Particle Rolling MCMC with Double-Block Sampling

Naoki Awaya

Graduate School of Economics,

The University of Tokyo

Yasuhiro Omori

The University of Tokyo

Particle rolling MCMC with double-block sampling

Naoki Awaya[*] and   Yasuhiro Omori[†]

September 16, 2021

Abstract

An efficient particle Markov chain Monte Carlo methodology is proposed for the rolling-window estimation of state space models. The particles are updated to approximate the long sequence of posterior distributions as we move the estimation window. To overcome the well-known weight degeneracy problem that causes the poor approximation, we introduce a practical double-block sampler with the conditional sequential Monte Carlo update where we choose one lineage from multiple candidates for the set of current state variables. Our proposed sampler is justified in the augmented space through theoretical discussions. In the illustrative examples, it is shown to be successful to accurately estimate the posterior distributions of the model parameters.

Keywords: Double-block sampler; 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, this time point is usually unknown, and the change may occur gradually from one state to another. Although there are various statistical models for the structural change in the literature, the rolling-window estimation is the simple and common way to reflect the recent change in the forecasting without delay 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 a long sequence of posterior distributions.

To overcome this difficulty, we take an alternative approach based on the sequential Monte Carlo (SMC) sampler discussed in Del Moral et al. (2006). 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 reiterating the same MCMC algorithm with the slightly different dataset. The particles consist of realized values of state variables and static parameters, which are updated when including a new observation and excluding the old observation. As we shall show in the illustrative examples of Section 4, a simple rolling-window sampler that is derived in a straightforward manner from the previous literature leads to the severe weight degeneracy problem, suggesting that the updating step should be constructed carefully. To fix this problem, 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 simultaneously when learning new information. It is highly efficient in the sense that it substantially increases the effective sample size. Based on this idea, we propose the novel sampling method, called the double-block sampler, where we sample a block of state variables when both including and excluding the 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 conditional SMC of the particle MCMC (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 the particle Gibbs instead of the particle MH (Metropolis-Hastings) algorithm. As a special case of our new method, our proposed double-block sampler can be used to implement the ordinary sequential analysis by keeping all past observations. It contrasts with SMC^2 in that it originates from different types of the particle MCMC algorithms.

The remainder of the paper is organized as follows. In Section 2, we introduce the simple rolling-window sampler for state space models and point out that such a sampler derived from the conventional filtering algorithm causes the serious weight degeneracy phenomenon. Section 3 introduces a double-block sampler to overcome this difficulty. Section 4 provides illustrative examples and, in Section 5, theoretical justifications of the proposed method are provided. Section 6 concludes the paper.

2 Particle rolling MCMC in general state space models

2.1 Rolling-window estimation in general 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 x_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 $x_{s:t} \equiv (x_s, x_{s+1}, \ldots, x_t)$ and $y_{s:t} \equiv (y_s, y_{s+1}, \ldots, y_t)$. We assume that the distribution of y_t given $(y_{1:t-1}, x_{1:t}, \theta)$ depends exclusively on x_t and θ and that the distribution of x_t given $(x_{1:t-1}, \theta)$ depends only on x_{t-1} and θ. The corresponding probability density functions are noted as follows:

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

where $p(x_1 \mid \theta) \equiv \mu_\theta(x_1)$ denotes a known density function of the stationary distribution given θ.

We also incorporate the correlation between y_t and x_{t+1}, which is conditional on x_t since we consider such an example, the realized stochastic volatility (RSV) model, for the financial time series (see e.g. a seminal work by Takahashi et al. (2009)) in our illustrative example. It is a stochastic volatility model with an additional measurement equation for the realized volatility. Let $y_t = (y_{1,t}, y_{2,t})'$ where $y_{1,t}$ and $y_{2,t}$ denote the daily log return and the logarithm of the realized volatility (variance) at time t. Let x_t denote the latent log volatility which is assumed to follow the

stationary AR(1) process. The RSV model is defined as follows:

$$y_{1,t} = \exp(x_t/2)\epsilon_t, \ \epsilon_t \sim \mathcal{N}(0,1), \ t = 1, \ldots, T \tag{3}$$

$$y_{2,t} = x_t + \xi + u_t, \ u_t \sim \mathcal{N}(0, \sigma_u^2), \ t = 1, \ldots, T \tag{4}$$

$$x_{t+1} = \mu + \phi(x_t - \mu) + \eta_t, \ \eta_t \sim \mathcal{N}(0, \sigma_\eta^2), \ t = 1, \ldots, T, \tag{5}$$

$$x_1 = \mu + \frac{1}{\sqrt{1 - \phi^2}} \eta_0, \ \eta_0 \sim \mathcal{N}(0, \sigma_\eta^2), \quad |\phi| < 1, \tag{6}$$

where

$$\begin{pmatrix} \epsilon_t \\ u_t \\ \eta_t \end{pmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & \rho\sigma_\eta \\ 0 & \sigma_u^2 & 0 \\ \rho\sigma_\eta & 0 & \sigma_\eta^2 \end{bmatrix} \right), \tag{7}$$

$\mathcal{N}(\mu, \Sigma)$ denotes a normal distribution with mean μ and covariance matrix Σ, and $\theta = (\mu, \phi, \sigma_\eta^2, \xi, \sigma_u^2, \rho)'$ is the static parameter vector. The correlation ρ 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 (e.g. Omori et al. (2007)). In this case, we express the dependence of y_t on x_{t+1} (or x_{t+1} on y_t) as follows.

$$p(y_t \mid x_{1:t+1}, y_{1:t-1}, \theta) = p(y_t \mid x_t, x_{t+1}, \theta) \equiv g_\theta(y_t \mid x_t, x_{t+1}), \quad t = 1, \ldots, n, \tag{8}$$

$$p(x_{t+1} \mid x_{1:t}, y_{1:t}, \theta) = p(x_{t+1} \mid x_t, y_t, \theta) \equiv f_\theta(x_{t+1} \mid x_t, y_t), \quad t = 1, \ldots, n-1. \tag{9}$$

