CIRJE-F-1126

Particle Rolling MCMC

Naoki Awaya
Graduate School of Economcis, The University of Tokyo

Yasuhiro Omori
The University of Tokyo

September 2019

Particle rolling MCMC

Naoki Awaya* and Yasuhiro Omori†

September 27, 2019

Abstract

An efficient simulation-based methodology is proposed for the rolling window estimation of state space models, called particle rolling Markov chain Monte Carlo (MCMC) with double block sampling. In our method, which is based on Sequential Monte Carlo (SMC), particles are sequentially updated to approximate the posterior distribution for each window by learning new information and discarding old information from observations. Th particles are refreshed with an MCMC algorithm when the importance weights degenerate. To avoid degeneracy, which is crucial for reducing the computation time, we introduce a block sampling scheme and generate multiple candidates by the algorithm based on the conditional SMC. The theoretical discussion shows that the proposed methodology with a nested structure is expressed as SMC sampling for the augmented space to provide the justification. The computational performance is evaluated in illustrative examples, showing that the posterior distributions of the model parameters are accurately estimated. The proofs and additional discussions (algorithms and experimental results) are provided in the Supplementary Material.

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. These models are flexible and capture the dynamics of the complex economic structure. However, several structural changes have been noted in long-term economic series. 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, we use the rolling window estimation where we fix the number of observations to estimate model parameters and update the dataset to improve the forecasting performance.

In non-linear or non-Gaussian state space models, the likelihood is often not obtained analytically and the maximum likelihood estimation is difficult to implement. The Markov Chain Monte Carlo (MCMC) method is a popular and powerful technique used 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. However, for rolling estimation, simply applying the MCMC method would be too time-consuming given the need to estimate numerous (e.g. thousands) posterior distributions.

To overcome this difficulty, we take an alternative approach based on the sequential Monte Carlo (SMC) method (e.g. Doucet et al. (2001)). This is effective because, in the rolling window estimation, we can utilize the weighted samples from one posterior distribution to approximate the next posterior distribution instead of iterating the same MCMC algorithm. The particles consist of realized values of state variables and static parameters, which are updated by incorporating new observation and discarding the oldest observation. As noted in the numerical experiment in Section 2, the updating step should be constructed carefully since a simple method that could be directly derived from the previous literature leads to the severe weight degeneracy problem. Hence we adopt the idea of block sampling (e.g. Doucet et al. (2006), Polson et al. (2008)), in which state variables at multiple time points are updated when learning new information. This framework is highly efficient in the sense that it substantially increases the effective sample size. We also utilize the same idea to discard old information. However, unless the time series model has a relatively simple form, finding an appropriate proposal distribution for these update steps may be difficult. Hence, instead of generating only one candidate from the proposal distribution, we generate multiple candidates and choose one of them using the algorithm of the conditional SMC of Particle

MCMC (PMCMC) (Andrieu et al. (2010)). This nested structure is similar to that of SMC^2 (Chopin et al. (2013), Fulop and Li (2013)) and nested SMC (Naesseth et al. 2015), but our proposed algorithm differs in that it is derived from Particle Gibbs instead of Particle MH (Metropolis-Hastings) algorithm.

We can obtain an algorithm for the ordinary sequential analysis, in which we simply incorporate new information sequentially, as a special case of our new method. This algorithm contrasts with SMC^2 in that it originates from different PMCMC algorithms. Of note, Whiteley et al. (2010) employed PMCMC for a Markov switching state-space model that includes structural changes with a finite number of states.

The remainder 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 method directly derived from the conventional filtering algorithm causes the serious weight degeneracy phenomenon. Section 3 introduces a new methodology to overcome this difficulty. Theoretical justifications of the proposed method are provided in Section 4. Section 5 provides illustrative examples, and 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, a state equation with an observation vector y_t, and 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 $\alpha_{s:t} \equiv (\alpha_s, \alpha_{s+1}, \dots, \alpha_t)$ and $y_{s:t} \equiv (y_s, y_{s+1}, \dots, y_t)$. We assume that the distribution of y_t given $(y_{1:t-1}, \alpha_{1:t}, \theta)$ depends exclusively 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 noted as follows:

$$
\begin{aligned}
p(y_t \mid y_{1:t-1}, \alpha_{1:t}, \theta) &= p(y_t \mid \alpha_t, \theta) \equiv g_\theta(y_t \mid \alpha_t), \quad t = 1, \dots, n, &(1)\\
p(\alpha_t \mid \alpha_{1:t-1}, \theta) &= p(\alpha_t \mid \alpha_{t-1}, \theta) \equiv f_\theta(\alpha_t \mid \alpha_{t-1}), \quad t = 2, \dots, n, &(2)
\end{aligned}
$$

where $p(\alpha_1 \mid \theta) \equiv \mu_\theta(\alpha_1)$ denotes a known density function of the stationary distribution given θ. We also consider the correlation between y_t and α_{t+1}, which is conditional on α_t, as demonstrated in our empirical studies of the financial time series. Thus we formulate the process as follows:

$$p(y_t \mid y_{1:t-1}, \alpha_{1:t+1}, \theta) = p(y_t \mid \alpha_t, \alpha_{t+1}, \theta) \equiv g_\theta(y_t \mid \alpha_t, \alpha_{t+1}), \quad t = 1, \ldots, n, \tag{3}$$

$$p(\alpha_{t+1} \mid \alpha_{1:t}, y_{1:t}, \theta) = p(\alpha_{t+1} \mid \alpha_t, y_t, \theta) \equiv f_\theta(\alpha_{t+1} \mid \alpha_t, y_t), \quad t = 1, \ldots, n-1. \tag{4}$$

In the rolling window estimation of time series, the number of observations (or the window size) in the sample period is fixed and is set equal to, e.g., $L+1$. 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 as follows:

$$\pi(\theta, \alpha_{s:t} \mid y_{s:t}) \propto p(\theta)\mu_\theta(\alpha_s) g_\theta(y_s \mid \alpha_s) \left\{ \prod_{j=s+1}^{t} f_\theta(\alpha_j \mid \alpha_{j-1}, y_{j-1}) g_\theta(y_j \mid \alpha_j) \right\} \tag{5}$$

$$\propto p(\theta)\mu_\theta(\alpha_s) \left\{ \prod_{j=s+1}^{t} f_\theta(\alpha_j \mid \alpha_{j-1}) g_\theta(y_{j-1} \mid \alpha_j, \alpha_{j-1}) \right\} g_\theta(y_t \mid \alpha_t). \tag{6}$$

2.2 Simple particle rolling MCMC

One of most straightforward approaches to construct the particle rolling estimation is to use the idea of resample-move algorithms (e.g. Gilks and Berzuini (2001)). In other words, we basically repeat moving the data window by generating new values for a state variable at the new time point and updating the importance weights. Additionally, we refresh all the particles with the MCMC method when we observe the weight degeneracy. The procedure is described in Algorithm 1.

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 $\pi(\theta, \alpha_{s-1:t-1} \mid y_{s-1:t-1})$. In Step 1, we follow the standard algorithm to learn the new information. Specifically, we generate α^n_t using some proposal density $q_{t,\theta^n}(\cdot \mid \alpha^n_{t-1}, y_t)$ and compute the importance weight

$$W^n_{s-1:t} \propto \frac{f_{\theta^n}(\alpha^n_t \mid \alpha^n_{t-1}, y_{t-1}) g_{\theta^n}(y_t \mid \alpha^n_t)}{q_{t,\theta^n}(\alpha^n_t \mid \alpha^n_{t-1}, y_t)} W^n_{s-1:t-1}, \tag{7}$$

where we add a new observation y_t to the information set. Then we compute some degeneracy criteria, such as the effective sample size (ESS), which is defined as follows.

$$ESS_{s-1:t} = \frac{1}{\sum_{n=1}^{N} (W^n_{s-1:t})^2} \tag{8}$$

Then, the particles are resampled if ESS $< cN$ (e.g. $c = 0.5$). In Step 2, we set the target density to be $p(\tilde{\alpha}_{s-1} \mid \tilde{\alpha}_s, \tilde{\theta})\pi(\tilde{\theta}, \tilde{\alpha}_{s:t} \mid y_{s:t})$, where

$$p(\alpha_{s-1} \mid \alpha_s, \theta) = \frac{\mu_\theta(\alpha_{s-1}) f_\theta(\alpha_s \mid \alpha_{s-1})}{\mu_\theta(\alpha_s)}. \tag{9}$$

Then, we update the importance weight as follows:

$$W_{s:t} \propto g_{\tilde{\theta}}(y_{s-1} \mid \tilde{\alpha}_{s-1}, \tilde{\alpha}_s)^{-1} W_{s-1:t}, \tag{10}$$

where the incremental weight is derived from

$$\begin{aligned} \frac{p(\tilde{\alpha}_{s-1} \mid \tilde{\alpha}_s, \tilde{\theta})\pi(\tilde{\theta}, \tilde{\alpha}_{s:t} \mid y_{s:t})}{\pi(\tilde{\theta}, \tilde{\alpha}_{s-1:t} \mid y_{s-1:t})} &= \frac{p(\tilde{\alpha}_{s-1} \mid \tilde{\alpha}_s, \tilde{\theta})\mu_{\tilde{\theta}}(\tilde{\alpha}_s)}{\mu_{\tilde{\theta}}(\tilde{\alpha}_{s-1}) f_{\tilde{\theta}}(\tilde{\alpha}_s \mid \tilde{\alpha}_{s-1}) g_{\tilde{\theta}}(y_{s-1} \mid \tilde{\alpha}_{s-1}, \tilde{\alpha}_s)} \\ &= g_{\tilde{\theta}}(y_{s-1} \mid \tilde{\alpha}_{s-1}, \tilde{\alpha}_s)^{-1}. \end{aligned} \tag{11}$$

Then, we discard α^n_{s-1}. If some degeneracy criteria are fulfilled, resample all the particles by implementing MCMC algorithm as noted in Step 1.

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

Generate α^n_t given $(\theta^n, \alpha^n_{s-1:t-1})$ using some proposal distribution 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 $\pi(\theta, \alpha_{s-1:t} \mid y_{s-1:t})$. If some degeneracy criteria are fulfilled, resample all the particles and set $W^n_{s-1:t} = 1/N$. Further, update $(\theta^n, \alpha^n_{s-1:t})$ 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: *Remove the oldest observation y_{s-1} from the information set.*

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

Table 1: **Algorithm 1: Simple particle rolling MCMC**

Remark 1. In the additional MCMC implementation of Step 1, we update all the particles with the transition kernel K (called MCMC kernel) which satisfies

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

where we suppress n for convenience. This update step can also be regarded as an importance sampling step in SMC (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

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

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 follows:

$$\tilde{\pi}_2(x_1, x_2) = \pi_2(x_2)L(x_2, x_1), \quad L(x_2, x_1) = \frac{\pi_1(x_1)K(x_1, x_2)}{\pi_2(x_2)}. \tag{12}$$

Here, the so-called (unnormalized) incremental weight $\tilde{w}_2$ is defined as follows:

$$\tilde{w}_2(x_1, x_2) = \frac{\pi_2(x_2)L(x_2, x_1)}{\pi_1(x_1)K(x_1, x_2)} = 1. \tag{13}$$

Hence, there is no change in the weight $W_{s-1:t}$ 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 using Particle Gibbs sampler (Andrieu et al. (2010)), which leaves the artificial target distribution invariant in the augmented space and the posterior distribution $\pi_2(x_2)$ is obtained as its marginal distribution.

