.035

Table 3: Summary statistics for R_{1t} and R_{2t} ($t = 1988, \ldots, 4248$) in RSV model with leverage using $K = 5, 10$ and 15.

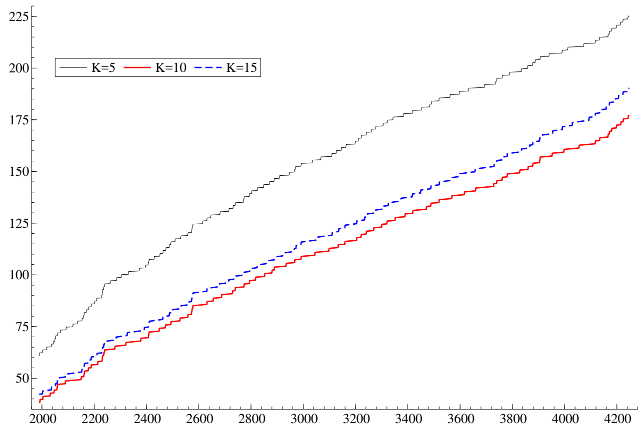

Figure 10: Cumulative computation times (wall time, unit time =1000 seconds) using $K = 5, 10$ and 15 ($t = 1988, \ldots, 4248$) in RSV model.

Figure 10 shows three cumulative computation times (wall time) for the same period corresponding to $K = 5, 10$ and 15. The computational time with $K = 5$ is much longer than those with $K = 10$ and $K = 15$, while the difference in the computational times seems to be relatively small for $K = 10$ or 15. It indicates that $K = 10$ is a reasonable choice in the RSV model in contrast to $K = 2$ in the linear Gaussian state model in Example 1. It is expected to reduce the weight degeneracy phenomenon and hence the number of MCMC update steps which is computationally expensive.

We investigate the estimation accuracy of the proposed sampling algorithm using the posterior distribution function of $\theta = (\mu, \phi, \sigma_\eta^2, \xi, \sigma_u^2, \rho)'$ for the first period from January 1, 2000 to December 31, 2007 ($t = 1, \ldots, 1988$) and the last period from February 10, 2009 to December 30, 2016 ($t = 2261, \ldots, 4248$). First, the MCMC sampling is conducted for these two periods to obtain the accurate estimates of the distribution functions (see Takahashi et al. (2009) for the details of MCMC sampling). Then we apply our proposed sampling algorithm with the size of block $K = 10$, $M = 300$ and $N = 1000$ where both θ^n and α^n are updated in MCMC steps by drawing from the full conditional posterior distribution. Three cases for the number of iterations are considered in MCMC steps: (1) one iteration as in Section 5.1, (2) 5 iterations and (3) 10 iterations. Figure 11 and Figure 12 show the estimation results for the first and the last estimation periods.

Among three case, the estimates obtained by iterating MCMC algorithm 5 or 10 times in MCMC update steps are close to those obtained by ordinary MCMC sampling algorithm in both Figures. If we iterate only once in MCMC update step, the estimation results are found to be inaccurate. This is because the MCMC iterations not only diversify the the particles but also correct approximation errors by the particles. The estimation errors for the distribution function of μ are most serious, probably because the mixing property of MCMC sampling in the RSV model is poor especially with respect to μ as discussed in the numerical studies of Takahashi et al. (2009). Thus these results suggest that MCMC iterations should be implemented a sufficient number of times in MCMC update steps so that the particles can trace the correct posterior distributions.

Figure 13 shows the traceplot of estimated posterior means and 95% credible intervals for $\theta = (\mu, \phi, \sigma_\eta^2, \xi, \sigma_u^2, \rho)'$ from December 31, 2007 ($t = 1988$) to December 30, 2016 ($t = 4248$). By implementing the rolling estimation, we are able to observe the transition of the economic structure and the effect of the financial crisis ($t = 2150, \ldots, 2213$ correspond to September, October and November in 2008) . The posterior distribution of μ seems to be stable before $t = 4000$ (January 7, 2016), but its mean and 95% intervals decrease after $t = 4000$. The average level of log volatility started to decrease toward the end of the sample period. The autoregressive parameter, ϕ, continues to decrease throughout the sample period indicating that the latent log volatility becomes less persistent. The variances, σ_η^2 and σ_u^2, of error terms in the state equation and the measurement equation of the log realized volatility continue to increase, while the bias adjustment term, ξ, and the leverage effect, ρ, become closer to zero during the sample period. The leverage effects in the stock market are found to become weaker after the financial crisis.

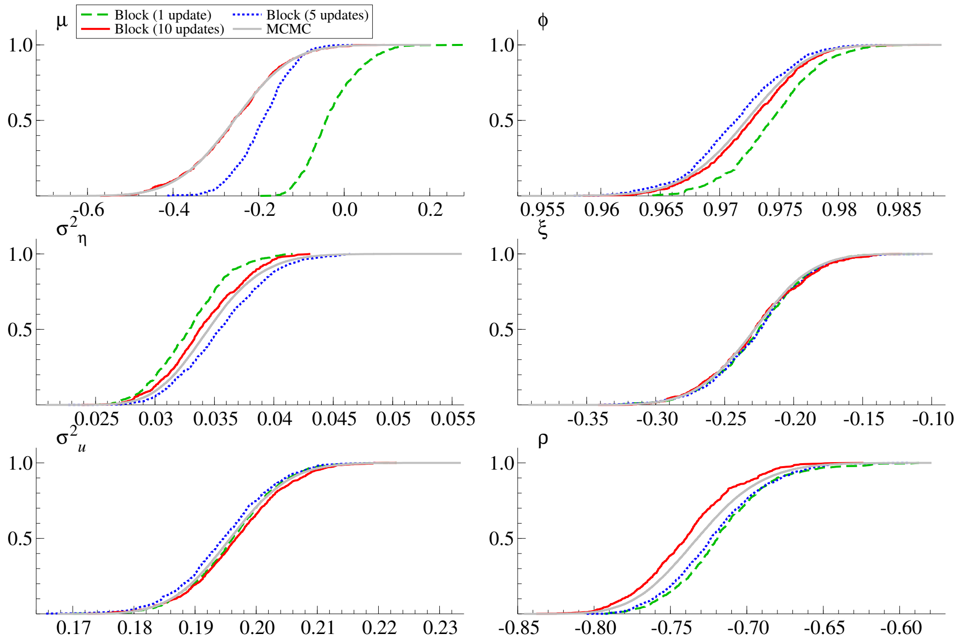

Figure 11: The estimated posterior distribution functions of θ for $t = 1, \ldots, 1988$. MCMC and Particle rolling MCMC: 1, 5 and 10 iterations.