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 $x_{s:t}$ given the observations $y_{s:t}$ with $t = s + L$ for $s = 1, 2 \ldots$, and its probability density function is given by

$$\pi(x_{s:t}, \theta \mid y_{s:t}) \propto p(\theta)\mu_\theta(x_s)g_\theta(y_s \mid x_s) \left\{ \prod_{j=s+1}^{t} f_\theta(x_j \mid x_{j-1}, y_{j-1})g_\theta(y_j \mid x_j) \right\}, \tag{10}$$

or, equivalently,

$$\pi(x_{s:t}, \theta \mid y_{s:t}) \propto p(\theta)\mu_\theta(x_s) \left\{ \prod_{j=s+1}^{t} f_\theta(x_j \mid x_{j-1})g_\theta(y_{j-1} \mid x_{j-1}, x_j) \right\} g_\theta(y_t \mid x_t). \tag{11}$$

2.2 Simple rolling-window sampler

We first describe a simple rolling-window sampler that is derived in a straightforward manner from the previous literature. The estimation procedure consists of two steps, each of which can be described in the framework of the SMC sampler in Del Moral et al. (2006) as follows. In Step 1,

suppose we have samples from the old target density $\pi(x_{s-1:t-1}, \theta \mid y_{s-1:t-1})$ with importance weight $W_{[s-1,t-1]}$ at time $t-1$ where the subscript $[s-1,t-1]$ implies that the weight is based on observations $y_{s-1:t-1}$. After we include an observation y_t, our new target density is $\pi(x_{s-1:t}, \theta \mid y_{s-1:t})$. Using the proposal kernel $K((x_{s-1:t-1}, \theta), (x_{s-1:t}, \theta))$, we update the weight

$$W_{[s-1,t]} = \frac{\pi(x_{s-1:t}, \theta \mid y_{s-1:t})L((x_{s-1:t}, \theta), (x_{s-1:t-1}, \theta))}{\pi(x_{s-1:t-1}, \theta \mid y_{s-1:t-1})K((x_{s-1:t-1}, \theta), (x_{s-1:t}, \theta))} \times W_{[s-1,t-1]}, \tag{12}$$

where $L((x_{s-1:t}, \theta), (x_{s-1:t-1}, \theta))$ is the artificial backward Markov kernel with $L \equiv 1$ in this step. The first factor on the right hand side of Equation (12) is the incremental weight to adjust that of the previous step.

In Step 2, we have samples from the old target density $\pi(x_{s-1:t}, \theta \mid y_{s-1:t})$ with the importance weight $W_{[s-1,t]}$ from Step 1. After we exclude the observation y_{s-1}, our new target density is $\pi(x_{s-1:t}, \theta \mid y_{s:t})$. Using the backward kernel $L((x_{s:t}, \theta), (x_{s-1:t}, \theta)) = \pi(x_{s-1:t}, \theta \mid y_{s:t})/\pi(x_{s:t}, \theta \mid y_{s:t}) = \pi(x_{s-1} \mid x_{s:t}, y_{s:t}, \theta)$, we update the weight

$$W_{[s,t]} = \frac{\pi(x_{s:t}, \theta \mid y_{s:t})L((x_{s:t}, \theta), (x_{s-1:t}, \theta))}{\pi(x_{s-1:t}, \theta \mid y_{s-1:t})K((x_{s-1:t}, \theta), (x_{s-1:t}, \theta))} \times W_{[s-1,t]}, \tag{13}$$

where $K((x_{s-1:t}, \theta), (x_{s-1:t}, \theta))$ is the artificial proposal kernel with $K \equiv 1$, and discard x_{s-1}. Additionally, we can refresh all the particles with the MCMC when we observe the weight degeneracy, which is also regarded as an importance sampling step in the SMC sampler with the unnormalized weight equal to one. We note that one can also update particles using the particle Gibbs sampler (Andrieu et al. (2010)). Details are given below.

Step 1. Assume that, at time $t-1$, we have a collection of particles $(x^n_{s-1:t-1}, \theta^n)$ with the importance weight $W^n_{[s-1,t-1]}$, $(n = 1, \ldots, N)$ which is a discrete approximation of $\pi(x_{s-1:t-1}, \theta \mid y_{s-1:t-1})$. We include a new observation y_t in the information set and aim to sample from $\pi(x_{s-1:t}, \theta \mid y_{s-1:t})$. Given the current sample $(x_{s-1:t-1}, \theta)$ from $\pi(x_{s-1:t-1}, \theta \mid y_{s-1:t-1})$, we propose a candidate x_t using some proposal density $q_{t,\theta}(x_t \mid x_{t-1}, y_t)$. Since the incremental weight is

$$\frac{\pi(x_{s-1:t}, \theta \mid y_{s-1:t})}{\pi(x_{s-1:t-1}, \theta \mid y_{s-1:t-1})q_{t,\theta}(x_t \mid x_{t-1}, y_t)} = \frac{p(x_t, y_t \mid x_{s-1:t-1}, y_{s-1:t-1}, \theta)}{q_{t,\theta}(x_t \mid x_{t-1}, y_t)p(y_t \mid y_{s-1:t-1})}$$

$$\propto \frac{f_\theta(x_t \mid x_{t-1}, y_{t-1})g_\theta(y_t \mid x_t)}{q_{t,\theta}(x_t \mid x_{t-1}, y_t)},$$

we generate $x^n_t \sim q_{t,\theta^n}(x^n_t \mid x^n_{t-1}, y_t)$ and compute the importance weight

$$W^n_{[s-1,t]} \propto \frac{f_{\theta^n}(x^n_t \mid x^n_{t-1}, y_{t-1})g_{\theta^n}(y_t \mid x^n_t)}{q_{t,\theta^n}(x^n_t \mid x^n_{t-1}, y_t)} \times W^n_{[s-1,t-1]}. \tag{14}$$

Finally, we compute some degeneracy criteria such as the effective sample size (ESS),

$$\text{ESS}_{[s-1:t]} \equiv \left[\sum_{n=1}^N \left\{ W^n_{[s-1,t]} \right\}^2 \right]^{-1}, \tag{15}$$