Remark 2. Note that Equation (10) in Step 2 includes the inverse of the likelihood, g. The weight is expected to be very unstable when g is close to zero. It will cause the weight degeneracy problem and the corresponding posterior distribution would have heavier tails than the proposal distribution. The next subsection will show this phenomenon in more details using a real data example.

2.3 Serious weight degeneracy when removing observations

Using the importance weights in Step 2 is obviously problematic because they would take extremely high values when $g_{\tilde{\theta}}$ is close to 0. This causes the ESS to rapidly drop and triggers the MCMC update steps many times, which is time-consuming. On the other hand, in Step 1, one might think it will work without any problem as long as we choose an appropriate proposal distribution $q_{t,\theta}$. However, as noted in the numerical example below, this step also causes a problem similar to that noted in Step 2. To illustrate these weight degeneracy problems, we consider the rolling estimation of the realized stochastic volatility (RSV) model for the financial time series (additional illustrative examples using the linear Gaussian state space model and the RSV model are given in the Supplementary Material). The RSV model is a stochastic volatility model with an additional measurement equation for the realized volatility (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. Let α_t denote the latent log volatility which is assumed to follow the AR(1) process. The RSV model is defined as follows:

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

where

$$
\begin{pmatrix} \epsilon_t \\ u_t \\ \eta_t \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & \rho\sigma_\eta \\ 0 & \sigma_u^2 & 0 \\ \rho\sigma_\eta & 0 & \sigma_\eta^2 \end{bmatrix} \right). \tag{14}
$$

The correlation between ϵ_t and η_t is introduced to express the leverage effect. The effect is often negative in empirical studies, which implies that the decrease in the today's log return is followed by the increase in the log volatility on the next day. The static parameters are unknown. Thus, $\theta = (\mu, \phi, \sigma_\eta^2, \xi, \sigma_u^2, \rho)'$, where we assume the same prior distributions as noted in Takahashi et al. (2009). Thus, we consider the following transformation:

$$
\sigma_\epsilon = \exp(\mu/2),\ c = \xi + \mu,\ \Sigma = \begin{bmatrix} \sigma_\epsilon^2 & \rho\sigma_\epsilon\sigma_\eta \\ \rho\sigma_\epsilon\sigma_\eta & \sigma_\eta^2 \end{bmatrix}. \tag{15}
$$

We assume the following:

$$
\frac{\phi+1}{2} \sim Beta(20, 1.5),\ c \sim N(0,10),\ \sigma_u^2 \sim IG(5/2, 0.05/2), \tag{16}
$$

$$
\Sigma \sim IW(5, \Sigma_0),\ \Sigma_0 = \left(5 \begin{bmatrix} 1 & -0.3\sqrt{1\cdot 0.01} \\ -0.3\sqrt{1\cdot 0.01} & 0.01 \end{bmatrix} \right)^{-1}. \tag{17}
$$

For y_{1t} and y_{2t}, we use Standard and Poor's (S&P) 500 index data, which are obtained from the 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

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

($t = 1$) 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, 2008 ($T = 2248$). Thus the first estimation period is before the financial crisis caused by the bankruptcy of Lehman Brothers and the last estimation period includes the crisis.

If the ESS is less than the threshold ($0.5 \times N$), the particles are refreshed with the MCMC update 10 times. (see Takahashi et al. (2009) for the details of MCMC sampling). We set $N = 1000$ and construct the proposal density $q_{t,\theta}(\alpha_t \mid \alpha_{t-1}, y_{s-1:t})$ based on the normal mixture approximation (see Omori et al. (2007)), which is expected to improve the weight degeneracy. To evaluate the weight degeneracy in each of Steps 1 and 2, we define two ratios:

$$R_{1t} = \frac{ESS_{s-1:t}}{ESS_{s-1:t-1}}, \quad R_{2t} = \frac{ESS_{s:t}}{ESS_{s-1:t}}. \tag{18}$$

The ratio R_{1t} measures the relative magnitude of ESS in Step 1 after adding a new observation when compared with that of the previous step. If the distribution of particles is close to the posterior distribution from which we aim to sample in the step, R_{1t} would be close to 1. On the other hand, in the presence of the weight degeneracy problem, it will be close to 0. Similarly, the ratio R_{2t} measures the relative magnitude of ESS in Step 2 after removing the oldest observation compared with that of the previous step.

	Mean	Median	Std. dev.
R_{1t}	0.837	0.912	0.193
R_{2t}	0.227	0.197	0.176

Table 2: Summary statistics for R_{1t} and R_{2t} ($t = 1988, \ldots, 2248$).

Table 2 presents a summary of computed R_{1t} and R_{2t}. As expected, R_{2t}'s takes low values, so the update with MCMC kernel should be implemented in almost every step. The results for R_{1t}'s also show that the ESS will be often less than the threshold to resample all the particles. In fact, due to these problems, the resampling steps are implemented 271 times for 260 data windows.

3 Particle rolling MCMC with double block sampling

To overcome this difficulty of the weight degeneracy, we propose a novel sampling approach called, Particle rolling MCMC with double block sampling. First, we consider incorporating the new information y_t. Our method is a combination of the block sampling and the conditional SMC update as follows. (1) We sample a block of state variables when we add the new observation. In other words, we update values of $\{\alpha^n_{t-K:t-1}\}_{n=1}^N$ in addition to generating $\{\alpha^n_t\}_{n=1}^N$ when we learn the information of y_t. We call this process forward block sampling. The blocking method addresses the weight degeneracy problem by reducing the path dependence between the new particle and the old particle values that are not updated, in the context of particle filtering with the known parameter θ (Doucet et al. (2006)). However, we cannot directly adopt the block sampling here because it is often difficult to find an appropriate K-dimensional proposal distribution to update $\{\alpha^n_{t-K:t-1}\}_{n=1}^N$ including the stochastic volatility model. Hence, (2) we adopt the approach of the conditional SMC update (Andrieu et al. (2010)). For the n-th particle, we generate multiple (M, say) candidates $\{\alpha^{n,m}_{t-K:t}\}_{m=1}^M$ with the fixed previous values of $\alpha^n_{t-K:t-1}$. Among the candidates $\{\alpha^{n,m}_{t-K:t}\}_{m=1}^M$, we randomly select one path to store the next values for $\alpha^n_{t-K:t}$.

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

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 $W^n_{s-1:t}$ $(n = 1, \dots, N)$ to approximate the posterior distribution with the density $\pi(\theta, \alpha_{s-1:t} \mid y_{s-1:t})$. The particle simulation smoother may be implemented to improve the mixing property. If some degeneracy criteria are 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 $\pi(\theta, \alpha_{s-1:t} \mid y_{s-1:t})$, $(n = 1, \dots, N)$.

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

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 $W^n_{s:t}$ $(n = 1, \dots, N)$ 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 may be implemented to improve the mixing property. If some degeneracy criteria are 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 $\pi(\theta, \alpha_{s:t} \mid y_{s:t})$, $(n = 1, \dots, N)$.

Table 3: **Algorithm 2: Particle rolling MCMC with double block sampling**

Second, to discard the old information y_{s-1}, we follow a similar but slightly different procedure. We not only discard $\{\alpha_{s-1}\}_{n=1}^N$ but also update $\{\alpha_{s:s+K-1}^n\}_{n=1}^N$. A set of $K+1$ state variables $\{\alpha_{s-1:s+K-1}^n\}$ are sampled as a block in a similar manner, but we construct a candidate path sequentially from α_{s+K-1}^n to α_{s-1}^n in the opposite direction. We call it the backward block sampling, and the general algorithm is described in Table 3.

3.1 Forward block sampling (Step 1)

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 numerous 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) involve a nested SMC algorithm that 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.

This subsection describes the procedure to obtain the n-th particle when including the new information y_t. We first generate a number of candidates $\alpha_{t-K:t-1}^{n,m}$ $(m=1,\dots,M)$ with the current values $\alpha_{t-K:t-1}^n$ fixed using the conditional SMC. Then, for each $\alpha_{t-K:t-1}^{n,m}$, we generate $\alpha_t^{n,m}$. In this 'local particle filtering', we resample the particles at $j=t-K+1,\dots,t$. This operation is equivalent to choosing the 'parent' $\alpha_j^{n,m}$ for $\alpha_{j+1}^{n,m}$. Using this terminology, if we choose one particle $\alpha_t^{n,m}$, its 'ancestors' are uniquely determined from $\alpha_j^{n,m}$ $(j=t-K,\dots,t-1)$. We call this descendant and its ancestors the 'lineage'. In the conditional SMC step, fixing the current values $\alpha_{t-K:t-1}^n$ is seen as fixing one lineage by choosing their indices k_j $(j=t-K,\dots,t-1)$ (where we drop the superscript n for simplicity) which follows the rule

$$a_j^{k_{j+1}} = k_j, \quad j=t-K,\dots,t-1. \tag{19}$$

In addition, the index of their descendant is determined as $k_t=1$.

After generating $\alpha_{t-K:t}^{n,m}$ $(m=1,\dots,M)$, we choose one lineage to store as the next values of $\alpha_{t-K:t}^n$. This is equivalent to sampling a random index k_t^* for the candidate $\alpha_t^{n,m}$ and identifying the ancestors for which indices are obtained by following the rule

$$a_j^{k_{j+1}^*} = k_j^*, \quad j=t-K,\dots,t-1. \tag{20}$$

Moreover, we can improve its efficiency by implementing 'smoothing' for the generated candidates following the algorithm reported in Whiteley et al. (2010). In this smoothing step, we again choose k_j^* for $j = t-K, \ldots, t-1$ randomly. This manipulation of breaking the relationship between the parent and the child in the lineage is effective in improving the mixing property, or sampling values of $\alpha_{t-K:t-1}^n$ that may be different from the lineages obtained in the previous step.

The detailed algorithm is provided below. We fix one lineage in (1) and implement the conditional SMC in (2) and (3). The candidates for α_t^n are generated in (4) and the smoothing is implemented in (6). In parallel, we compute the importance weight for the n-th particle in the 'global particle filtering' in (5).

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

$$(\alpha_{t-K}^{n,k_{t-K}}, \ldots, \alpha_{t-1}^{n,k_{t-1}}) = \alpha_{t-K:t-1}^n, \quad (a_{t-K}^{k_{t-K+1}}, \ldots, a_{t-1}^{k_t}) = (k_{t-K}, \ldots, k_{t-1}),$$

where $\alpha_{t-K:t-1}^n$ is a current sample with the importance weight $W_{s-1:t-1}^n$.

(2) Set $\alpha_{t-K-1}^{n,a_{t-K-1}^m} = \alpha_{t-K-1}^n$ for all m according to the convention, and sample $\alpha_{t-K}^{n,m} \sim q_{t-K,\theta^n}(\cdot \mid \alpha_{t-K-1}^n, y_{t-K})$ 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 $m \in \{1, \ldots, M\} \setminus \{k_j\}$ where $V_{j-1,\theta^n}^{1:M} \equiv (V_{j-1,\theta^n}^1, \ldots, V_{j-1,\theta^n}^M)$ and

$$V_{j,\theta^n}^m = \frac{v_{j,\theta^n}(\alpha_{j-1}^{n,a_{j-1}^m}, \alpha_j^{n,m})}{\sum_{i=1}^M v_{j,\theta^n}(\alpha_{j-1}^{n,a_{j-1}^i}, \alpha_j^{n,i})}, \tag{21}$$

$$v_{j,\theta^n}(\alpha_{j-1}^{n,a_{j-1}^m}, \alpha_j^{n,m}) = \frac{f_{\theta^n}(\alpha_j^{n,m} \mid \alpha_{j-1}^{n,a_{j-1}^m}, y_{j-1}) g_{\theta^n}(y_j \mid \alpha_j^{n,m})}{q_{j,\theta^n}(\alpha_j^{n,m} \mid \alpha_{j-1}^{n,a_{j-1}^m}, y_j)}, \quad m = 1, \ldots, M. \tag{22}$$

(4) If $j < t-1$, set $j \leftarrow j+1$ and go to (3). Otherwise, sample $\alpha_t^{n,m}$ ($m = 1, \ldots, M$) 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 $m \in \{2, \ldots, M\}$.