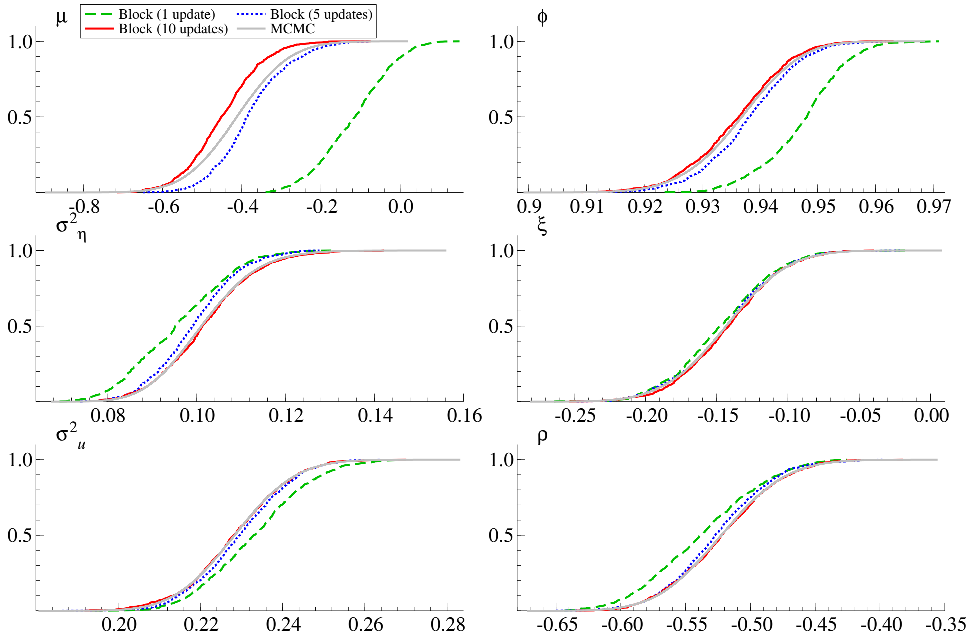

Figure 12: The estimated posterior distribution functions of θ for $t = 2261, \ldots, 4248$. MCMC and Particle rolling MCMC: 1, 5 and 10 iterations.

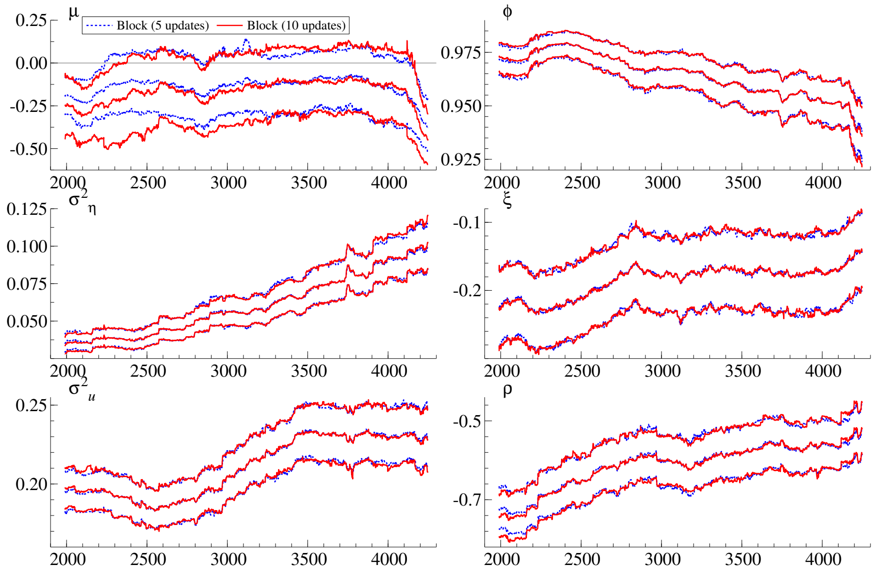

Figure 13: Traceplot of estimated posterior means and 95% credible intervals for parameters using S&P500 return in RSV model (from December 31, 2007 to December 30, 2016).

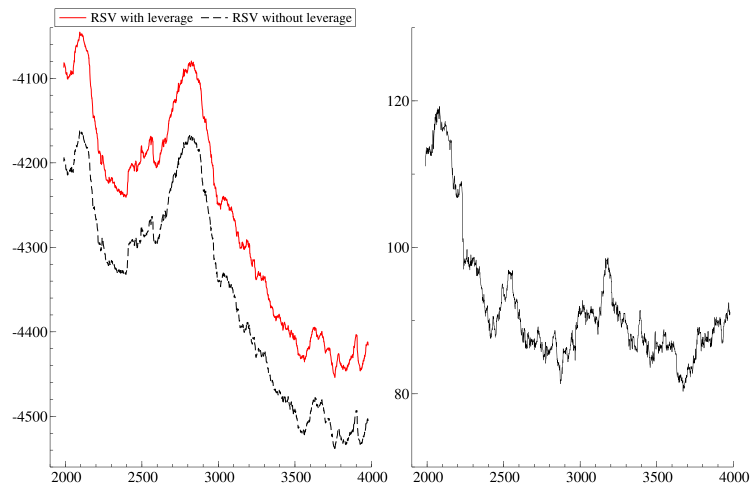

Figure 14: Left: Estimates of $\log p(y_{t-1987:t})$ ($t = 1988, \ldots, 4248$) for S&P 500 index return in RSV model with leverage (solid red) and in RSV model without leverage (dotted black). Right: Difference between two log marginal likelihoods.

The log marginal likelihoods, $\log p(y_{t-1987:t})$, of the RSV model with and without leverage effect are shown in Figure 14 for the period from December 31, 2007 ($t = 1988$) to December 30, 2016 ($t = 4248$). The log marginal likelihood for the RSV model with leverage effect is always larger than the other model, supporting the RSV model with leverage effect. This is consistent with the estimation result of ρ in Figure 13 where ρ is negative throughout the sample period. The difference between two log marginal likelihoods decreases until $t = 2400$ (August 28, 2009), and seems to become stable after $t = 2400$.