4

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

Step 2. We exclude the old observation y_{s-1} from the information set, and aim to sample from $\pi(x_{s:t}, \theta \mid y_{s:t})$ where the backward kernel is $\pi(x_{s-1:t}, \theta \mid y_{s:t})/\pi(x_{s:t}, \theta \mid y_{s:t}) = \pi(x_{s-1} \mid x_{s:t}, y_{s:t}, \theta) = p(x_{s-1} \mid x_s, \theta)$ given by

$$p(x_{s-1} \mid x_s, \theta) = \frac{\mu_\theta(x_{s-1}) f_\theta(x_s \mid x_{s-1})}{\mu_\theta(x_s)}. \tag{16}$$

Since the current sample is from $\pi(x_{s-1:t}, \theta \mid y_{s-1:t})$ where

$$\pi(x_{s-1:t}, \theta \mid y_{s-1:t}) \;\; = \;\; \frac{p(x_{s-1:t}, y_{s-1}, \theta \mid y_{s:t})}{p(y_{s-1} \mid y_{s:t})} = \frac{\pi(x_{s-1:t}, \theta \mid y_{s:t}) p(y_{s-1} \mid x_{s-1:t}, y_{s:t}, \theta)}{p(y_{s-1} \mid y_{s:t})},$$

the (unnormalized) incremental weight is

$$\frac{\pi(x_{s:t}, \theta \mid y_{s:t}) p(x_{s-1} \mid x_s, \theta)}{\pi(x_{s-1:t}, \theta \mid y_{s-1:t})} \;\; = \;\; \frac{p(y_{s-1} \mid y_{s:t})}{p(y_{s-1} \mid x_{s-1:t}, y_{s:t}, \theta)} \propto g_\theta(y_{s-1} \mid x_{s-1}, x_s)^{-1},$$

Thus, we update the importance weight

$$W_{[s,t]}^n \;\; \propto \;\; g_{\theta^n}(y_{s-1} \mid x_{s-1}^n, x_s^n)^{-1} \times W_{[s-1,t]}^n, \tag{17}$$

and discard x_{s-1}^n. If some degeneracy criteria are fulfilled, resample all the particles by implementing the MCMC algorithm as in Step 1. The above procedure is summarized in Algorithm 1.

Algorithm 1: Simple rolling-window sampler

Let $(x_{s-1:t-1}^n, \theta^n)$ denote the sample from $\pi(x_{s-1:t-1}, \theta \mid y_{s-1:t-1})$ with the weight $W_{[s-1,t-1]}^n$ $(n = 1, \ldots, N)$.

Step 1: Generate $x_t^n \sim q_{t,\theta^n}(x_t^n \mid x_{t-1}^n, y_t)$ and set

$$W_{[s-1,t]}^n \;\; \propto \;\; \frac{f_{\theta^n}(x_t^n \mid x_{t-1}^n, y_{t-1}) g_{\theta^n}(y_t \mid x_t^n)}{q_{t,\theta^n}(x_t^n \mid x_{t-1}^n, y_t)} \times W_{[s-1,t-1]}^n. \tag{18}$$

Step 2: Update the weight

$$W_{[s,t]}^n \;\; \propto \;\; g_{\theta^n}(y_{s-1} \mid x_{s-1}^n, x_s^n)^{-1} \times W_{[s-1,t]}^n, \tag{19}$$

and discard x_{s-1}^n.

2.3 Weight degeneracy problem

Using the importance weight (17) in Step 2 is obviously problematic because it would take an extremely high value when g_θ is close to 0. This causes the ESS to rapidly drop and triggers the MCMC update steps many times, which makes the estimation time-consuming. Further, 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 we shall see in illustrative examples in Section 4, this step also causes a serious degeneracy problem. In Section 3, we overcome this difficulty of the weight degeneracy by proposing a novel sampling method called "a double-block sampler"with the conditional SMC update.

3 Particle rolling MCMC with double-block sampler

We consider sampling a block of state variables when we add the new observation or remove the old observation. For example, we update values of $\{x^n_{t-K:t-1}\}^N_{n=1}$ in addition to generating $\{x^n_t\}^N_{n=1}$ when we learn the information of y_t. We call this process the forward block sampling (Step 1), and the backward block sampling (Step 2) can also be defined in a similar manner. The double-block sampler addresses the weight degeneracy problem by reducing the path dependence between the new particle and the old particle values that are not updated.

3.1 Idealised double-block sampler

We first consider the 'idealised' double-block sampler where we assume an appropriate $K + 1$ dimensional proposal distribution is available for the block sampling. In the framework of the SMC sampler, in Step 1, we have samples from the old target density $\pi(x_{s-1:t-1}, \theta \mid y_{s-1:t-1})$ and generate $x^\dagger_{t-K:t}$ from the proposal kernel $\pi(x^\dagger_{t-K:t}|x_{s-1:t-K-1}, y_{s-1:t}, \theta)$. The new target density is $\pi(x_{s-1:t-K-1}, x^\dagger_{t-K:t}, \theta \mid y_{s-1:t})$ with the backward kernel $\pi(x_{t-K:t-1} \mid x_{s-1:t-K-1}, y_{s-1:t-1}, \theta)$. In Step 2, we have samples from the old target density $\pi(x_{s-1:t}, \theta \mid y_{s-1:t})$ and generate a candidate $x^\dagger_{s:s+K-1}$ using the proposal kernel $\pi(x^\dagger_{s:s+K-1}|x_{s+K:t}, y_{s:t}, \theta)$. Our new target density is $\pi(x^\dagger_{s:s+K-1}, x_{s+K:t}, \theta \mid y_{s:t})$ with the backward kernel $\pi(x_{s-1:s+K-1} \mid x_{s+K:t}, y_{s-1:t}, \theta)$. Finally, we discard x_{s-1}. Details are given below.