(iii) Sample $k_t^* \sim \mathcal{M}(V_{t,\theta^n}^{1:M})$ and obtain k_j^* ($j = t-1, \ldots, t-K$) using (20).

(5) Let $\alpha^n_{s-1:t} = (\alpha^n_{s-1}, \ldots, \alpha^n_{t-K-1}, \alpha^{n,k^*_{t-K}}_{t-K}, \ldots, \alpha^{n,k^*_t}_t)$ and compute the importance weight[2]

$$W^n_{s-1:t} \propto \hat{p}(y_t \mid y_{s-1:t-1}, \alpha^n_{t-K-1}, \theta^n) \times W^n_{s-1:t-1}. \tag{23}$$

$$\hat{p}(y_t \mid y_{s-1:t-1}, \alpha^n_{t-K-1}, \theta^n) = \frac{1}{M}\sum_{m=1}^{M} v_{t,\theta^n}(\alpha^{n,a^m_{t-1}}_{t-1}, \alpha^{n,m}_t). \tag{24}$$

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

$$\bar{V}^m_{j,\theta} \equiv \frac{V^m_{j,\theta} f_\theta(\alpha^{k^*_{j+1}}_{j+1} \mid \alpha^m_j, y_{j+1})}{\sum_{i=1}^M V^i_{j,\theta} f_\theta(\alpha^{k^*_{j+1}}_{j+1} \mid \alpha^i_j, y_{j+1})}, \quad m = 1, \ldots, M, \tag{25}$$

and set $\alpha^n_{s-1:t} = (\alpha^n_{s-1}, \ldots, \alpha^n_{t-K-1}, \alpha^{n,k^*_{t-K}}_{t-K}, \ldots, \alpha^{n,k^*_t}_t)$.

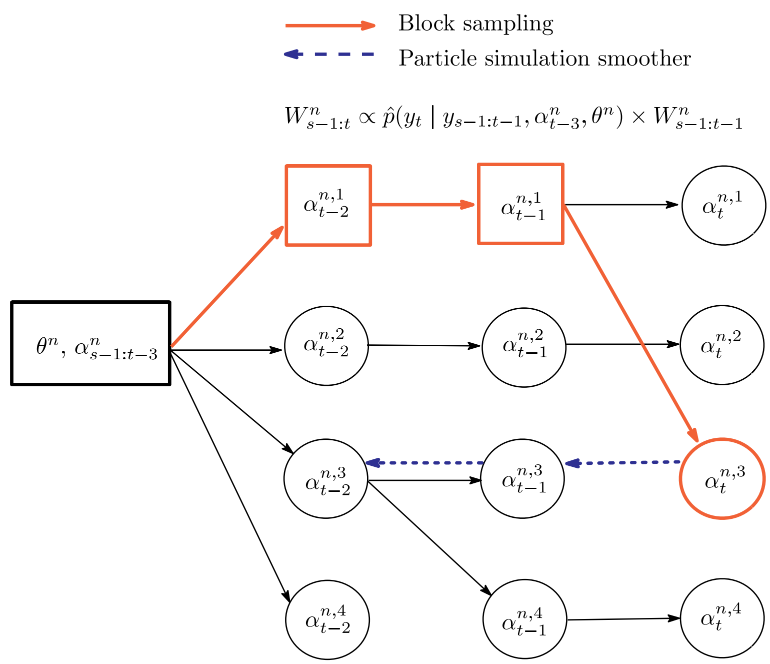

Figure 1: Forward block sampling and particle simulation smoother.

Figure 1 illustrates an example with $K = 2$, $M = 4$ and the current sample $(\theta^n, \alpha^n_{s-1:t-1})$.

(1) Sample k_{t-2} and k_{t-1} from $1:4$ with probability $1/4$ and suppose $k_{t-2} = k_{t-1} = 1$. We set $\alpha^{n,1}_{t-2} = \alpha^n_{t-2}$, $\alpha^{n,1}_{t-1} = \alpha^n_{t-1}$ (with the red rectangle) and $(a^1_{t-2}, a^1_{t-1}) = (1, 1)$.

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

(2) Set $\alpha_{t-3}^{n,a_{t-3}^m} = \alpha_{t-3}^n$ for all m (with the black rectangle), and sample $\alpha_{t-2}^{n,m} \sim q_{t-2,\theta^n}(\cdot \mid \alpha_{t-3}^n, y_{t-2})$ for each $m \in \{2,3,4\}$ (with the black circle).

(3) Sample $a_{t-2}^m \sim \mathcal{M}(V_{t-2,\theta^n}^{1:4})$ for $m \in \{2,3,4\}$ and suppose $a_{t-2}^2 = 2$, $a_{t-2}^3 = 3$, $a_{t-2}^4 = 3$. Generate $\alpha_{t-1}^{n,m} \sim q_{t-1,\theta^n}(\cdot \mid \alpha_{t-2}^{n,a_{t-2}^m}, y_{t-1})$ for $m \in \{2,3,4\}$ (with the black circle).

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

(ii) Sample $a_{t-1}^{n,m} \sim \mathcal{M}(V_{t-1,\theta^n}^{1:4})$ for $m \in \{2,3,4\}$ and suppose $a_{t-1}^2 = 2$, $a_{t-1}^3 = 1$, $a_{t-1}^4 = 4$. Generate and $\alpha_t^{n,m} \sim q_{t,\theta^n}(\cdot \mid \alpha_{t-1}^{n,a_{t-1}^m}, y_t)$ for $m \in \{2,3,4\}$.

(iii) Sample $k_t^* \sim \mathcal{M}(V_{t,\theta^n}^{1:4})$ and suppose $k_t^* = 3$. Using (20), we obtain $k_{t-1}^* = k_{t-2}^* = 1$ and select $(\alpha_t^{n,3}, \alpha_{t-1}^{n,1}, \alpha_{t-2}^{n,1})$ with red lines.

(5) Let $\alpha_{s-1:t}^n = (\alpha_{s-1}^n, \ldots, \alpha_{t-3}^n, \alpha_{t-2}^{n,1}, \alpha_{t-1}^{n,1}, \alpha_t^{n,3})$ and compute the importance weight.

(6) Implement the particle simulation smoother to sample $(k_{t-2}^*, k_{t-1}^*, k_t^*)$ jointly. Generate $k_j^* \sim \mathcal{M}(\bar{V}_{j,\theta}^{1:4})$, $j = t-1, t-2$, recursively (with dotted lines) and suppose $k_{t-1}^* = 3$ and $k_{t-2}^* = 3$. We set $\alpha_{s-1:t}^n = (\alpha_{s-1}^n, \ldots, \alpha_{t-3}^n, \alpha_{t-2}^{n,3}, \alpha_{t-1}^{n,3}, \alpha_t^{n,3})$.

Remark 3. As the proposal density $q_{j,\theta}$, we can either use the prior density f_θ or more sophisticated density that incorporates the information of the likelihood g_θ. Even if we use the prior f_θ as the proposal, the above sampling becomes much more efficient than the simple particle rolling algorithm as shown in Section 5.

3.2 Backward block sampling (Step 2)

Before we describe the backward block sampling which generates a cloud of particles based on $(\alpha_{s+K:t}^n, \theta^n)$, we define the notation for the particle index as noted in the forward block sampling but in the reverse order. A 'parent' particle of α_j^m is chosen from $\alpha_{j+1}^{1:M}$ (not from $\alpha_{j-1}^{1:M}$) and consequently a_{j+1}^m denotes its parent's index. In this case, the relationship of a_{j+1}^m and k_j is given as follows:

$$a_{j+1}^{k_j} = k_{j+1}, \quad j = s+K-2, \ldots, s-2. \tag{26}$$

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

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

$$(\alpha_{s-1}^{n,k_{s-1}}, \ldots, \alpha_{s+K-1}^{n,k_{s+K-1}}) = \alpha_{s-1:s+K-1}^{n}, \quad (a_{s-1}^{k_{s-2}}, \ldots, a_{s+K-1}^{k_{s+K-2}}) = (k_{s-1}, \ldots, k_{s+K-1}),$$

where $\alpha_{s-1:s+K-1}^{n}$ is a current sample with the importance weight $W_{s-1:t}^{n}$.

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

(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 $m \in \{1, \ldots, M\} \setminus \{k_j\}$ where $V_{j+1,\theta^n}^{1:M} = (V_{j+1,\theta^n}^1, \ldots, V_{j+1,\theta^n}^M)$ and

$$V_{j,\theta^n}^m = \frac{v_{j,\theta^n}(\alpha_j^{n,m}, \alpha_{j+1}^{n,a_{j+1}^m})}{\sum_{i=1}^M v_{j,\theta^n}(\alpha_j^{n,i}, \alpha_{j+1}^{n,a_{j+1}^i})}, \tag{27}$$

$$v_{j,\theta^n}(\alpha_j^{n,m}, \alpha_{j+1}^{n,a_{j+1}^m}) = \frac{p(\alpha_j^{n,m} \mid \alpha_{j+1}^{n,a_{j+1}^m}, \theta) g_{\theta^n}(y_j \mid \alpha_j^{n,m}, \alpha_{j+1}^{n,a_{j+1}^m})}{q_{j,\theta^n}(\alpha_j^{n,m} \mid \alpha_{j+1}^{n,a_{j+1}^m}, y_j)}, \quad m = 1, \ldots, M. \tag{28}$$

(4) If $j > s-1$, set $j \leftarrow j-1$ and go to (3). Otherwise, sample $k_s^* \sim \mathcal{M}(V_{s,\theta^n}^{1:M})$ and obtain k_j^* ($j = s+1, \ldots, s+K-1$) using (26).

(5) Let $\alpha_{s:t}^n = (\alpha_s^{n,k_s^*}, \ldots, \alpha_{s+K-1}^{n,k_{s+K-1}^*}, \alpha_{s+K}^n, \ldots, \alpha_t^n)$ and compute its importance weight

$$W_{s:t}^n \propto \begin{cases} \frac{1}{\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}^n, \theta^n)} W_{s-1:t}^n, & \text{if } \hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}^n, \theta^n) \neq 0, \\ 0, & \text{if } \hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}^n, \theta^n) = 0, \end{cases} \tag{29}$$

where

$$\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}^n, \theta^n) \;=\; \frac{1}{M} \sum_{m=1}^M v_{s-1,\theta^n}(\alpha_{s-1}^{n,m}, \alpha_s^{n,a_s^m}). \tag{30}$$

(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

$$\bar{V}_{j,\theta^n}^m = \frac{V_{j,\theta^n}^m p(\alpha_{j-1}^{k_{j-1}^*} \mid \alpha_j^m, \theta^n)}{\sum_{i=1}^M V_{j,\theta^n}^i p(\alpha_{j-1}^{k_{j-1}^*} \mid \alpha_j^i, \theta^n)}, \quad m = 1, \ldots, M. \tag{31}$$

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 2 illustrates an example with $K = 2$, $M = 4$ and the current sample $(\theta^n, \alpha_{s-1:t}^n)$.