Finally, we focus on the special case of our applications, the sequential Bayesian estimation of parameter θ given $y_{1:t}$ ($t = 1, 2, \ldots, 1988$) to compare our proposed methodology, PRMCMC, with the promising alternative sequential MCMC algorithm, SMC^2 (Chopin et al. (2013)). The simulation experiments are repeated with different random seeds to see whether the obtained estimation results are stable. The PRMCMC is implemented using $N = 1000$, $M = 300$ and $K = 10$ as in the previous section, and we iterate MCMC update steps 10 times to refresh particles. Similarly, for the SMC^2, we use $N_\theta = 1000$ particles for θ, and $N_x = 100$ particles for $\alpha_{1:t}$ as the initial number of particles to approximate $p(y_{1:t}|\theta)$. N_x is doubled when the acceptance rate in the particle MH update steps becomes smaller than 0.2 following Chopin et al. (2013). In the particle MH update steps, the parameter θ is generated using the random walk MH algorithm using the normal proposal $\mathcal{N}(\theta^*, c\hat{\Sigma})$ where θ^* is a current value of the parameter θ, c is a tuning parameter (0.3 or 0.5), and $\hat{\Sigma}$ is the estimated covariance matrix.

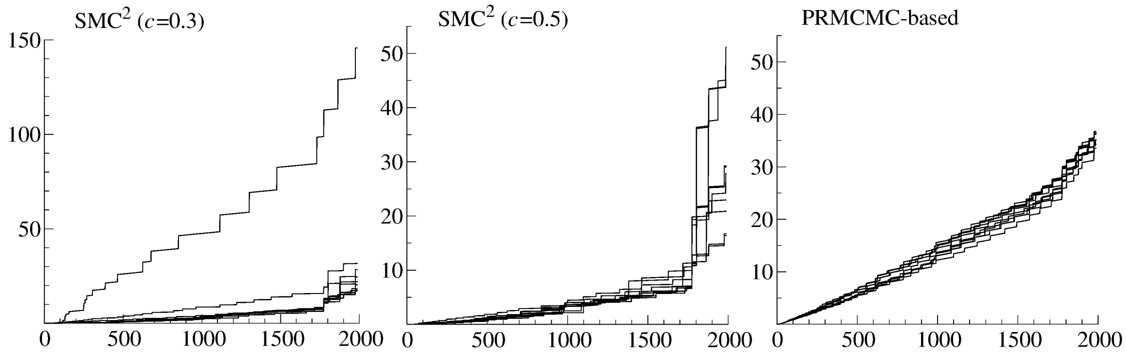

Figure 15: Cumulative computation times (wall time, unit time =1000 seconds) over 10 runs with SMC^2 with $c = 0.3$ (left), 0.5 (middle) and with the PRMCMC-based algorithm (right).

Figure 15 shows computational times for (1) SMC^2 with $c = 0.3$ (left) (2) SMC^2 with $c = 0.5$ (middle) and (3) PRMCMC (right), using 10 different random seeds respectively. The computational times for SMC^2 are found to be very large for some random seeds and are not stable, especially for $c = 0.3$, and jump up quickly for $t > 1800$ for $c = 0.3$ and 0.5. On the other hand, they are very stable and gradually increase as t increases for PRMCMC. Further, the estimated posterior cumulative distribution functions of $\theta = (\mu, \phi, \sigma_\eta^2, \sigma_u^2, \rho)'$

are shown for (1) SMC^2 with $c = 0.3$ (top two rows) and (2) SMC^2 with $c = 0.5$ (bottom two rows) in Figure 16, and for (3) PRMCMC in Figure 17.

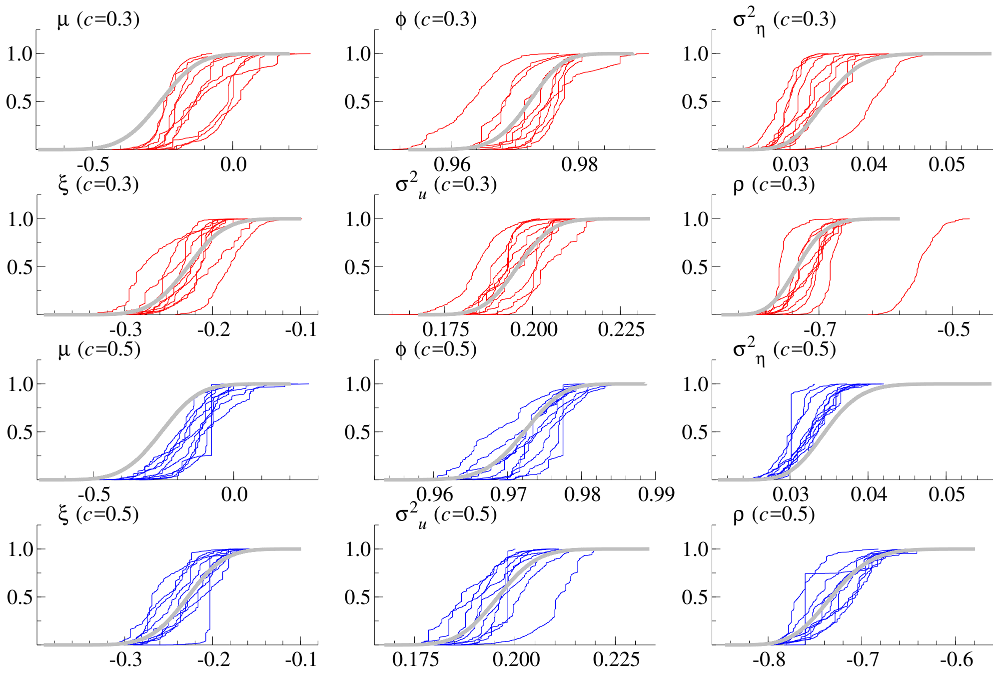

Figure 16: Estimated posterior cumulative distribution functions of θ given $y_{1:1988}$ over 10 runs with SMC^2 with $c = 0.3, 0.5$(red and blue, respectively) compared to the result of MCMC (gray).

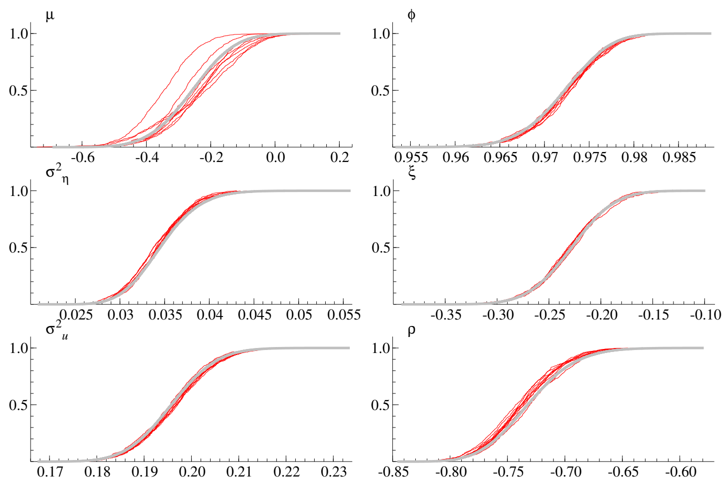

Figure 17: Estimated posterior cumulative distribution functions of θ given $y_{1:1988}$ over 10 runs with the PRMCMC-based method (red) compared to the result of MCMC (gray).