Step 1. We include a new observation y_t in the information set, and sample from $\pi(x_{s-1:t}, \theta \mid y_{s-1:t})$. Given the current sample $(x_{s-1:t-K-1}, \theta)$ from $\pi(x_{s-1:t-K-1}, \theta \mid y_{s-1:t-1})$, we generate $x^\dagger_{t-K:t} \sim \pi(x^\dagger_{t-K:t} \mid x_{s-1:t-K-1}, y_{s-1:t}, \theta)$. Since

$$
\pi(x_{s-1:t-K-1}, x^\dagger_{t-K:t}, \theta|y_{s-1:t})
$$
$$
= \frac{p(x_{s-1:t-K-1}, x^\dagger_{t-K:t}, y_t, \theta|y_{s-1:t-1})}{p(y_t|y_{s-1:t-1})}
$$
$$
= \pi(x_{s-1:t-K-1}, \theta|y_{s-1:t-1}) \times \frac{p(x^\dagger_{t-K:t}, y_t|x_{s-1:t-K-1}, y_{s-1:t-1}, \theta)}{p(y_t|y_{s-1:t-1})}
$$
$$
= \pi(x_{s-1:t-K-1}, \theta|y_{s-1:t-1})\pi(x^\dagger_{t-K:t}|x_{s-1:t-K-1}, y_{s-1:t}, \theta) \times \frac{p(y_t|x_{s-1:t-K-1}, y_{s-1:t-1}, \theta)}{p(y_t|y_{s-1:t-1})}
$$

the unnormalized incremental weight is

$$\frac{\pi(x_{s-1:t-K-1}, x^{\dagger}_{t-K:t}, \theta | y_{s-1:t})\pi(x_{t-K:t-1} | x_{s-1:t-K-1}, y_{s-1:t-1}, \theta)}{\pi(x_{s-1:t-1}, \theta | y_{s-1:t-1})\pi(x^{\dagger}_{t-K:t} | x_{s-1:t-K-1}, y_{s-1:t}, \theta)} = \frac{p(y_t | x_{s-1:t-K-1}, y_{s-1:t-1}, \theta)}{p(y_t | y_{s-1:t-1})}.$$

Thus we let $x_{s-1:t} = (x_{s-1:t-K-1}, x^{\dagger}_{t-K:t})$ and update the importance weight as

$$W^n_{[s-1,t]} \quad \propto \quad p(y_t | x^n_{t-K-1}, y_{s-1:t-1}, \theta^n) \times W^n_{[s-1,t-1]}, \tag{20}$$

noting that $p(y_t | x_{s-1:t-K-1}, y_{s-1:t-1}, \theta) = p(y_t | x_{t-K-1}, y_{s-1:t-1}, \theta)$.

Step 2. We remove the old observation y_{s-1} from the information set, and sample from $\pi(x_{s-1:t}, \theta \mid y_{s:t})$. Given the current sample from $\pi(x_{s-1:t}, \theta | y_{s-1:t})$, we generate $x^{\dagger}_{s:s+K-1} \sim \pi(x^{\dagger}_{s:s+K-1} | x_{s+K:t}, y_{s:t}, \theta)$. The unnormalized incremental weight is

$$\frac{\pi(x^{\dagger}_{s:s+K-1}, x_{s+K:t}, \theta \mid y_{s:t})\pi(x_{s-1:s+K-1} \mid x_{s+K:t}, y_{s-1:t}, \theta)}{\pi(x_{s-1:t}, \theta | y_{s-1:t})\pi(x^{\dagger}_{s:s+K-1} | x_{s+K:t}, y_{s:t}, \theta)} = \frac{p(y_{s-1} | y_{s:t})}{p(y_{s-1} | x_{s+K:t}, y_{s:t}, \theta)},$$

since

$$\pi(x^{\dagger}_{s:s+K-1}, x_{s+K:t}, \theta \mid y_{s:t}) \quad = \quad \pi(x^{\dagger}_{s:s+K-1} | x_{s+K:t}, y_{s:t}, \theta)\pi(x_{s+K:t}, \theta \mid y_{s:t})$$

and

$$\begin{aligned}
\pi(x_{s-1:t}, \theta | y_{s-1:t}) &= \pi(x_{s-1:s+K-1} | x_{s+K:t}, y_{s-1:t}, \theta)\pi(x_{s+K:t}, \theta | y_{s-1:t}) \\
&= \pi(x_{s-1:s+K-1} | x_{s+K:t}, y_{s-1:t}, \theta) \times \frac{p(x_{s+K:t}, y_{s-1}, \theta | y_{s:t})}{p(y_{s-1} | y_{s:t})} \\
&= \pi(x_{s-1:s+K-1} | x_{s+K:t}, y_{s-1:t}, \theta)\pi(x_{s+K:t}, \theta | y_{s:t}) \times \frac{p(y_{s-1} | x_{s+K:t}, y_{s:t}, \theta)}{p(y_{s-1} | y_{s:t})}.
\end{aligned}$$

Noting that $p(y_{s-1} \mid x_{s+K:t}, y_{s:t}, \theta) = p(y_{s-1} \mid x_{s+K}, y_{s:t}, \theta)$, we set $x_{s:t} = (x^{\dagger}_{s:s+K-1}, x_{s+K:t})$ and update the importance weight

$$W^n_{[s,t]} \quad \propto \quad p(y_{s-1} | x^n_{s+K}, y_{s:t}, \theta^n)^{-1} \times W^n_{[s-1,t]}, \tag{21}$$

and discard x^n_{s-1}. The sampling algorithm is summarized in Algorithm 2.

Algorithm 2: Idealised double-block sampler

Let $(x^n_{s-1:t-1}, \theta^n)$ denote the sample from $\pi(x_{s-1:t-1}, \theta \mid y_{s-1:t-1})$ with the weight $W^n_{[s-1,t-1]}$ $(n = 1, \ldots, N)$.

Step 1: Generate $x^n_{t-K:t} \sim \pi(x^n_{t-K:t} \mid x^n_{s-1:t-K-1}, y_{s-1:t}, \theta^n)$ and set

$$W^n_{[s-1,t]} \quad \propto \quad p(y_t | x^n_{t-K-1}, y_{s-1:t-1}, \theta^n) \times W^n_{[s-1,t-1]}. \tag{22}$$

Step 2. Generate $x^n_{s:s+K-1} \sim \pi(x^n_{s-1:s+K-1} \mid x^n_{s+K:t}, y_{s:t}, \theta^n)$ and set

$$W^n_{[s,t]} \quad \propto \quad p(y_{s-1} \mid x^n_{s+K}, y_{s:t}, \theta^n)^{-1} \times W^n_{[s-1,t]}, \tag{23}$$

and discard x^n_{s-1}.