(1) Sample indices k_{s+1}, k_s, k_{s-1} from $1:4$ with probability $1/4$ and suppose $k_{s+1} = 1$, $k_s = 1$, $k_{s-1} = 1$. We set $(\alpha_{s-1}^{n,1}, \alpha_s^{n,1}, \alpha_{s+1}^{n,1}) = \alpha_{s-1:s+1}^n$ (with the rectangle) and $(a_{s-1}^1, a_s^1, a_{s+1}^1) = (1, 1, 1)$.

(2) Set $\alpha_{s+2}^{n,a_{s+2}^m} = \alpha_{s+2}^n$ for all m (with the thick black rectangle), and sample $\alpha_{s+1}^{n,m} \sim q_{s+1,\theta^n}(\cdot \mid \alpha_{s+1}^n, y_{s+1})$ for $m \in \{2, 3, 4\}$ (with the black circle).

(3) Sample $a_{s+1}^m \sim \mathcal{M}(V_{s+1,\theta^n}^{1:4})$ and suppose $a_{s+1}^2 = 1$, $a_{s+1}^3 = 3$, $a_{s+1}^4 = 3$. Generate $\alpha_s^{n,m} \sim q_{s,\theta^n}(\cdot \mid \alpha_{s+1}^{n,a_{s+1}^m}, y_s)$ for $m \in \{2, 3, 4\}$.

(4) Sample $k_s^* \sim \mathcal{M}(V_{s,\theta^n}^{1:4})$ and suppose $k_s^* = 2$. Using (26), we obtain $k_{s+1}^* = 1$, and select $(\alpha_s^{n,2}, \alpha_{s+1}^{n,1})$ with red lines.

(5) Let $\alpha_{s:t}^n = (\alpha_s^{n,2}, \alpha_{s+1}^{n,1}, \alpha_{s+2}^n, \ldots, \alpha_t^n)$ and compute its importance weight.

(6) Implement the particle simulation smoother to sample (k_s^*, k_{s+1}^*) jointly. Generate $k_{s+1}^* \sim \mathcal{M}(\bar{V}_{s+1,\theta^n}^{1:4})$, and suppose $k_{s+1}^* = 2$. We set $\alpha_{s:t}^n = (\alpha_s^{n,2}, \alpha_{s+1}^{n,2}, \alpha_{s+2}^n, \ldots, \alpha_t^n)$.

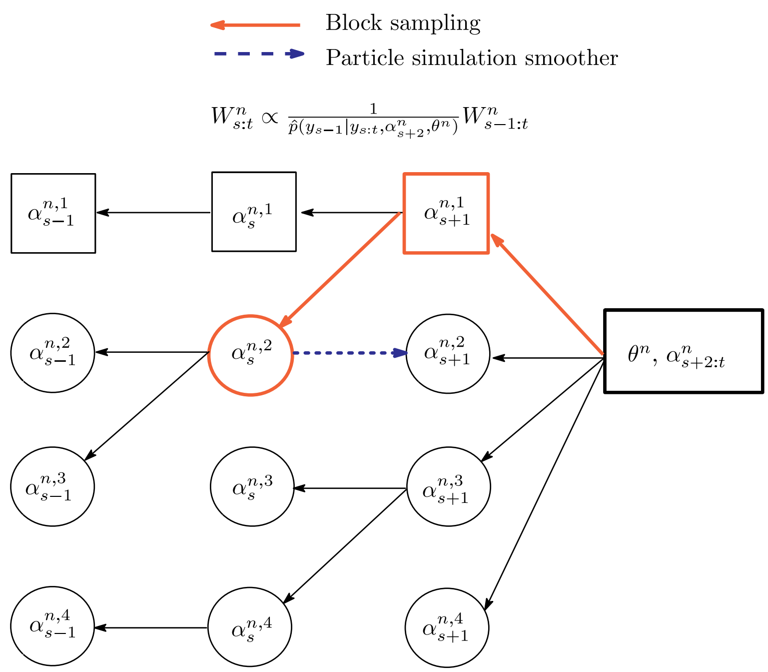

Figure 2: Backward block sampling and particle simulation smoother.

Remark 4. In the above algorithm, we assume we can evaluate $p(\alpha_{j-1} \mid \alpha_j, \theta)$ given in (9).

Remark 5. To remove the oldest observation in Step 2, we reweight the particles according to the likelihood $g_\theta(y_{s-1}|\alpha_{s-1}, \alpha_s)$ for the simple particle rolling MCMC. On the other hand, for the particle rolling MCMC with double block sampling, we reweight the particles according to the unbiased estimate of the conditional likelihood $\hat{p}(y_{s-1} \mid y_{s:t}, \alpha^n_{s+K}, \theta^n)$ where we condition on $y_{s:t}$ but $(\alpha_{s-1}, \alpha_s, \ldots, \alpha_{s+K-1})$ are integrated out, which results in substantial improvement in the weight degeneracy problem.

3.3 Initializing the rolling estimation

In the above discussion, it is implicitly assumed that the particles approximating $\pi(\theta, \alpha_{1:L+1} \mid y_{1:L+1})$ are obtained. To sample from this initial posterior distribution, using MCMC-based methods is straightforward as in the warm-up period for the practical filtering described in Polson et al. (2008). Moreover, we could simply use MCMC samples generated from the MCMC algorithm targeting the initial posterior distribution. However, based on our proposed method for the rolling estimation, we can obtain samples of $\alpha_{1:L+1}$ and θ simply by skipping the discarding steps. The advantage of using our SMC-based method is that we can obtain the estimate of marginal likelihood $p(y_{1:L+1})$ as a by-product (the initializing algorithm and the marginal likelihood estimator are described in detail in the Supplementary Material). This initializing algorithm can be used for the ordinary sequential learning of $\pi(\theta, \alpha_{1:t} \mid y_{1:t})$ $(t = 1, \ldots, T)$. We note that this approach is derived from the particle Gibbs scheme in Andrieu et al. (2010). Our approach is different from that of SMC^2 which applies the particle MH scheme as noted in Chopin et al. (2013) and Fulop and Li (2013).

4 Theoretical justification

Theoretical justifications of our proposed algorithm 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 $(\alpha^n_{s-1:t}, \theta^n)$ 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 1 (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

$$p(\alpha_{t-K:t-1}, a_{t-K}^{k_{t-K+1}}, \ldots, a_{t-1}^{k_t} \mid \alpha_{t-K-1}, y_{t-K:t-1}, \theta) = \frac{\pi(\alpha_{t-K:t-1} \mid \alpha_{t-K-1}, y_{t-K:t-1}, \theta)}{M^K}. \tag{32}$$

Let $a_j^{1:M} = (a_j^1, \ldots, a_j^M)$ and $a_j^{-k_{j+1}} \equiv a_j^{1:M} \setminus a_j^{k_{j+1}} = a_j^{1:M} \setminus k_j$ for $j = t-K, \ldots, t-1$ where we note $a_j^{k_{j+1}} = k_j$ and $k_t = 1$ in (19). 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 (2)(3) and (4) of Step 1, 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

$$\begin{aligned}
&\psi_\theta\left(\alpha_{t-K}^{-k_{t-K}}, \ldots, \alpha_{t-1}^{-k_{t-1}}, \alpha_t^{1:M}, a_{t-K}^{-k_{t-K+1}}, \ldots, a_{t-1}^{-k_t}, k_t^* \mid \alpha_{t-K-1:t-1}, a_{t-K}^{k_{t-K+1}}, \ldots, a_{t-1}^{k_t}, y_{t-K:t}\right) \\
&\quad= \prod_{\substack{m=1\\ m\neq k_{t-K}}}^{M} q_{t-K,\theta}(\alpha_{t-K}^m \mid \alpha_{t-K-1}, y_{t-K}) \times \prod_{j=t-K+1}^{t-1} \prod_{\substack{m=1\\ m\neq k_j}}^{M} V_{j-1,\theta}^{a_{j-1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_{j-1}^m}, y_j) \\
&\qquad \times q_{t,\theta}(\alpha_t^1 \mid \alpha_{t-1}^{k_{t-1}}, y_t) \times \prod_{m=2}^{M} V_{t-1,\theta}^{a_{t-1}^m} q_{t,\theta}(\alpha_t^m \mid \alpha_{t-1}^{a_{t-1}^m}, y_t) \times V_{t,\theta}^{k_t^*}.
\end{aligned} \tag{33}$$

In Step 1 (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 (before the particle smoother step) is written as

$$\begin{aligned}
&\hat{\pi}(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K:t}^{1:M}, a_{t-K:t-1}^{1:M}, k_t^* \mid y_{s-1:t}) \\
&\equiv \frac{\pi(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{k_{t-K}}, \ldots, \alpha_{t-1}^{k_{t-1}} \mid y_{s-1:t-1})}{M^K} \\
&\quad \times \psi_\theta(\alpha_{t-K}^{-k_{t-K}}, \ldots, \alpha_{t-1}^{-k_{t-1}}, \alpha_t^{1:M}, a_{t-K}^{-k_{t-K+1}}, \ldots, a_{t-1}^{-k_t}, k_t^* \mid \alpha_{t-K-1:t-1}, a_{t-K}^{k_{t-K+1}}, \ldots, a_{t-1}^{k_t}, y_{t-K:t}) \\
&\quad \times \frac{\hat{p}(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta)}{p(y_t \mid y_{s-1:t-1})} \\
&= \frac{\pi(\theta, \alpha_{s-1:t-1} \mid y_{s-1:t-1})}{M^K} \\
&\quad \times \prod_{\substack{m=1\\ m\neq k_{t-K}}}^{M} q_{t-K,\theta}(\alpha_{t-K}^m \mid \alpha_{t-K-1}, y_{t-K}) \times \prod_{j=t-K+1}^{t-1} \prod_{\substack{m=1\\ m\neq k_j}}^{M} V_{j-1,\theta}^{a_{j-1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_{j-1}^m}, y_j) \\
&\quad \times \quad q_{t,\theta}(\alpha_t^1 \mid \alpha_{t-1}^{k_{t-1}}, y_t) \times \prod_{m=2}^{M} V_{t-1,\theta}^{a_{t-1}^m} q_{t,\theta}(\alpha_t^m \mid \alpha_{t-1}^{a_{t-1}^m}, y_t) \times V_{t,\theta}^{k_t^*} \\
&\quad \times \frac{\hat{p}(y_t \mid y_{s-1:t-1}, \alpha_{t-K-1}, \theta)}{p(y_t \mid y_{s-1:t-1})}.
\end{aligned} \tag{34}$$

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

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

$$\begin{aligned}
&\hat{\pi}(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K:t}^{1:M}, a_{t-K:t-1}^{1:M}, k_t^* \mid y_{s-1:t}) \\
&\quad = \frac{\pi(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{k_{t-K}^*}, \ldots, \alpha_t^{k_t^*} \mid y_{s-1:t})}{M^{K+1}} \times \prod_{\substack{m=1 \\ m \neq k_{t-K}^*}}^{M} q_{t-K,\theta}(\alpha_{t-K}^m \mid \alpha_{t-K-1}, y_{t-K}) \\
&\qquad \times \prod_{j=t-K+1}^{t} \prod_{\substack{m=1 \\ m \neq k_j^*}}^{M} V_{j-1,\theta}^{a_{j-1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_{j-1}^m}, y_j),
\end{aligned} \tag{35}$$

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 the Supplementary Material.