The estimates of posterior cumulative distribution functions using SMC^2 are not stable regardless of the selection of c and random seeds, and sometimes deviate very far from those based on the MCMC which are supposed to give more accurate estimates at the expense of additional computational costs. On the other hand, the estimates using PRMCMC are stable and close to those based on the MCMC regardless of the random seeds. Overall, we found our PRMCMC outperforms SMC^2 in simulation experiments, partly owing to the MCMC update steps which are implemented to refresh θ^n and $\alpha_{1:t}^n$ in PRMCMC.

6 Conclusion

In this paper, we propose the novel efficient estimation method to implement the rolling window particle MCMC simulation using the framework of the conditional SMC update. The weighted particles are updated to learn and forget the information of the new and old observations by the forward and backward block sampling with particle simulation smoother, and further propagated by the MCMC update step. The proposed estimation methodology is also applicable to the ordinary sequential estimation with parameter uncertainty. Its computational performance is evaluated in illustrative examples, using the linear Gaussian state space model with simulated data and the realized stochastic volatility model with S&P500 index returns. It is shown that the posterior distributions of model parameters and marginal likelihoods are estimated with accuracy and the new methodology is prospective also for ordinary sequential analysis . The empirical study of S&P500 index returns shows the gradual transition of the economic structure by observing the posterior distributions which are obtained from the particle rolling MCMC estimation.

Acknowledgement

All computational results in this paper are generated using Ox metrics 7.0 (see Doornik (2009)). This work was supported by JSPS KAKENHI Grant Numbers 25245035, 26245028, 17H00985, 15H01943.

References

Andrieu, C., A. Doucet, and R. Holenstein (2010). Particle Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B: Statistical Methodology 72(3)*, 269–342.

Chopin, N., P. E. Jacob, and O. Papaspiliopoulos (2013). SMC2: An efficient algorithm for sequential analysis of state space models. *Journal of the Royal Statistical Society.*

Series B: Statistical Methodology 75(3), 397–426.

Chopin, N. and S. S. Singh (2015). On particle Gibbs sampling. *Bernoulli 21*(3), 1855–1883.

de Jong, P. and N. Shephard (1995). The simulation smoother for time series models. *Biometrika 82*, 339–350.

Del Moral, P., A. Doucet, and A. Jasra (2006). Sequential Monte Carlo samplers. *Journal of the Royal Statistical Society. Series B: Statistical Methodology 68*(3), 411–436.

Doornik, J. A. (2009). *An Object-oriented Matrix Programming Language - Ox 6*. London: Timberlake Consultants Press and Oxford. www.doornik.com.

Doucet, A., M. Briers, and S. Sénécal (2006). Efficient block sampling strategies for sequential Monte Carlo methods. *Journal of Computational and Graphical Statistics 15*(3), 693–711.

Doucet, A., N. De Freitas, and N. Gordon (2001). An introduction to sequential monte carlo methods. In *Sequential Monte Carlo methods in practice*, pp. 3–14. Springer.

Durbin, J. and S. J. Koopman (2002). A simple and efficient simulation smoother for state space time series analysis. *Biometrika 89*, 603–616.

Fulop, A. and J. Li (2013). Efficient learning via simulation: A marginalized resample-move approach. *Journal of Econometrics 176*(2), 146–161.

Heber, G., A. Lunde, N. Shephard, and K. Sheppard (2009). Oxford-man Institute's realized library. version 0.2, Oxford-Man Institute, University of Oxford.

Omori, Y., S. Chib, N. Shephard, and J. Nakajima (2007). Stochastic volatility with leverage: fast and efficient likelihood inference. *Journal of Econometrics 140*(2), 425–449.

Pitt, M. K. and N. Shephard (1999). Filtering via simulation: auxiliary particle filters. *Journal of the American Statistical Association 94*(446), 590–599.

Polson, N. G., J. R. Stroud, and P. Müller (2008). Practical filtering with sequential parameter learning. *Journal of the Royal Statistical Society. Series B: Statistical Methodology 70*(2), 413–428.

Shephard, N. and M. K. Pitt (1997). Likelihood analysis of non-Gaussian measurement time series. *Biometrika 84*, 653–667.

Shephard, N. and K. Sheppard (2010). Realising the future: forecasting with high-frequency-based volatility (HEAVY) models. *Journal of Applied Econometrics 25*(2), 197–231.

Takahashi, M., Y. Omori, and T. Watanabe (2009). Estimating stochastic volatility models using daily returns and realized volatility simultaneously. *Computational Statistics and Data Analysis 53*(6), 2404–2426.

Watanabe, T. and Y. Omori (2004). A multi-move sampler for estimating non-Gaussian time series models: Comments on Shephard and Pitt (1997). *Biometrika 91*, 246–248.

Whiteley, N., C. Andrieu, and A. Doucet (2010). Efficient Bayesian inference for switching state-space models using discrete particle Markov chain Monte Carlo methods. arXiv preprint arXiv:1011.2437.

A Proofs

A.1 Proof of Lemma 4.1

Using Bayes' theorem and

$$v_{j,\theta}(\alpha_{j-1}^{a_{j-1}^m}, \alpha_j^m) = \frac{f_\theta(\alpha_j^m \mid \alpha_{j-1}^{a_{j-1}^m}, y_{j-1}) g_\theta(y_j \mid \alpha_j^m)}{q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_{j-1}^m}, y_j)}, \quad j = 1, \ldots, M,$$

the numerator of the first term in (43) is

$$\begin{aligned}
&\pi(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{k_{t-K}}, \ldots, \alpha_{t_0}^{k_{t_0}} \mid y_{s-1:t_0}) \\
&= \frac{\pi(\theta, \alpha_{s-1:t-K-1} \mid y_{s-1:t-K-1})}{p(y_{t-K:t_0} \mid y_{s-1:t-K-1})} \prod_{j=t-K}^{t_0} f_\theta(\alpha_j^{k_j} \mid \alpha_{j-1}^{k_{j-1}}, y_{j-1}) g_\theta(y_j \mid \alpha_j^{k_j}) \\
&= \frac{\pi(\theta, \alpha_{s-1:t-K-1} \mid y_{s-1:t-K-1})}{p(y_{t-K:t_0} \mid y_{s-1:t-K-1})} \prod_{j=t-K}^{t_0} v_{j,\theta}(\alpha_{j-1}^{k_{j-1}}, \alpha_j^{k_j}) \prod_{j=t-K}^{t_0} q_{j,\theta}(\alpha_j^{k_j} \mid \alpha_{j-1}^{k_{j-1}}, y_j). \quad (56)
\end{aligned}$$