3.2 Practical double-block sampler

In practice, it is often difficult to find an 'idealised' proposal distribution for the block sampling. Hence, we adopt the approach of the conditional SMC update for the particle Gibbs sampler (Andrieu et al. (2010)), which considers the artificial target density ad generates a cloud of values for one particle path.

In Step 1, we have samples from the old target density $\pi(x_{s-1:t-1}, \theta \mid y_{s-1:t-1})$, and generate the indices $k_{t-K:t-1} = (k_{t-K}, \ldots, k_{t-1})$ and a cloud of particles from the proposal kernel ψ_θ defined in (43). The new target density is $\pi(x_{s-1:t-K-1}, x^\dagger_{t-K:t}, \theta \mid y_{s-1:t})$ with the backward kernel $\hat{\pi}/\pi(x_{s-1:t-K-1}, x^\dagger_{t-K:t}, \theta \mid y_{s-1:t})$ where $\hat{\pi}$ is defined in (44)[1]. We set $x_{s-1:t} = \left(x_{s-1:t-K-1}, x^\dagger_{t-K:t}\right)$ which is the sample from the new target density with the unnormalized incremental weight $\hat{p}(y_t \mid x^n_{t-K-1}, y_{s-1:t-1}, \theta^n)$ in (29).

In Step 2, we have samples from the old target density $\pi(x_{s-1:t}, \theta \mid y_{s-1:t})$, and generate the indices $k_{s-1:s+K-1}$ and a cloud of particles from the proposal kernel $\bar{\psi}_\theta$ defined in (48). The new target density is $\pi(x^\dagger_{s:s+K-1}, x_{s+K:t}, \theta \mid y_{s:t})$ with the backward kernel $\check{\pi}/\pi(x^\dagger_{s:s+K-1}, x_{s+K:t}, \theta \mid y_{s:t})$ where $\check{\pi}$ defined in (49)[2]. We set $x_{s:t} = \left(x^\dagger_{s:s+K-1}, x_{s+K:t}\right)$ which is the sample from the new target density with the unnormalized incremental weight $\hat{p}(y_{s-1} \mid x^n_{s+K}, y_{s:t}, \theta^n)^{-1}$ in (35). Details are given below.

3.2.1 Forward block sampling (Step 1)

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

$$a^{k_{j+1}}_j = k_j, \quad j = t - K, \ldots, t - 1. \tag{24}$$

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

After generating $x^{n,m}_{t-K:t}$ $(m = 1, \ldots, M)$, we choose one lineage to store as the next values of $x^n_{t-K:t}$. This is equivalent to sampling a random index k^*_t for the candidate $x^{n,m}_t$ and identifying

---

[1]The marginal density of $\hat{\pi}$ is $\pi(x_{s-1:t-K-1}, x^\dagger_{t-K:t}, \theta \mid y_{s-1:t})$ as shown in Proposition 5.1 with $x^\dagger_{t-K:t} = (x^{k^*_{t-K}}_{t-K}, \ldots, x^{k^*_t}_t)$.

[2]The marginal density of $\check{\pi}$ is $\pi(x^\dagger_{s:s+K-1}, x_{s+K:t}, \theta \mid y_{s:t})$ as shown in Proposition 5.3 with $x^\dagger_{s:s+K-1} = (x^{k^*_s}_s, \ldots, x^{k^*_{s+K-1}}_{s+K-1})$.

the ancestors for which indices are obtained by following the rule

$$a_j^{k_{j+1}^*} = k_j^*, \quad j = t - K, \ldots, t - 1. \tag{25}$$

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 $x_{t-K:t-1}^n$ that may be different from the lineages obtained in the previous step.

The detailed algorithm is provided below. We fix one lineage in Step 1-1(a) and implement the conditional SMC in Steps 1-1(b) and 1-1(c). The candidates for x_t^n are generated in Step 1-1(d) and we compute the importance weight for the n-th particle in the 'global particle filtering' in Step 1-2. The smoothing is implemented in Step 1-3.

1. We generate $x_{t-K:t}^n \sim \pi(x_{t-K:t}^n \mid x_{s-1:t-K-1}^n, y_{s-1:t}, \theta^n)$ using the conditional SMC update:

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

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

   where $x_{t-K:t-1}^n$ is a current sample with the importance weight $W_{[s-1,t-1]}^n$.

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

   (c) Sample $a_{j-1}^m \sim \mathcal{M}(V_{j-1,\theta^n}^{1:M})$ and $x_j^{n,m} \sim q_{j,\theta^n}(\cdot \mid x_{j-1}^{n,a_{j-1}^m}, y_j)$ 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

   $$V_{j,\theta^n}^m = \frac{v_{j,\theta^n}(x_{j-1}^{n,a_{j-1}^m}, x_j^{n,m})}{\sum_{i=1}^{M} v_{j,\theta^n}(x_{j-1}^{n,a_{j-1}^i}, x_j^{n,i})}, \tag{26}$$

   $$v_{j,\theta^n}(x_{j-1}^{n,a_{j-1}^m}, x_j^{n,m}) = \frac{f_{\theta^n}(x_j^{n,m} \mid x_{j-1}^{n,a_{j-1}^m}, y_{j-1}) g_{\theta^n}(y_j \mid x_j^{n,m})}{q_{j,\theta^n}(x_j^{n,m} \mid x_{j-1}^{n,a_{j-1}^m}, y_j)}, \quad m = 1, \ldots, M. \tag{27}$$

   (d) If $j < t - 1$, set $j \leftarrow j + 1$ and go to (c). Otherwise, sample $x_t^{n,m}$ $(m = 1, \ldots, M)$ and k_t^* as follows.