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 6. We note that k_j's do not appear in (35). 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 this weight can be considered an unbiased estimator.

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 (33), 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 the Supplementary Material.

This shows that 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 obtain the decomposition of the variance as follows.

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

The variance of the incremental weight consists of two components, including variance of the conditional likelihood and (expected) variance which is introduced using M particles to approximate the conditional likelihood. This decomposition identifies factors that influences the ESS of the particles. Regarding 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

$$E\left[p(y_t \mid y_{s-1:t-1}, \alpha_{t-K_1-1}, \theta) \mid \alpha_{s-1:t-K_2-1}, \theta\right] = p(y_t \mid y_{s-1:t-1}, \alpha_{t-K_2-1}, \theta). \tag{36}$$

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 affects the variance of incremental weights in practice and show that large K actually reduces the variance in each step of sampling.

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 2 (1) of Section 3.2, 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

$$p(\alpha_{s-1:s+K-1}, a_{s-1}^{k_{s-2}}, \ldots, a_{s+K-1}^{k_{s+K-2}} \mid \alpha_{s+K}, \theta, y_{s-1:t}) \quad = \quad \frac{\pi(\alpha_{s-1:s+K-1} \mid \alpha_{s+K}, y_{s-1:t}, \theta)}{M^{K+1}}. \quad (37)$$

In (2)(3) and (4) of Steps 2, 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

$$\begin{aligned}
&\bar{\psi}_\theta(\alpha_{s-1}^{-k_{s-1}}, \ldots, \alpha_{s+K-1}^{-k_{s+K-1}}, a_s^{-k_{s-1}}, \ldots, a_{s+K-1}^{-k_{s+K-2}}, k_s^* \mid \alpha_{s-1:s+K}, a_{s-1}^{k_{s-2}}, \ldots, a_{s+K-1}^{k_{s+K-2}}, y_{s-1:s+K-1}) \\
&= \prod_{\substack{m=1 \\ m \neq k_{s+K-1}}}^{M} q_{s+K-1,\theta}(\alpha_{s+K-1}^m \mid \alpha_{s+K}, y_{s+K-1}) \times \prod_{j=s-1}^{s+K-2} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j+1}^{a_{j+1}^m}, y_j) \times V_{s,\theta}^{k_s^*}.
\end{aligned} \quad (38)$$

In Step2 (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

$$\begin{aligned}
&\breve{\pi}(\theta, \alpha_{s-1:s+K-1}^{1:M}, \alpha_{s+K:t}, a_{s:s+K-1}^{1:M}, k_{s-1}, k_s^* \mid y_{s-1:t}) \\
&\equiv \frac{\pi(\theta, \alpha_{s-1:t} \mid y_{s-1:t})}{M^{K+1}} \\
&\times \prod_{\substack{m=1 \\ m \neq k_{s+K-1}}}^{M} q_{s+K-1,\theta}(\alpha_{s+K-1}^m \mid \alpha_{s+K}, y_{s+K-1}) \times \prod_{j=s-1}^{s+K-2} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j+1}^{a_{j+1}^m}, y_j) \\
&\times \quad V_{s,\theta}^{k_s^*} \times \frac{p(y_{s-1} \mid y_{s:t})}{\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)},
\end{aligned} \quad (39)$$

where $p(y_{s-1} \mid y_{s:t})^{-1}$ is the normalizing constant of this target density as shown in Proposition 4.4. Below we state Proposition 4.3 for the backward block sampling, which correspond to Proposition 4.1 for the forward block sampling.

Proposition 4.3. *The artificial target density $\breve{\pi}$ for the backward block sampling can be rewritten as*

$$
\begin{aligned}
&\breve{\pi}(\theta, \alpha_{s-1:s+K-1}^{1:M}, \alpha_{s+K:t}, a_{s:s+K-1}^{1:M}, k_{s-1}, k_s^* \mid y_{s-1:t}) \\
&\quad = \frac{\pi(\theta, \alpha_s^{k_s^*}, \dots, \alpha_{s+K-1}^{k_{s+K-1}^*}, \alpha_{s+K:t} \mid y_{s:t})}{M^K} \times \prod_{\substack{m=1 \\ m \neq k_{s+K-1}^*}}^{M} q_{s+K-1,\theta}(\alpha_{s+K-1}^m \mid \alpha_{s+K}, y_{s+K-1:t}) \\
&\quad \times \prod_{j=s}^{s+K-2} \prod_{\substack{m=1 \\ m \neq k_j^*}}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j+1}^{a_{j+1}^m}, y_j) \times \prod_{m=1}^{M} V_{s,\theta}^{a_s^m} q_{s-1,\theta}(\alpha_{s-1}^m \mid \alpha_s^{a_s^m}, y_{s-1}) \times V_{s-1,\theta}^{k_{s-1}}, \qquad (40)
\end{aligned}
$$

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

Proof. See the Supplementary Material.

Although the probability density (40) in Proposition 4.3 has a bit different form from that of (35) 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 4.4.

Proposition 4.4. *If*

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

and

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

where $\bar{\psi}_\theta$ is given in (38), then

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

Proof. See the Supplementary Material.

4.3 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 improve the mixing property of Particle Gibbs, which is also 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 proposition.

Proposition 4.5. *The joint conditional density of* $(k_{t-K}^*,\ldots,k_t^*)$ *is given by*

$$\begin{aligned}
&\hat{\pi}(k_{t-K}^*,\ldots,k_t^*|\theta,\alpha_{s-1:t-K-1},\alpha_{t-K:t}^{1:M},a_{t-K:t-1}^{1:M},y_{s-1:t})\\
&= \hat{\pi}(k_t^*|\theta,\alpha_{s-1:t-K-1},\alpha_{t-K:t}^{1:M},a_{t-K:t-1}^{1:M},y_{s-1:t})\\
&\times \prod_{t_0=t-1}^{t-K} \hat{\pi}(k_{t_0}^*|\theta,\alpha_{s-1:t-K-1},\alpha_{t-K:t_0}^{1:M},a_{t-K:t_0-1}^{1:M},\alpha_{t_0+1}^{k_{t_0+1}^*},\ldots,\alpha_t^{k_t^*},k_{t_0+1:t}^*,y_{s-1:t}),
\end{aligned} \tag{41}$$

where

$$\begin{aligned}
&\hat{\pi}(k_{t_0}^*|\theta,\alpha_{s-1:t-K-1},\alpha_{t-K:t_0}^{1:M},a_{t-K:t_0-1}^{1:M},\alpha_{t_0+1}^{k_{t_0+1}^*},\ldots,\alpha_t^{k_t^*},k_{t_0+1:t}^*,y_{s-1:t})\\
&\quad = \bar{V}_{t_0,\theta}^{k_{t_0}^*}, \quad \bar{V}_{j,\theta}^m \equiv \frac{V_{j,\theta}^m f_\theta(\alpha_{j+1}^{k_{j+1}^*}\mid\alpha_j^m,y_{j+1})}{\sum_{i=1}^M V_{j,\theta}^i f_\theta(\alpha_{j+1}^{k_{j+1}^*}\mid\alpha_j^i,y_{j+1})}.
\end{aligned} \tag{42}$$

Proof. See the Supplementary Material.

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 (34). In Step 1 (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.

5 Real data example

We illustrate our proposed method using the RSV model in Section 2.3 and set $M=300$ and $N=1000$ (we also tried using other values of M but the computation time is the shortest with $M=300$). Further we always implement 10 MCMC iterations for the comparison below unless otherwise stated. As a proposal density, we simply use a prior density $q_{t,\theta}(\alpha_t\mid\alpha_{t-1},y_{s-1:t})=f_\theta(\alpha_t\mid\alpha_{t-1})$ to demonstrate that the block sampling improves even with the simplest proposal. The summary statistics of R_{1t} and R_{2t} are shown in Table 4 where we use $K=5,10$ and 15. In contrast to the simple particle rolling MCMC algorithm, although we use a simple prior density as a proposal density, both means are close to 1 demonstrating that our sampling algorithm succeeded in overcoming the weight degeneracy problem. 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 4: Summary statistics for R_{1t} and R_{2t} ($t = 1988, \ldots, 4248$).

Figure 3 shows the trace plot 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 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 weaker after the financial crisis.

Figure 4 shows three cumulative computation times (wall time) for the same period corresponding to $K = 5, 10$ and 15. The computation times with $K = 5$ and $K = 15$ are longer than that with $K = 10$. This finding implies that, when $K = 5$, the effect of the blocking is not sufficient to reduce the path dependence between α_t and α_{t-K-1} (similarly, α_{s-1} and α_{s+K}). When $K = 15$, the Monte Carlo error in the local conditional SMC increased the variance of the importance weights (recall that the variance of the weights is derived from both the effect of blocking and the Monte Carlo error in the conditional SMC in Section 4).

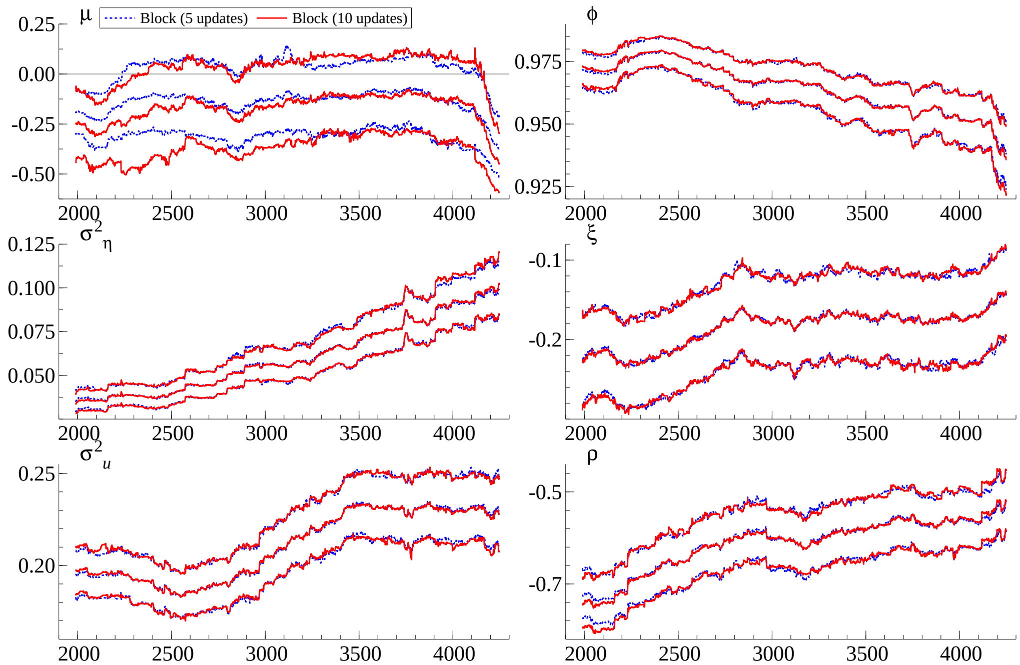

Figure 3: Trace plot of estimated posterior means and 95% credible intervals for parameters using S&P500 return in RSV model (from December 31, 2007 to December 30, 2016).