Thus we obtain

$$\begin{aligned}
(43) \quad &= \frac{\pi(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{k_{t-K}}, \ldots, \alpha_{t_0}^{k_{t_0}} \mid y_{s-1:t_0})}{M^{t_0-(t-K)+1}} \times \prod_{j=t-K}^{t_0} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_{j-1}^m}, y_j) \\
&\quad \times \prod_{j=t-K+1}^{t_0} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j-1,\theta}^{a_{j-1}^m} \\
&= \frac{\pi(\theta, \alpha_{s-1:t-K-1} \mid y_{s-1:t-K-1})}{M^{t_0-(t-K)+1} p(y_{t-K:t_0} \mid y_{s-1:t-K-1})} \times \prod_{j=t-K}^{t_0} \prod_{m=1}^{M} q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_{j-1}^m}, y_j) \\
&\quad \times \prod_{j=t-K+1}^{t_0} v_{j-1,\theta}(\alpha_{j-2}^{k_{j-2}}, \alpha_{j-1}^{k_{j-1}}) \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j-1,\theta}^{a_{j-1}^m} \times v_{t_0,\theta}(\alpha_{t_0-1}^{k_{t_0-1}}, \alpha_{t_0}^{k_{t_0}}) \\
&= \frac{\pi(\theta, \alpha_{s-1:t-K-1} \mid y_{s-1:t-K-1})}{\prod_{j=t-K}^{t_0} p(y_j \mid y_{s-1:j-1})} \times \prod_{j=t-K}^{t_0} \prod_{m=1}^{M} q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_{j-1}^m}, y_j) \\
&\quad \times \prod_{j=t-K+1}^{t_0} \prod_{m=1}^{M} V_{j-1,\theta}^{a_{j-1}^m} \times V_{t_0,\theta}^{k_{t_0}} \times \prod_{j=t-K}^{t_0} \hat{p}(y_j \mid y_{s-1:j-1}, \alpha_{t-K-1}, \theta)
\end{aligned}$$

and the result follows where we substitute (56) in the second equality, and used the definition of $\hat{p}(y_j \mid y_{s-1:j-1}, \alpha_{t-K-1}, \theta)$ in the third equality. □

A.2 Proof of Proposition 4.1

By applying Lemma 4.1 with $t_0 = t-1$ to the first three terms of (42), we obtain

$$
\begin{aligned}
&\hat{\pi}(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K:t}^{1:M}, a_{t-K:t-1}^{1:M}, k_t^* \mid y_{s-1:t}) \\
&= \pi(\theta, \alpha_{s-1:t-K-1} \mid y_{s-1:t-K-1}) \times \prod_{m=1}^{M} q_{t-K,\theta}(\alpha_{t-K}^m \mid \alpha_{t-K-1}, y_{t-K}) \\
&\quad \times \prod_{j=t-K+1}^{t-1} \prod_{m=1}^{M} V_{j-1,\theta}^{a_{j-1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_{j-1}^m}, y_j) \times V_{t-1,\theta}^{k_{t-1}} \times \prod_{j=t-K}^{t-1} \frac{\hat{p}(y_j \mid y_{s-1:j-1}, \theta, \alpha_{t-K-1})}{p(y_j \mid y_{s-1:j-1})} \\
&\quad \times\ q_{t,\theta}(\alpha_t^1 \mid \alpha_{t-1}^{k_{t-1}}, y_t) \times \prod_{m=2}^{M} V_{t-1,\theta}^{a_{t-1}^m} q_{t,\theta}(\alpha_t^m \mid \alpha_{t-1}^{a_{t-1}^m}, y_t) \times V_{t,\theta}^{k_t^*} \times \frac{\hat{p}(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta)}{p(y_t \mid y_{s-1:t-1})} \\
&= \pi(\theta, \alpha_{s-1:t-K-1} \mid y_{s-1:t-K-1}) \times \prod_{m=1}^{M} q_{t-K,\theta}(\alpha_{t-K}^m \mid \alpha_{t-K-1}, y_{t-K}) \\
&\quad \times \prod_{j=t-K+1}^{t} \prod_{m=1}^{M} V_{j-1,\theta}^{a_{j-1}^m} q_{j,\theta}(\alpha_j^m, \alpha_{j-1}^{a_{j-1}^m} \mid y_j) \times \prod_{j=t-K}^{t} \frac{\hat{p}(y_j \mid y_{s-1:j-1}, \theta, \alpha_{t-K-1})}{p(y_j \mid y_{s-1:j-1})} \times V_{t,\theta}^{k_t^*},
\end{aligned}
$$

where we note $a_{t-1}^1 = k_{t-1}$ and $k_t = 1$. Apply Lemma 4.1 with $t_0 = t$ and $k_{t_0} = k_t^*$ to the last equation and the result follows. □

A.3 Proof of Proposition 4.2

Proof. We first define the probability density function

$$
\begin{aligned}
&\psi_{\theta,0}(\alpha_{t-K:t}^{1:M}, a_{t-K:t-1}^{1:M}, k_t^* \mid \alpha_{t-K-1}, y_{s-1:t}) \\
&\equiv \frac{\pi(\alpha_{t-K:t-1} \mid \alpha_{t-K-1}, y_{s-1:t-1}, \theta)}{M^K} \\
&\quad \times \psi_\theta(\alpha_{t-K}^{-k_{t-K}}, \ldots, \alpha_{t-1}^{-k_{t-1}}, \alpha_t^{1:M}, a_{t-K}^{-k_{t-K+1}}, \ldots, a_{t-1}^{-k_t}, k_t^* \mid \alpha_{t-K-1:t-1}, a_{t-K}^{k_{t-K+1}}, \ldots, a_{t-1}^{k_t}, y_{t-K:t}),
\end{aligned}
$$

where $(\alpha_{t-K}^{k_{t-K}}, \ldots, \alpha_{t-1}^{k_{t-1}}) = \alpha_{t-K:t-1}$ and

$$
\pi(\alpha_{t-K:t-1} \mid \alpha_{t-K-1}, y_{s-1:t-1}, \theta) = \frac{\pi(\theta, \alpha_{s-1:t-1} \mid y_{s-1:t-1})}{\pi(\theta, \alpha_{s-1:t-K-1} \mid y_{s-1:t-1})}.
$$