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

   (ii) Sample $a_{t-1}^{n,m} \sim \mathcal{M}(V_{t-1,\theta^n}^{1:M})$ and $x_t^{n,m} \sim q_{t,\theta^n}(\cdot \mid x_{t-1}^{n,a_{t-1}^m}, y_t)$ 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 (25).

2. Let $x_{s-1:t}^n = (x_{s-1}^n, \ldots, x_{t-K-1}^n, x_{t-K}^{n,k_{t-K}^*}, \ldots, x_t^{n,k_t^*})$ and compute the importance weight[3]

$$W_{[s-1,t]}^n \quad \propto \quad \hat{p}(y_t \mid x_{t-K-1}^n, y_{s-1:t-1}, \theta^n) \times W_{[s-1,t-1]}^n. \tag{28}$$

$$\hat{p}(y_t \mid x_{t-K-1}^n, y_{s-1:t-1}, \theta^n) = \frac{1}{M} \sum_{m=1}^{M} v_{t,\theta^n}(x_{t-1}^{n,a_{t-1}^m}, x_t^{n,m}), \tag{29}$$

where $\hat{p}(y_t \mid x_{t-K-1}^n, y_{s-1:t-1}, \theta^n)$ can be seen as the estimate of the intractable incremental weight $p(y_t \mid x_{t-K-1}^n, y_{s-1:t-1}, \theta^n)$ in (20) for the idealised double-block sampler.

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

$$\bar{V}_{j,\theta}^m \equiv \frac{V_{j,\theta}^m f_\theta(x_{j+1}^{k_{j+1}^*} \mid x_j^m, y_{j+1})}{\sum_{i=1}^{M} V_{j,\theta}^i f_\theta(x_{j+1}^{k_{j+1}^*} \mid x_j^i, y_{j+1})}, \quad m = 1, \ldots, M, \tag{30}$$

and set $x_{s-1:t}^n = (x_{s-1}^n, \ldots, x_{t-K-1}^n, x_{t-K}^{n,k_{t-K}^*}, \ldots, x_t^{n,k_t^*})$.

Algorithm 3 (Step 1) : Practical double-block sampler

Let $(x_{s-1:t-1}^n, \theta^n)$ denote the sample from $\pi(x_{s-1:t-1}, \theta \mid y_{s-1:t-1})$ with the weight $W_{[s-1,t-1]}^n$ $(n = 1, \ldots, N)$.

1. We generate $x_{t-K:t}^n \sim \pi(x_{t-K:t}^n \mid x_{s-1:t-K-1}^n, y_{s-1:t}, \theta^n)$ using the conditional SMC update:

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

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

   (b) Set $x_{t-K-1}^{n,a_{t-K-1}^m} = x_{t-K-1}^n$ and sample $x_{t-K}^{n,m} \sim q_{t-K,\theta^n}(\cdot \mid x_{t-K-1}^n, y_{t-K})$, $m \in \{1, \ldots, M\} \setminus \{k_{t-K}\}$. Let $j = t-K+1$.

   (c) Sample $a_{j-1}^m \sim \mathcal{M}(V_{j-1,\theta^n}^{1:M})$ and $x_j^{n,m} \sim q_{j,\theta^n}(\cdot \mid x_{j-1}^{n,a_{j-1}^m}, y_j)$, $m \in \{1, \ldots, M\} \setminus \{k_j\}$ where $V_{j-1,\theta^n}^{1:M}$ is given by (26).

   (d) If $j < t-1$, set $j \leftarrow j+1$ and go to (c). Otherwise, sample $x_t^{n,m}$ $(m = 1, \ldots, M)$ and k_t^* as follows.

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

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

      (iii) Sample $k_t^* \sim \mathcal{M}(V_{t,\theta^n}^{1:M})$ and obtain k_j^* $(j = t-1, \ldots, t-K)$ using (25).

2. Let $x_{s-1:t}^n = (x_{s-1}^n, \ldots, x_{t-K-1}^n, x_{t-K}^{n,k_{t-K}^*}, \ldots, x_t^{n,k_t^*})$ and update the weight

$$W_{[s-1,t]}^n \quad \propto \quad \hat{p}(y_t \mid x_{t-K-1}^n, y_{s-1:t-1}, \theta^n) \times W_{[s-1,t-1]}^n.$$

where $\hat{p}$ is defined in (29).

3. The particle simulation smoother. Generate $k_j^* \sim \mathcal{M}(\bar{V}_{j,\theta}^{1:M})$, $j = t-1, \ldots, t-K$, where $\bar{V}_{j,\theta}^m$ is given in (30), and set $x_{s-1:t}^n = (x_{s-1}^n, \ldots, x_{t-K-1}^n, x_{t-K}^{n,k_{t-K}^*}, \ldots, x_t^{n,k_t^*})$.

---

[3] we use the notation $\hat{p}(y_t \mid x_{t-K-1}^n, y_{s-1:t-1}, \theta^n)$ since it is an unbiased estimator of $p(y_t \mid x_{t-K-1}^n, y_{s-1:t-1}, \theta^n)$ as we shall show in Proposition 5.2.

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

1. (a) Sample k_{t-2} and k_{t-1} from $\{1, 2, 3, 4\}$ with probability $1/4$ and suppose $k_{t-2} = k_{t-1} = 1$. We set $x^{n,1}_{t-2} = x^n_{t-2}$, $x^{n,1}_{t-1} = x^n_{t-1}$ (with the red rectangle) and $(a^1_{t-2}, a^1_{t-1}) = (1, 1)$.

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

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

   (d) (i) Sample $x^{n,1}_t \sim q_{t,\theta^n}(\cdot \mid x^{n,1}_{t-1}, y_t)$.

      (ii) Sample $a^{n,m}_{t-1} \sim \mathcal{M}(V^{1:4}_{t-1,\theta^n})$ for $m \in \{2, 3, 4\}$ and suppose $a^2_{t-1} = 2$, $a^3_{t-1} = 1$, $a^4_{t-1} = 4$. Generate and $x^{n,m}_t \sim q_{t,\theta^n}(\cdot \mid x^{n,a^m_{t-1}}_{t-1}, y_t)$ for $m \in \{2, 3, 4\}$.