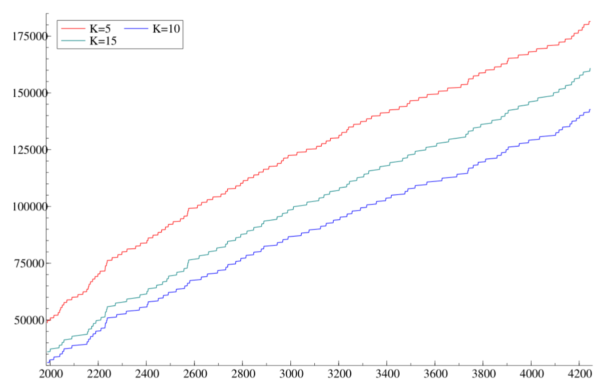

Figure 4: Cumulative computation times (wall time, unit time = second) (t = $1988, \ldots, 4248$).

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. Then we apply our proposed sampling algorithm with block size of $K = 10$, $M = 300$ 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 (2) 5 iterations and (3) 10 iterations. Figure 5 shows the estimation results for the first period (the figure for the last estimation period is similar and hence omitted).

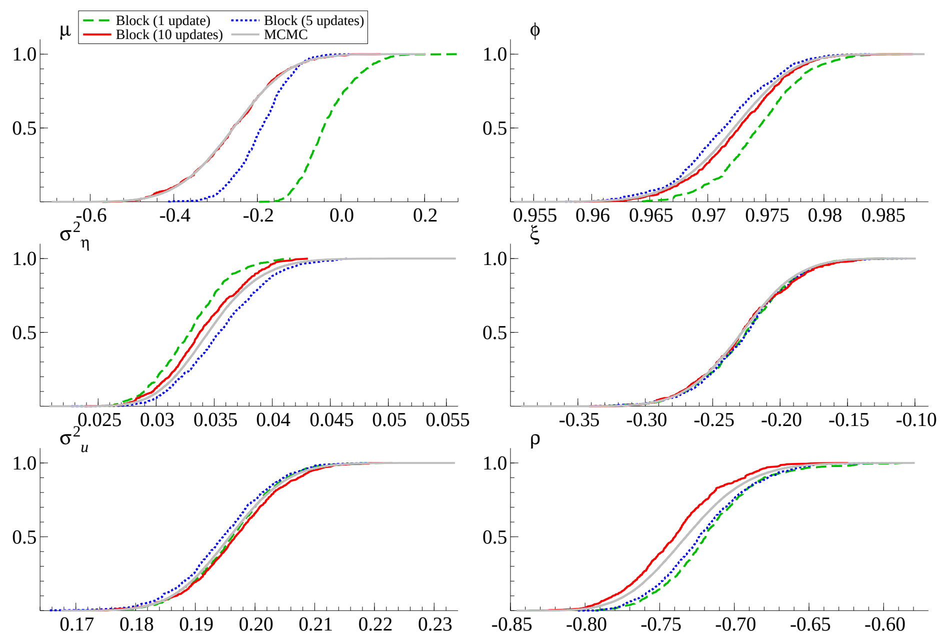

Figure 5: The estimated posterior distribution functions of θ for $t = 1, \ldots, 1988$. MCMC and Particle rolling MCMC: 1, 5 and 10 iterations.

Among three cases, the estimates obtained by iterating MCMC algorithm 5 or 10 times in the MCMC update steps are close to those obtained by the ordinary MCMC sampling algorithm. If only one iteration is performed in the MCMC update step, the estimation results are found to be inaccurate because the MCMC iterations not only diversify the particles but also correct approximation errors introduced by the particle algorithm, which basically update only a part of the vector $\alpha^n_{s-1:t}$. 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 the MCMC update steps such that the particles can trace the correct posterior distributions.

6 Conclusion

In this paper, we propose a novel efficient estimation method to implement the rolling window particle MCMC simulation using a Sequential Monte Carlo framework and refreshing steps with MCMC kernel. The weighted particles are updated to learn and discard the information of the new and old observations using the forward and backward block sampling based on the conditional SMC algorithm, which effectively circumvent the weight degeneracy problem. 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 realized stochastic volatility model with S&P500 index returns.

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, 19H00588.

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.

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.

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

Gilks, W. R. and C. Berzuini (2001). Following a moving target- Monte Carlo inference for dynamic Bayesian models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 63*(1), 127–146.

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.

Naesseth, C. A., F. Lindsten, and T. B. Schön (2015). Nested sequential monte carlo methods. *arXiv preprint arXiv:1502.02536*.

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.

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 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.

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.

Particle rolling MCMC: Supplementary Material

Naoki Awaya* and Yasuhiro Omori†

September 27, 2019

A Proofs

A.1 Proof of Proposition 4.1

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

Lemma A.1. *For any t and t_0 $(t-K \le t_0 \le t)$,*

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

$$= \pi(\theta, \alpha_{s-1:t-K-1} \mid y_{s-1:t-K-1}) \times \prod_{m=1}^{M} q_{t-K,\theta}(\alpha_{t-K}^m \mid \alpha_{t-K-1}, y_{t-K}) \times \prod_{j=t-K+1}^{t_0} \prod_{m=1}^{M} V_{j-1,\theta}^{a_{j-1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_{j-1}^m}, y_j) \times V_{t_0,\theta}^{k_{t_0}} \times \prod_{j=t-K}^{t_0} \frac{\hat{p}(y_j \mid y_{s-1:j-1}, \alpha_{t-K-1}, \theta)}{p(y_j \mid y_{s-1:j-1})}, \tag{45}$$

where

$$\hat{p}(y_j \mid y_{s-1:j-1}, \alpha_{t-K-1}, \theta) = \frac{1}{M} \sum_{m=1}^{M} v_{j,\theta}(\alpha_{j-1}^{a_{j-1}^m}, \alpha_j^m), \quad j = t-K, \ldots, t_0, \tag{46}$$

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

*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.

The probability density (44) 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.

Proof of Lemma A.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 (44) is

$$\begin{aligned}
&\pi(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{k_{t-K}}, \ldots, \alpha_{t_0}^{k_{t_0}} \mid y_{s-1:t_0}) \\
&= \frac{\pi(\theta, \alpha_{s-1:t-K-1} \mid y_{s-1:t-K-1})}{p(y_{t-K:t_0} \mid y_{s-1:t-K-1})} \prod_{j=t-K}^{t_0} f_\theta(\alpha_j^{k_j} \mid \alpha_{j-1}^{k_{j-1}}, y_{j-1}) g_\theta(y_j \mid \alpha_j^{k_j}) \\
&= \frac{\pi(\theta, \alpha_{s-1:t-K-1} \mid y_{s-1:t-K-1})}{p(y_{t-K:t_0} \mid y_{s-1:t-K-1})} \prod_{j=t-K}^{t_0} v_{j,\theta}(\alpha_{j-1}^{k_{j-1}}, \alpha_j^{k_j}) \prod_{j=t-K}^{t_0} q_{j,\theta}(\alpha_j^{k_j} \mid \alpha_{j-1}^{k_{j-1}}, y_j). \qquad (47)
\end{aligned}$$

Thus we obtain

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

and the result follows where we substitute (47) 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. □

Using Lemma A.1, we obtain Proposition 4.1 as follows.

Proof of Proposition 4.1.
By applying Lemma A.1 with $t_0 = t-1$ to the first three terms of (35), 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 A.1 with $t_0 = t$ and $k_{t_0} = k_t^*$ to the last equation and the result follows. □

A.2 Proof of Proposition 4.2

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 (35),

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

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.3 Proof of Proposition 4.3

We first establish the following lemma as in the proof of Proposition 4.1.

Lemma A.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

$$
\hat{p}(y_j \mid y_{j+1:t}, \alpha_{s+K}, \theta) = \frac{1}{M} \sum_{m=1}^{M} v_{j,\theta}(\alpha_j^m, \alpha_{j+1}^{a_{j+1}^m}). \tag{48}
$$

Proof of Lemma A.2.

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

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

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

□

Proof of Proposition 4.3.

By applying Lemma A.2 with $s_0 = s - 1$ to the first three terms of the target distribution in

(40), we have

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

where we again applied Lemma A.2 with $s_0 = s$ and $k_{s-1} = k_{s-1}^*$ in the last equality. □

A.4 Proof of Proposition 4.4

Proof of Proposition 4.4.

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

$$
\pi(\alpha_{s-1:s+K-1} \mid \alpha_{s+K}, y_{s-1:t}, \theta) = \frac{\pi(\theta, \alpha_{s-1:s+K-1}, \alpha_{s+K:t} \mid y_{s-1:t})}{\pi(\theta, \alpha_{s+K:t} \mid y_{s-1:t})}. \tag{49}
$$

Since

$$\begin{aligned}
&\frac{1}{\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)} \bar{\psi}_{\theta,0}(\alpha^{1:M}_{s-1:s+K-1}, a^{1:M}_{s:s+K-1}, k_{s-1}, k^*_s \mid \alpha_{s+K}, y_{s-1:s+K-1}) \\
&\quad = \breve{\pi}(\theta, \alpha^{1:M}_{s-1:s+K-1}, \alpha_{s+K:t}, a^{1:M}_{s:s+K-1}, k_{s-1}, k^*_s \mid y_{s-1:t}) \frac{1}{p(y_{s-1} \mid y_{s:t}) \pi(\theta, \alpha_{s+K:t} \mid y_{s-1:t})},
\end{aligned}$$

where we used the definition of $\breve{\pi}$ in (40), we obtain

$$\begin{aligned}
&E_{\bar{\psi}_{\theta,0}}\left[\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)^{-1} \mid \alpha_{s+K}, y_{s-1:t}, \theta\right] \\
&\quad = \int \frac{1}{\hat{p}(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)} \bar{\psi}_{\theta,0}(\alpha^{1:M}_{s-1:s+K-1}, a^{1:M}_{s:s+K-1}, k_{s-1}, k^*_s \mid \alpha_{s+K}, y_{s-1:s+K-1}) \\
&\qquad\qquad d\alpha^{1:M}_{s-1:s+K-1} da^{1:M}_{s:s+K-1} dk_{s-1} dk^*_s \\
&\quad = \int \breve{\pi}(\theta, \alpha^{1:M}_{s-1:s+K-1}, \alpha_{s+K:t}, a^{1:M}_{s:s+K-1}, k_{s-1}, k^*_s \mid y_{s-1:t}) d\alpha^{1:M}_{s-1:s+K-1} da^{1:M}_{s:s+K-1} dk_{s-1} dk^*_s \\
&\qquad \times \frac{1}{p(y_{s-1} \mid y_{s:t}) \pi(\theta, \alpha_{s+K:t} \mid y_{s-1:t})} \\
&\quad = \frac{\pi(\theta, \alpha_{s+K:t} \mid y_{s:t})}{p(y_{s-1} \mid y_{s:t}) \pi(\theta, \alpha_{s+K:t} \mid y_{s-1:t})} = \frac{1}{p(y_{s-1} \mid y_{s:t}, \alpha_{s+K:t}, \theta)} = \frac{1}{p(y_{s-1} \mid y_{s:t}, \alpha_{s+K}, \theta)}
\end{aligned}$$

where we use Proposition 4.3 in the third equality. Further,

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

□

Proof of Proposition 4.5

Proof of Proposition 4.5.