Noting that

$$
\begin{aligned}
&\hat{p}(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta)\psi_{\theta,0}(\alpha_{t-K:t}^{1:M}, a_{t-K:t-1}^{1:M}, k_t^* \mid \alpha_{t-K-1}, y_{s-1:t}) \\
&= \hat{\pi}(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K:t}^{1:M}, a_{t-K:t-1}^{1:M}, k_t^* \mid y_{s-1:t}) \frac{p(y_t \mid y_{s-1:t-1})}{\pi(\theta, \alpha_{s-1:t-K-1} \mid y_{s-1:t-1})},
\end{aligned}
$$

where we used the definition of $\hat{\pi}$ in (42),

$$
\begin{aligned}
&E_{\psi_{\theta,0}}\left[\hat{p}(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta) \mid \alpha_{t-K-1}, y_{s-1:t}, \theta\right] \\
&= \int \hat{p}(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta)\psi_{\theta,0}(\alpha_{t-K:t}^{1:M}, a_{t-K:t-1}^{1:M}, k_t^* \mid \alpha_{t-K-1}, y_{s-1:t}) d\alpha_{t-K:t}^{1:M} da_{t-K:t-1}^{1:M} dk_t^* \\
&= \int \hat{\pi}(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K:t}^{1:M}, a_{t-K:t-1}^{1:M}, k_t^* \mid y_{s-1:t}) d\alpha_{t-K:t}^{1:M} da_{t-K:t-1}^{1:M} dk_t^* \frac{p(y_t \mid y_{s-1:t-1})}{\pi(\theta, \alpha_{s-1:t-K-1} \mid y_{s-1:t-1})} \\
&= \frac{\pi(\theta, \alpha_{s-1:t-K-1} \mid y_{s-1:t}) p(y_t \mid y_{s-1:t-1})}{\pi(\theta, \alpha_{s-1:t-K-1} \mid y_{s-1:t-1})} \\
&= p(y_t \mid y_{s-1:t-1}, \alpha_{s-1:t-K-1}, \theta) = p(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta).
\end{aligned}
$$

Also it is easy to see

$$
\begin{aligned}
&E[p(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta) \mid y_{s-1:t}] \\
&= \int p(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta)\pi(\theta, \alpha_{t-K-1} \mid y_{s-1:t-1}) d\theta d\alpha_{t-K-1} = p(y_t \mid y_{s-1:t-1}).
\end{aligned}
$$

□

A.4 Proof of Proposition 4.3

Proof. Consider the joint marginal density of (46):

$$
\begin{aligned}
&\hat{\pi}(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K:t_0}^{1:M}, a_{t-K:t_0-1}^{1:M}, \alpha_{t_0+1}^{k_{t_0+1}^*}, \ldots, \alpha_t^{k_t^*}, k_{t_0:t}^* \mid y_{s-1:t}) \\
&= \frac{\pi(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{k_{t-K}^*}, \ldots, \alpha_t^{k_t^*} \mid y_{s-1:t})}{M^{K+1}} \times \prod_{\substack{m=1 \\ m \neq k_{t-K}^*}}^{M} q_{t-K,\theta}(\alpha_{t-K}^m \mid \alpha_{t-K-1}, y_{t-K}) \\
&\quad \times \prod_{j=t-K+1}^{t_0} \prod_{\substack{m=1 \\ m \neq k_j^*}}^{M} V_{j-1}^{a_j^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_j^m}, y_j),
\end{aligned} \tag{57}
$$

for $t_0 = t-1, \ldots, t-K+1$, and

$$
\begin{aligned}
&\hat{\pi}(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{1:M}, \alpha_{t-K+1}^{k_{t-K+1}^*}, \ldots, \alpha_t^{k_t^*}, k_{t-K:t}^* \mid y_{s-1:t}) \\
&= \frac{\pi(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{k_{t-K}^*}, \ldots, \alpha_t^{k_t^*} \mid y_{s-1:t})}{M^{K+1}} \times \prod_{\substack{m=1 \\ m \neq k_{t-K}^*}}^{M} q_{t-K,\theta}(\alpha_{t-K}^m \mid \alpha_{t-K-1}, y_{t-K}).
\end{aligned} \tag{58}
$$

Then we obtain

$$
\begin{aligned}
&\hat{\pi}(k_{t_0}^* \mid \theta, \alpha_{s-1:t-K-1}^{1:M}, \alpha_{t-K:t_0}^{1:M}, a_{t-K:t_0-1}^{1:M}, \alpha_{t_0+1}^{k_{t_0+1}^*}, \ldots, \alpha_t^{k_t^*}, k_{t_0+1:t}^*, y_{s-1:t}) \\
&\quad\propto \hat{\pi}(\theta, \alpha_{s-1:t-K-1}^{1:M}, \alpha_{t-K:t_0}^{1:M}, a_{t-K:t_0-1}^{1:M}, \alpha_{t_0+1}^{k_{t_0+1}^*}, \ldots, \alpha_t^{k_t^*}, k_{t_0:t}^* \mid y_{s-1:t}) \\
&\quad\propto \frac{\pi(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{k_{t-K}^*}, \ldots, \alpha_{t_0}^{k_{t_0}^*} \mid y_{s-1:t_0})}{M^{t_0-(t-K)+1}} \times \prod_{\substack{m=1 \\ m\neq k_{t-K}^*}}^{M} q_{t-K,\theta}(\alpha_{t-K}^m \mid \alpha_{t-K-1}, y_{t-K}) \\
&\quad\times \prod_{j=t-K+1}^{t_0} \prod_{\substack{m=1 \\ m\neq k_j^*}}^{M} V_{j-1}^{a_j^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_j^m}, y_j) \times \prod_{j=t_0+1}^{t} f_\theta(\alpha_j^{k_j^*} \mid \alpha_{j-1}^{k_{j-1}^*}, y_{j-1}) g_\theta(y_j \mid \alpha_j^{k_j^*}) \\
&\quad= \pi(\theta, \alpha_{s-1:t-K-1} \mid y_{s-1:t-K-1}) \times \prod_{m=1}^{M} q_{t-K,\theta}(\alpha_{t-K}^m \mid \alpha_{t-K-1}, y_{t-K}) \\
&\quad\times \prod_{j=t-K+1}^{t_0} \prod_{m=1}^{M} V_{j-1,\theta}^{a_{j-1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_{j-1}^m}, y_j) \times V_{t_0,\theta}^{k_{t_0}^*} \times \prod_{j=t-K}^{t_0} \frac{\hat{p}(y_j \mid y_{s-1:j-1}, \alpha_{t-K-1}, \theta)}{p(y_j \mid y_{s-1:j-1})}, \\
&\quad\times \prod_{j=t_0+1}^{t} f_\theta(\alpha_j^{k_j^*} \mid \alpha_{j-1}^{k_{j-1}^*}, y_{j-1}) g_\theta(y_j \mid \alpha_j^{k_j^*}) \\
&\quad\propto V_{t_0,\theta}^{k_{t_0}^*} \times f_\theta(\alpha_{t_0+1}^{k_{t_0+1}^*} \mid \alpha_{t_0}^{k_{t_0}^*}, y_{t_0}),
\end{aligned}
\tag{59}
$$