      (iii) Sample $k^*_t \sim \mathcal{M}(V^{1:4}_{t,\theta^n})$ and suppose $k^*_t = 3$. Using (25), we obtain $k^*_{t-1} = k^*_{t-2} = 1$ and select $(x^{n,3}_t, x^{n,1}_{t-1}, x^{n,1}_{t-2})$ with red lines.

2. Let $x^n_{s-1:t} = (x^n_{s-1}, \dots, x^n_{t-3}, x^{n,1}_{t-2}, x^{n,1}_{t-1}, x^{n,3}_t)$ and compute the importance weight.

3. Implement the particle simulation smoother to sample $(k^*_{t-2}, k^*_{t-1}, k^*_t)$ jointly. Generate $k^*_j \sim \mathcal{M}(\bar{V}^{1:4}_{j,\theta})$, $j = t-1, t-2$, recursively (with dotted lines) and suppose $k^*_{t-1} = 3$ and $k^*_{t-2} = 3$. We set $x^n_{s-1:t} = (x^n_{s-1}, \dots, x^n_{t-3}, x^{n,3}_{t-2}, x^{n,3}_{t-1}, x^{n,3}_t)$.

Figure 1: Forward block sampling and particle simulation smoother.
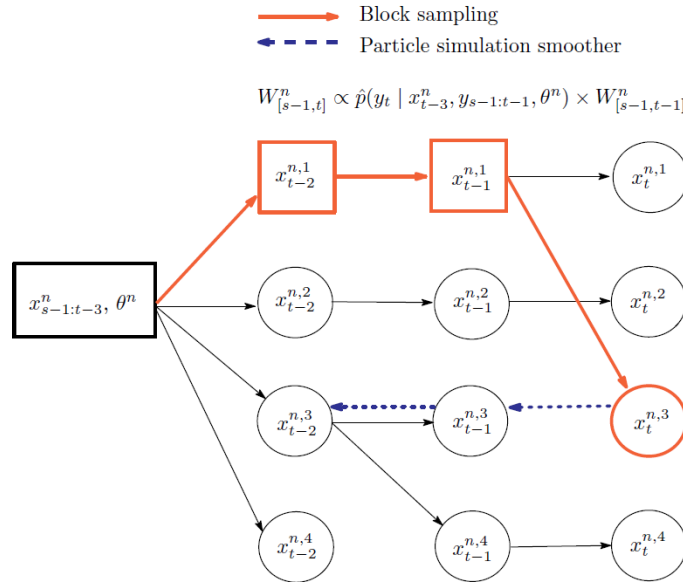
Remark 1. As the proposal density $q_{j,\theta}$, we can either use the prior density f_θ or more sophisticated density that incorporates the information of the likelihood g_θ. Even if we use the prior f_θ as the

proposal, the above sampling becomes much more efficient than the simple rolling-window sampler as shown in Section 4.

3.2.2 Backward block sampling (Step 2)

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

$$a^{k_j}_{j+1} = k_{j+1}, \quad j = s + K - 2, \ldots, s - 2. \tag{31}$$

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

1. We generate $x^n_{s-1:s+K-1} \sim \pi(x^n_{s-1:s+K-1} \mid x^n_{s+K:t}, y_{s:t}, \theta^n)$.

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

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

   where $x^n_{s-1:s+K-1}$ is a current sample with the importance weight $W^n_{[s-1,t]}$.

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

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

   $$V^m_{j,\theta^n} = \frac{v_{j,\theta^n}(x^{n,m}_j, x^{n,a^m_{j+1}}_{j+1})}{\sum_{i=1}^M v_{j,\theta^n}(x^{n,i}_j, x^{n,a^i_{j+1}}_{j+1})}, \tag{32}$$

   $$v_{j,\theta^n}(x^{n,m}_j, x^{n,a^m_{j+1}}_{j+1}) = \frac{p(x^{n,m}_j \mid x^{n,a^m_{j+1}}_{j+1}, \theta) g_{\theta^n}(y_j \mid x^{n,m}_j, x^{n,a^m_{j+1}}_{j+1})}{q_{j,\theta^n}(x^{n,m}_j \mid x^{n,a^m_{j+1}}_{j+1}, y_j)}, \quad m = 1, \ldots, M. \tag{33}$$

   (d) If $j > s - 1$, set $j \leftarrow j - 1$ and go to (c)[4]. Otherwise, sample $k^*_s \sim \mathcal{M}(V^{1:M}_{s,\theta^n})$ and obtain k^*_j $(j = s + 1, \ldots, s + K - 1)$ using (31).

---

[4]Note that we need to generate $x^{n,m}_{s-1}$ to compute $\hat{p}$ in (35).

2. Let $x_{s:t}^n = (x_s^{n,k_s^*}, \ldots, x_{s+K-1}^{n,k_{s+K-1}^*}, x_{s+K}^n, \ldots, x_t^n)$ and compute its importance weight

$$W_{[s,t]}^n \propto \begin{cases} \frac{1}{\hat{p}(y_{s-1}|x_{s+K}^n, y_{s:t}, \theta^n)} W_{[s-1,t]}^n, & \text{if } \hat{p}(y_{s-1} \mid x_{s+K}^n, y_{s:t}, \theta^n) \neq 0, \\ 0, & \text{if } \hat{p}(y_{s-1} \mid x_{s+K}^n, y_{s:t}, \theta^n) = 0, \end{cases} \tag{34}$$

where

$$\hat{p}(y_{s-1} \mid x_{s+K}^n, y_{s:t}, \theta^n) = \frac{1}{M} \sum_{m=1}^{M} v_{s-1,\theta^n}(x_{s-1}^{n,m}, x_s^{n,a_s^m}), \tag{35}$$

and $\hat{p}(y_{s-1} \mid x_{s+K}^n, y_{s:t}, \theta^n)^{-1}$ can be seen as the estimate of the intractable incremental weight $p(y_{s-1} \mid x_{s+K}^n, y_{s:t}, \theta^n)^{-1}$ in (21) for the idealised double-block sampler.