Consider the joint marginal density of (36):

$$\begin{aligned}
&\hat{\pi}(\theta, \alpha_{s-1:t-K-1}, \alpha^{1:M}_{t-K:t_0}, a^{1:M}_{t-K:t_0-1}, \alpha^{k^*_{t_0+1}}_{t_0+1}, \ldots, \alpha^{k^*_t}_t, k^*_{t_0:t} \mid y_{s-1:t}) \\
&\quad = \frac{\pi(\theta, \alpha_{s-1:t-K-1}, \alpha^{k^*_{t-K}}_{t-K}, \ldots, \alpha^{k^*_t}_t \mid y_{s-1:t})}{M^{K+1}} \times \prod_{\substack{m=1 \\ m \neq k^*_{t-K}}}^{M} q_{t-K,\theta}(\alpha^m_{t-K} \mid \alpha_{t-K-1}, y_{t-K}) \\
&\qquad \times \prod_{j=t-K+1}^{t_0} \prod_{\substack{m=1 \\ m \neq k^*_j}}^{M} V^{a^m_j}_{j-1} q_{j,\theta}(\alpha^m_j \mid \alpha^{a^m_j}_{j-1}, y_j), \qquad (50)
\end{aligned}$$

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

$$
\begin{aligned}
&\hat{\pi}(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{1:M}, \alpha_{t-K+1}^{k_{t-K+1}^*}, \dots, \alpha_t^{k_t^*}, k_{t-K:t}^* \mid y_{s-1:t}) \\
&\quad = \frac{\pi(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{k_{t-K}^*}, \dots, \alpha_t^{k_t^*} \mid y_{s-1:t})}{M^{K+1}} \times \prod_{\substack{m=1 \\ m \neq k_{t-K}^*}}^{M} q_{t-K,\theta}(\alpha_{t-K}^m \mid \alpha_{t-K-1}, y_{t-K}). \qquad (51)
\end{aligned}
$$

Then we obtain

$$
\begin{aligned}
&\hat{\pi}(k_{t_0}^* \mid \theta, \alpha_{s-1:t-K-1}^{1:M}, \alpha_{t-K:t_0}^{1:M}, a_{t-K:t_0-1}^{1:M}, \alpha_{t_0+1}^{k_{t_0+1}^*}, \dots, \alpha_t^{k_t^*}, k_{t_0+1:t}^*, y_{s-1:t}) \\
&\quad \propto \hat{\pi}(\theta, \alpha_{s-1:t-K-1}^{1:M}, \alpha_{t-K:t_0}^{1:M}, a_{t-K:t_0-1}^{1:M}, \alpha_{t_0+1}^{k_{t_0+1}^*}, \dots, \alpha_t^{k_t^*}, k_{t_0:t}^* \mid y_{s-1:t}) \\
&\quad \propto \frac{\pi(\theta, \alpha_{s-1:t-K-1}, \alpha_{t-K}^{k_{t-K}^*}, \dots, \alpha_{t_0}^{k_{t_0}^*} \mid y_{s-1:t_0})}{M^{t_0-(t-K)+1}} \times \prod_{\substack{m=1 \\ m \neq k_{t-K}^*}}^{M} q_{t-K,\theta}(\alpha_{t-K}^m \mid \alpha_{t-K-1}, y_{t-K}) \\
&\quad \times \prod_{j=t-K+1}^{t_0} \prod_{\substack{m=1 \\ m \neq k_j^*}}^{M} V_{j-1}^{a_j^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_j^m}, y_j) \times \prod_{j=t_0+1}^{t} f_\theta(\alpha_j^{k_j^*} \mid \alpha_{j-1}^{k_{j-1}^*}, y_{j-1}) g_\theta(y_j \mid \alpha_j^{k_j^*}) \\
&\quad = \pi(\theta, \alpha_{s-1:t-K-1} \mid y_{s-1:t-K-1}) \times \prod_{m=1}^{M} q_{t-K,\theta}(\alpha_{t-K}^m \mid \alpha_{t-K-1}, y_{t-K}) \\
&\quad \times \prod_{j=t-K+1}^{t_0} \prod_{m=1}^{M} V_{j-1,\theta}^{a_{j-1}^m} q_{j,\theta}(\alpha_j^m \mid \alpha_{j-1}^{a_{j-1}^m}, y_j) \times V_{t_0,\theta}^{k_{t_0}^*} \times \prod_{j=t-K}^{t_0} \frac{\hat{p}(y_j \mid y_{s-1:j-1}, \alpha_{t-K-1}, \theta)}{p(y_j \mid y_{s-1:j-1})}, \\
&\quad \times \prod_{j=t_0+1}^{t} f_\theta(\alpha_j^{k_j^*} \mid \alpha_{j-1}^{k_{j-1}^*}, y_{j-1}) g_\theta(y_j \mid \alpha_j^{k_j^*}) \\
&\quad \propto V_{t_0,\theta}^{k_{t_0}^*} \times f_\theta(\alpha_{t_0+1}^{k_{t_0+1}^*} \mid \alpha_{t_0}^{k_{t_0}^*}, y_{t_0}), \qquad (52)
\end{aligned}
$$

where we use Lemma A.1 at the equality. □

B Initializing the rolling estimation

In this section, we first give the initializing algorithm which is obtained by skipping the discarding step in the particle rolling algorithm. Next, we describe how to estimate the marginal likelihood.

B.1 Algorithm

(1) At time $j = 1$, sample (θ^n, α_1^n) from $\pi(\theta, \alpha_1 \mid y_1)$ for $n = 1, \dots, N$.

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

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

$$V_{1,\theta^n}^{n,m} = \frac{v_{1,\theta^n}(\alpha_1^{n,m})}{\sum_{i=1}^M v_{1,\theta^n}(\alpha_1^{n,i})}, \quad v_{1,\theta^n}(\alpha_1^{n,m}) = \frac{\mu_{\theta^n}(\alpha_1^{n,m}) g_{\theta^n}(y_1 \mid \alpha_1^{n,m})}{q_{1,\theta^n}(\alpha_1^{n,m} \mid y_1)}. \tag{53}$$

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

$$W_1^n \propto \hat{p}(y_1 \mid \theta^n), \quad \hat{p}(y_1 \mid \theta^n) = \sum_{m=1}^M v_{1,\theta^n}(\alpha_1^{n,m}). \tag{54}$$

(2) At time $j = 2, \ldots, L+1$, implement the forward block sampling to generate $\alpha_{1:j}^n$ and θ^n, and compute its importance weight

$$W_j^n \propto \hat{p}(y_j \mid y_{1:j-1}, \alpha_{j-K-1}^n, \theta^n) \times W_{j-1}^n, \tag{55}$$

$$\hat{p}(y_j \mid y_{1:j-1}, \alpha_{j-K-1}^n, \theta^n) = \frac{1}{M} \sum_{m=1}^M v_{j,\theta^n}(\alpha_{j-1}^{n,a_{j-1}^{n,m}}, \alpha_j^{n,m}). \tag{56}$$

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

Remark 1. 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.

B.2 Estimation of the marginal likelihood

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

$$p(y_{s:t}) = \int p(y_{s:t} \mid \alpha_{s:t}, \theta) p(\alpha_{s:t} \mid \theta) p(\theta) d\alpha_{s:t} d\theta, \tag{57}$$

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

$$p(y_{s:t}) = \frac{p(y_t \mid y_{s-1:t-1})}{p(y_{s-1} \mid y_{s:t})} p(y_{s-1:t-1}), \tag{58}$$

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

$$\hat{p}(y_{s:t}) = \frac{\hat{p}(y_t \mid y_{s-1:t-1})}{\hat{p}(y_{s-1} \mid y_{s:t})} \hat{p}(y_{s-1:t-1}), \tag{59}$$

where

$$\hat{p}(y_t \mid y_{s-1:t-1}) \;=\; \sum_{n=1}^{N} W^n_{s-1:t-1}\hat{p}(y_t \mid y_{s-1:t-1}, \alpha^n_{t-K-1}, \theta^n), \tag{60}$$

$$\hat{p}(y_{s-1} \mid y_{s:t}) \;=\; \sum_{n=1}^{N} W^n_{s-1:t}\hat{p}(y_{s-1} \mid y_{s:t}, \alpha^n_{s+K}, \theta^n), \tag{61}$$

using (23), (24), (29) and (30). The initial estimate $\hat{p}(y_{1:L+1})$, $L = t - s$ is given by

$$\hat{p}(y_{1:L+1}) \;=\; \hat{p}(y_1)\prod_{j=2}^{L+1}\hat{p}(y_j \mid y_{1:j-1}), \tag{62}$$

where we use (54), (55) and (56) to obtain

$$\hat{p}(y_1) \;=\; \sum_{n=1}^{N}\hat{p}(y_1 \mid \theta^n), \quad \hat{p}(y_j \mid y_{1:j-1}) = \sum_{n=1}^{N} W^n_{j-1}\hat{p}(y_j \mid y_{1:j-1}, \alpha^n_{j-K-1}, \theta^n). \tag{63}$$

C Additional illustrative numerical examples

C.1 Linear Gaussian state space model

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. We adopt conjugate priors, $\mu \mid \sigma^2 \sim \mathcal{N}(0, 10\sigma^2)$ and $\sigma^2 \sim \mathcal{IG}(5/2, 0.05/2)$. The rolling estimation is conducted with a 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. Because 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 block sampling in Section 3. We also use (3) the simple particle rolling MCMC sampler as a benchmark. In summary, we consider the following:

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

In the forward block sampling, we 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, we 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 5: The number of resampling steps in block sampling and simple sampling.

		(1) Fully	(2)	M		(3) Simple
Estimation period	K	adapted	100	300	500	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 5 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), respectively, in the rolling estimation (the initial estimation) period. Additionally, 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. The ratio R_{1t} (R_{2t}) measures the relative magnitude of ESS (effective sample size) 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} values for the block sampling are larger and less dispersed compared with the simple sampler suggesting that the forward block sampling is more efficient. Additionally, the R_{2t} values 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 more efficient. The scatter plots of R_{1t} and R_{2t} are shown at the bottom of Figure 6 for two sampling methods. These results demonstrate that our proposed block sampling is more efficient at both Steps 1 and 2 of each rolling step.

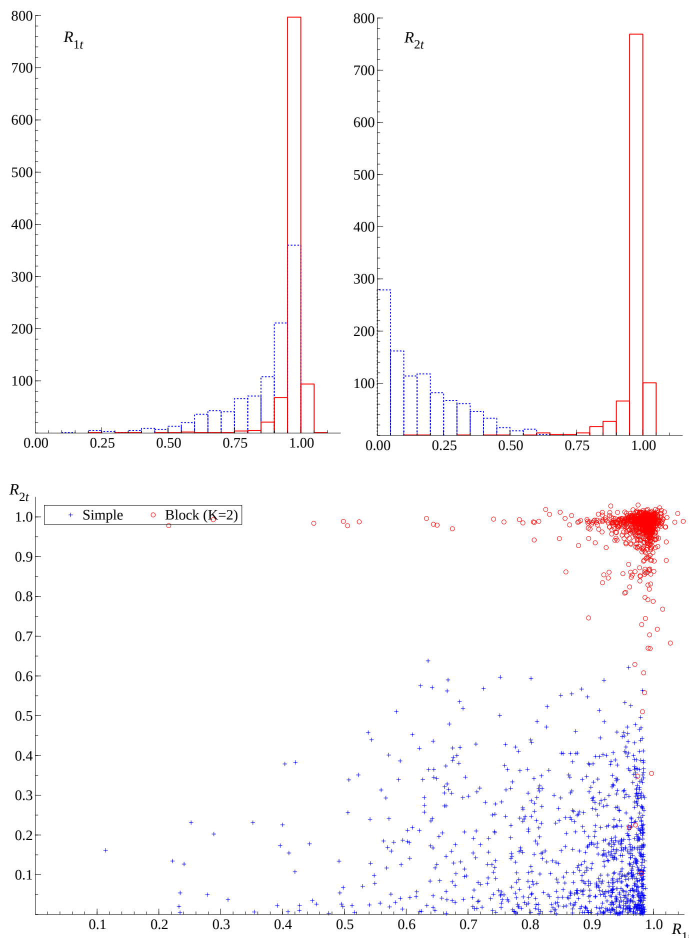

Figure 6: The histograms of R_{1t} (top left) and R_{2t} (top right) ($t = 1001, \ldots, 2000$) for the simple sampler (dotted blue) and the sampler with block sampling with $K = 2$ and $M = 100$ (solid red). The scatter plot of R_{1t} and R_{2t} (bottom).