where we use Lemma 4.1 at the equality.

□

A.5 Proof of Lemma 4.2

Proof. Since

$$
\begin{aligned}
&\pi(\theta, \alpha_{s_0}^{k_{s_0}}, \ldots, \alpha_{s+K-1}^{k_{s+K-1}}, \alpha_{s+K:t} \mid y_{s_0:t}) \\
&\quad= \frac{\pi(\theta, \alpha_{s+K:t} \mid y_{s+K:t})}{p(y_{s_0:s+K-1} \mid y_{s+K:t})} \prod_{j=s_0}^{s+K-1} p(\alpha_j^{k_j} | \alpha_{j+1}^{k_{j+1}}, \theta) g_\theta(y_j | \alpha_j^{k_j}, \alpha_{j+1}^{k_{j+1}}) \\
&\quad= \frac{\pi(\theta, \alpha_{s+K:t} \mid y_{s+K:t})}{p(y_{s_0:s+K-1} \mid y_{s+K:t})} \prod_{j=s_0}^{s+K-1} v_{j,\theta}(\alpha_j^{k_j}, \alpha_{j+1}^{k_{j+1}}) \times \prod_{j=s_0}^{s+K-1} q_{j,\theta}(\alpha_j^{k_j} | \alpha_{j+1}^{k_{j+1}}, y_j),
\end{aligned}
$$

we obtain

$$
\begin{aligned}
&\frac{\pi(\theta, \alpha_{s_0}^{k_{s_0}}, \ldots, \alpha_{s+K-1}^{k_{s+K-1}}, \alpha_{s+K:t} \mid y_{s_0:t})}{M^{(s+K-1)-s_0+1}} \\
&\quad\times \prod_{\substack{m=1 \\ m\neq k_{s+K-1}}}^{M} q_{s+K-1,\theta}(\alpha_{s+K-1}^m \mid \alpha_{s+K}, y_{s+K-1}) \times \prod_{j=s_0}^{s+K-2} \prod_{\substack{m=1 \\ m\neq k_j}}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j+1}^{a_{j+1}^m}, y_j)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\pi(\theta, \alpha_{s+K:t} \mid y_{s+K:t})}{M^{(s+K-1)-s_0+1} p(y_{s_0:s+K-1} \mid y_{s+K:t})} \times \prod_{j=s_0}^{s+K-1} \prod_{m=1}^{M} q_{j,\theta}(\alpha_j^m | \alpha_{j+1}^{a_{j+1}^m}, y_j) \\
&\quad \times \prod_{j=s_0-1}^{s+K-2} v_{j+1,\theta}(\alpha_{j+1}^{k_{j+1}}, \alpha_{j+2}^{k_{j+2}}) \times \prod_{j=s_0}^{s+K-2} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j+1,\theta}^{a_{j+1}^m} \\
&= \frac{\pi(\theta, \alpha_{s+K:t} \mid y_{s+K:t})}{p(y_{s_0:s+K-1} \mid y_{s+K:t})} \times \prod_{j=s_0}^{s+K-1} \prod_{m=1}^{M} q_{j,\theta}(\alpha_j^m | \alpha_{j+1}^{a_{j+1}^m}, y_j) \\
&\quad \times \prod_{j=s_0}^{s+K-1} \left\{ \frac{1}{M} \sum_{i=1}^{M} v_{j,\theta}(\alpha_j^i, \alpha_{j+1}^{a_{j+1}^i}) \right\} \times \prod_{j=s_0}^{s+K-2} \prod_{m=1}^{M} V_{j+1,\theta}^{a_{j+1}^m} \times V_{s_0,\theta}^{k_{s_0}} \\
&= \frac{\pi(\theta, \alpha_{s+K:t} \mid y_{s+K:t})}{\prod_{j=s_0}^{s+K-1} p(y_j \mid y_{j+1:t})} \times \prod_{j=s_0}^{s+K-1} \prod_{m=1}^{M} q_{j,\theta}(\alpha_j^m | \alpha_{j+1}^{a_{j+1}^m}, y_j) \\
&\quad \times \prod_{j=s_0}^{s+K-2} \prod_{m=1}^{M} V_{j+1,\theta}^{a_{j+1}^m} \times V_{s_0,\theta}^{k_{s_0}} \times \prod_{j=s_0}^{s+K-1} \hat{p}(y_j | y_{j+1:t}, \alpha_{s+K}, \theta)
\end{aligned}
$$

□

A.6 Proof of Proposition 4.4

Proof. By applying Lemma 4.2 with $s_0 = s-1$ to the first three terms of (52), we have

$$
\begin{aligned}
&\check{\pi}(\theta, \alpha_{s-1:s+K-1}^{1:M}, \alpha_{s+K:t}, a_{s:s+K-1}^{1:M}, k_{s-1}, k_s^* \mid y_{s-1:t}) \\
&= \pi(\theta, \alpha_{s+K:t} \mid y_{s+K:t}) \times \prod_{m=1}^{M} q_{s+K-1,\theta}(\alpha_{s+K-1}^m \mid \alpha_{s+K}, y_{s+K-1:t}) \\
&\quad \times \prod_{j=s-1}^{s+K-2} \prod_{m=1}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j+1}^{a_{j+1}^m}, y_j) \times V_{s-1,\theta}^{k_{s-1}} \times \prod_{j=s-1}^{s+K-1} \frac{\hat{p}(y_j \mid y_{j+1:t}, \alpha_{s+K}, \theta)}{p(y_j \mid y_{j+1:t})} \\
&\quad \times V_s^{k_s^*} \times \frac{p(y_{s-1} \mid y_{s:t})}{\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)} \\
&= \pi(\theta, \alpha_{s+K:t} \mid y_{s+K:t}) \times \prod_{m=1}^{M} q_{s+K-1,\theta}(\alpha_{s+K-1}^m \mid \alpha_{s+K}, y_{s+K-1:t}) \\
&\quad \times \prod_{j=s}^{s+K-2} \prod_{m=1}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j+1}^{a_{j+1}^m}, y_j) \times V_s^{k_s^*} \times \prod_{j=s}^{s+K-1} \frac{\hat{p}(y_j \mid y_{j+1:t}, \alpha_{s+K}, \theta)}{p(y_j \mid y_{j+1:t})} \\
&\quad \times \prod_{m=1}^{M} V_{s,\theta}^{a_s^m} q_{s-1,\theta}(\alpha_{s-1}^m \mid \alpha_s^{a_s^m}, y_{s-1}) \times V_{s-1,\theta}^{k_{s-1}}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\pi(\theta, \alpha_s^{k_s^*}, \ldots, \alpha_{s+K-1}^{k_{s+K-1}^*}, \alpha_{s+K:t} \mid y_{s:t})}{M^K} \times \prod_{\substack{m=1 \\ m \neq k_{s+K-1}^*}}^{M} q_{s+K-1,\theta}(\alpha_{s+K-1}^m \mid \alpha_{s+K}, y_{s+K-1:t}) \\
&\quad \times \prod_{j=s}^{s+K-2} \prod_{\substack{m=1 \\ m \neq k_j^*}}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j+1}^{a_{j+1}^m}, y_j) \times \prod_{m=1}^{M} V_{s,\theta}^{a_s^m} q_{s-1,\theta}(\alpha_{s-1}^m \mid \alpha_s^{a_s^m}, y_{s-1}) \times V_{s-1,\theta}^{k_{s-1}}
\end{aligned}
$$

where we again applied Lemma 4.2 with $s_0 = s$ and $k_{s-1} = k_{s-1}^*$ in the last equality. □

A.7 Proof of Proposition 4.5

Proof. We first define the probability density function

$$
\begin{aligned}
&\bar{\psi}_{\theta,0}(\alpha_{s-1:s+K-1}^{1:M}, a_{s:s+K-1}^{1:M}, k_{s-1}, k_s^* \mid \alpha_{s+K}, y_{s-1:s+K-1}) \\
&\equiv \frac{\pi(\alpha_{s-1:s+K-1} \mid \alpha_{s+K}, y_{s-1:t}, \theta)}{M^{K+1}} \\
&\times \bar{\psi}_\theta(\alpha_{s-1}^{-k_{s-1}}, \ldots, \alpha_{s+K-1}^{-k_{s+K-1}}, a_s^{-k_{s-1}}, \ldots, a_{s+K-1}^{-k_{s+K-2}}, k_s^* \mid \alpha_{s-1:s+K}, a_{s-1}^{k_{s-2}}, \ldots, a_{s+K-1}^{k_{s+K-2}}, y_{s-1:s+K-1}),
\end{aligned}
$$

and note that

$$
\pi(\alpha_{s-1:s+K-1} \mid \alpha_{s+K}, y_{s-1:t}, \theta) = \frac{\pi(\theta, \alpha_{s-1:s+K-1}, \alpha_{s+K:t} \mid y_{s-1:t})}{\pi(\theta, \alpha_{s+K:t} \mid y_{s-1:t})}. \tag{60}
$$

Since

$$
\begin{aligned}
&\frac{1}{\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)} \bar{\psi}_{\theta,0}(\alpha_{s-1:s+K-1}^{1:M}, a_{s:s+K-1}^{1:M}, k_{s-1}, k_s^* \mid \alpha_{s+K}, y_{s-1:s+K-1}) \\
&= \check{\pi}(\theta, \alpha_{s-1:s+K-1}^{1:M}, \alpha_{s+K:t}, a_{s:s+K-1}^{1:M}, k_{s-1}, k_s^* \mid y_{s-1:t}) \frac{1}{p(y_{s-1} \mid y_{s:t}) \pi(\theta, \alpha_{s+K:t} \mid y_{s-1:t})},
\end{aligned}
$$

where we used the definition of $\check{\pi}$ in (52), we obtain

$$
\begin{aligned}
&E_{\bar{\psi}_{\theta,0}}\left[\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)^{-1} \mid \alpha_{s+K}, y_{s-1:t}, \theta\right] \\
&= \int \frac{1}{\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)} \bar{\psi}_{\theta,0}(\alpha_{s-1:s+K-1}^{1:M}, a_{s:s+K-1}^{1:M}, k_{s-1}, k_s^* \mid \alpha_{s+K}, y_{s-1:s+K-1}) \\
&\qquad\qquad d\alpha_{s-1:s+K-1}^{1:M} da_{s:s+K-1}^{1:M} dk_{s-1} dk_s^* \\
&= \int \check{\pi}(\theta, \alpha_{s-1:s+K-1}^{1:M}, \alpha_{s+K:t}, a_{s:s+K-1}^{1:M}, k_{s-1}, k_s^* \mid y_{s-1:t}) d\alpha_{s-1:s+K-1}^{1:M} da_{s:s+K-1}^{1:M} dk_{s-1} dk_s^* \\
&\quad \times \frac{1}{p(y_{s-1} \mid y_{s:t}) \pi(\theta, \alpha_{s+K:t} \mid y_{s-1:t})} \\
&= \frac{\pi(\theta, \alpha_{s+K:t} \mid y_{s:t})}{p(y_{s-1} \mid y_{s:t}) \pi(\theta, \alpha_{s+K:t} \mid y_{s-1:t})} = \frac{1}{p(y_{s-1} \mid y_{s:t}, \alpha_{s+K:t}, \theta)} = \frac{1}{p(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)}
\end{aligned}
$$

where we use Proposition 4.4 in the third equality. Further,

$$
\begin{aligned}
E[p(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)^{-1} \mid y_{s-1:t}] &= \int \frac{\pi(\theta, \alpha_{s+K} \mid y_{s-1:t})}{p(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)} d\alpha_{s+K} d\theta \\
&= \int \frac{\pi(\theta, \alpha_{s+K} \mid y_{s:t})}{p(y_{s-1} \mid y_{s:t})} d\alpha_{s+K} d\theta = p(y_{s-1} \mid y_{s:t})^{-1}.
\end{aligned}
$$

□