Table 6 shows the summary statistics for the relative magnitudes of ESS in each step, R_{1t} and R_{2t}. In Step 1, the average R_t value for the block sampling slightly increased compared with that for the simple sampling, but the standard deviation for the former is less than half of that for the latter. Moreover, in Step 2, the average of R_t value for the block sampling is six times larger than that for the simple sampling, while the standard deviation for the former is approximately half of that for the latter. Thus the double block sampling drastically alleviate the weight degeneracy compared with the simple sampling method.

Table 6: 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**

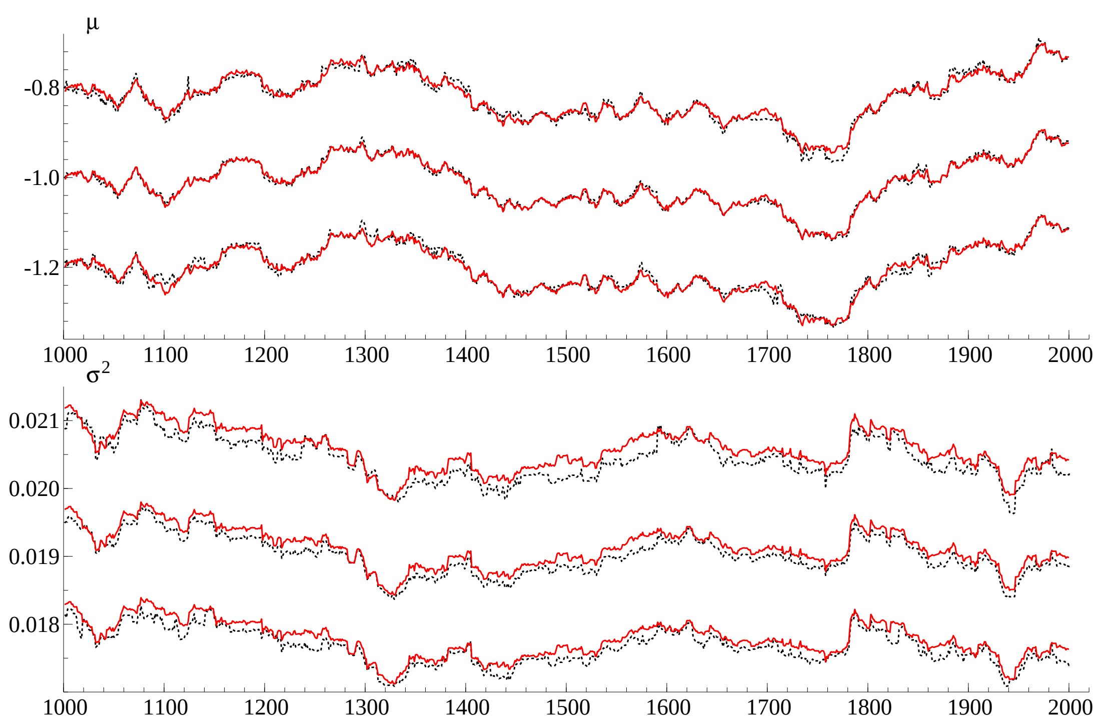

Figure 7: True posterior means and 95% credible intervals (dotted black) with their estimates (solid red) for μ and σ^2.

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

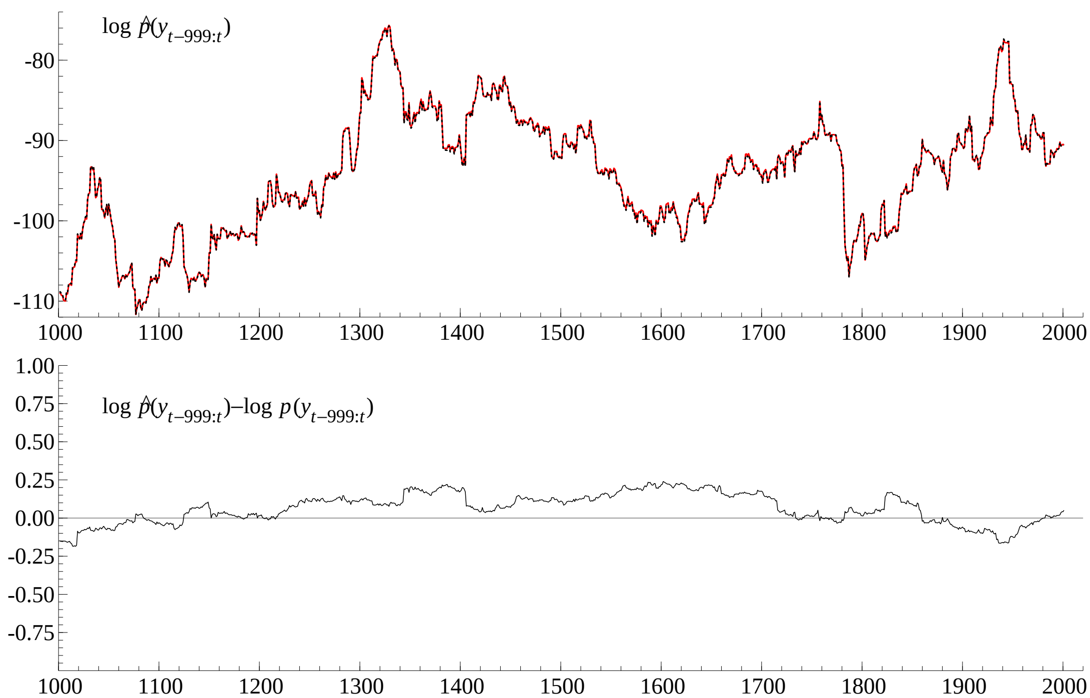

Figure 8: Top: true log marginal likelihoods $\log p(y_{t-999:t})$ (dotted black) and their estimates (solid red). Bottom: estimation errors $\log \hat{p}(y_{t-999:t}) - \log p(y_{t-999:t})$ for $t = 1001, \ldots, 2000$.

C.2 Realized Stochastic Volatility model

In this section, we first compare the computation time of the particle rolling MCMC with the iteration of MCMC. We then compare the marginal likelihoods $\hat{p}(y_{s:t})$ provided in Section B.2.

MCMC sampling is implemented for the initial data window (using $y_{1:1988}$) with 10,000 iteration (2,000 MCMC samples in the burn-in period are discarded). We estimate the total computation time for the MCMC and the PMCMC to complete the rolling window estimation by multiplying the computation time (for the first sample period) by 2261 for S&P500 index data. Table 7 shows the computation times and ESSs for three methods where the ESS for each parameter is computed as the MCMC sample size (10,000) divided by the inefficiency factor ($1 + 2\sum_{s=1}^{\infty} \rho_s$, where ρ_s is the MCMC sample correlation at lag s). The ESS for our proposed method is computed as the average ESS during the rolling estimations. The recursive estimation using the standard MCMC or the PMCMC takes 20-50 times longer than our proposed method. If we consider the ESS, the difference increases (400-900 times longer for the recursive estimations). These results show that the computation time for our proposed method is much smaller compared with recursive estimations using the standard MCMC or PMCMC.

	Time (seconds)	Param.	ESS
PRMCMC	**142,709**	-	**729**
MCMC	$1,293 \times 2,261$	μ	**35**
	$= \mathbf{2,923,473}$	ϕ	1764
		σ_η^2	189
		ξ	3942
		σ_u^2	636
		ρ	257
PMCMC	$3,189 \times 2,261$	μ	**40**
	$= \mathbf{7,210,329}$	ϕ	2184
		σ_η^2	221
		ξ	4878
		σ_u^2	656
		ρ	189

Table 7: Computation times for the PRMCMC, MCMC and PMCMC.

The log marginal likelihoods, $\log p(y_{t-1987:t})$, of the RSV model with and without leverage effects are shown in Figure 9 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 effects is always larger than the other model; thus, the RSV model with leverage effects is supported. This finding is consistent with the rolling estimation results, 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 $t = 2400$.

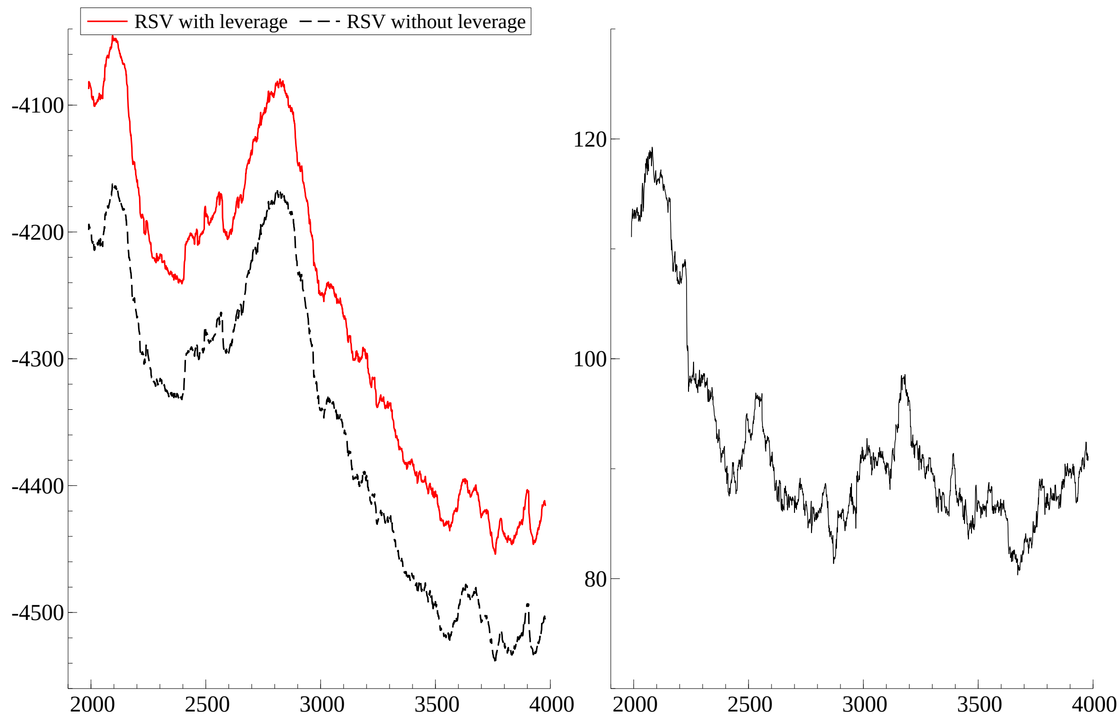

Figure 9: Left: Estimates of $\log p(y_{t-1987:t})$ ($t = 1988, \ldots, 4248$) for S&P 500 index return in RSV model with leverage (solid red) and in RSV model without leverage (dotted black). Right: Difference between two log marginal likelihoods.

References

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.