3. 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

$$\bar{V}_{j,\theta^n}^m = \frac{V_{j,\theta^n}^m p(x_{j-1}^{k_{j-1}^*} \mid x_j^m, \theta^n)}{\sum_{i=1}^M V_{j,\theta^n}^i p(x_{j-1}^{k_{j-1}^*} \mid x_j^i, \theta^n)}, \quad m = 1, \ldots, M. \tag{36}$$

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

Algorithm 3 (Step 2) : Practical double-block sampler

Let $(x_{s-1:t}^n, \theta^n)$ denote the sample from $\pi(x_{s-1:t}, \theta \mid y_{s-1:t})$ with the weight $W_{[s-1,t]}^n$ ($n = 1, \ldots, N$).

1. We generate $x_{s-1:s+K-1}^n \sim \pi(x_{s-1:s+K-1}^n \mid x_{s+K:t}^n, y_{s:t}, \theta^n)$ as follows:

   (a) Sample k_j from $\{1, \ldots, M\}$ with probability $1/M$ ($j = s+K-1, \ldots, s-1$) and set

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

   (b) Set $x_{s+K}^{n,a_{s+K}^m} = x_{s+K}^n$ and sample $x_{s+K-1}^{n,m} \sim q_{s+K-1,\theta^n}(\cdot \mid x_{s+K}^n, y_{s+K-1})$, $m \in \{1, \ldots, M\} \setminus \{k_{s+K-1}\}$. Let $j = s+K-2$.

   (c) Sample $a_{j+1}^m \sim \mathcal{M}(V_{j+1,\theta^n}^{1:M})$ and $x_j^{n,m} \sim q_{j,\theta^n}(\cdot \mid x_{j+1}^{n,a_{j+1}^m}, y_j)$, $m \in \{1, \ldots, M\} \setminus \{k_j\}$ where $V_{j+1,\theta^n}^{1:M}$ is given by (32).

   (d) If $j > s-1$, set $j \leftarrow j-1$ and go to (c). 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 (31).

2. Let $x_{s:t}^n = (x_s^{n,k_s^*}, \ldots, x_{s+K-1}^{n,k_{s+K-1}^*}, x_{s+K}^n, \ldots, x_t^n)$ and update the weight

   $$W_{[s,t]}^n \propto \hat{p}(y_{s-1} \mid x_{s+K}^n, y_{s:t}, \theta^n)^{-1} \times W_{[s-1,t]}^n, \tag{37}$$

   where $\hat{p}$ is defined in (35). If $\hat{p} = 0$, set $W_{[s,t]}^n = 0$.

3. The particle simulation smoother. Generate $k_j^* \sim \mathcal{M}(\bar{V}_{j,\theta^n}^{1:M})$, $j = s+1, \ldots, s+K-1$, where $\bar{V}_{j,\theta^n}^m$ is given in (36), and set $x_{s:t}^n = (x_s^{n,k_s^*}, \ldots, x_{s+K-1}^{n,k_{s+K-1}^*}, x_{s+K}^n, \ldots, x_t^n)$.

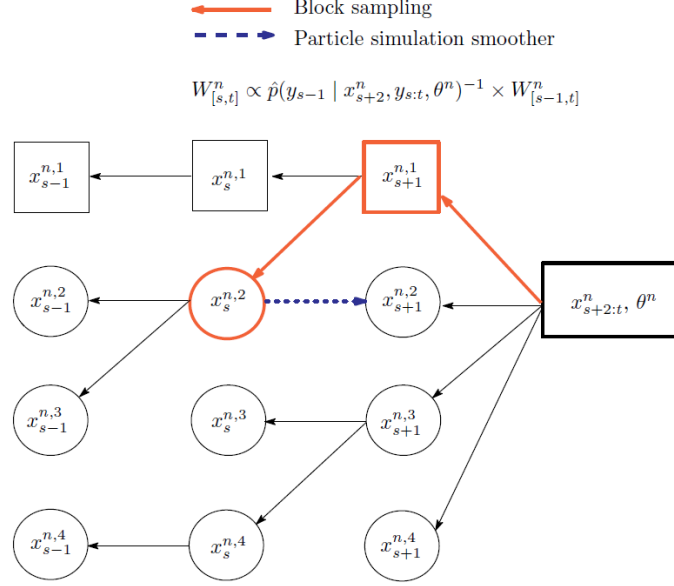Figure 2: Backward block sampling and particle simulation smoother.

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

1. (a) Sample indices k_{s+1}, k_s, k_{s-1} from $\{1, 2, 3, 4\}$ with probability $1/4$ and suppose $k_{s+1} = 1$, $k_s = 1$, $k_{s-1} = 1$. We set $(x_{s-1}^{n,1}, x_s^{n,1}, x_{s+1}^{n,1}) = x_{s-1:s+1}^n$ (with the rectangle) and $(a_{s-1}^1, a_s^1, a_{s+1}^1) = (1, 1, 1)$.

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

   (c) 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 $x_s^{n,m} \sim q_{s,\theta^n}(\cdot \mid x_{s+1}^{n,a_{s+1}^m}, y_s)$ for $m \in \{2, 3, 4\}$.

   (d) Sample $a_s^m \sim \mathcal{M}(V_{s,\theta^n}^{1:4})$ and suppose $a_s^2 = 2$, $a_s^3 = 2$, $a_s^4 = 4$. Generate $x_{s-1}^{n,m} \sim q_{s-1,\theta^n}(\cdot \mid x_s^{n,a_s^m}, y_{s-1})$ for $m \in \{2, 3, 4\}$.

   (e) Sample $k_s^* \sim \mathcal{M}(V_{s,\theta^n}^{1:4})$ and suppose $k_s^* = 2$. Using (31), we obtain $k_{s+1}^* = 1$, and select $(x_s^{n,2}, x_{s+1}^{n,1})$ with red lines.

2. Let $x_{s:t}^n = (x_s^{n,2}, x_{s+1}^{n,1}, x_{s+2}^n, \dots, x_t^n)$ and compute its importance weight.

3. 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 $x_{s:t}^n = (x_s^{n,2}, x_{s+1}^{n,2}, x_{s+2}^n, \dots, x_t^n)$.

Remark 2. In Algorithm 3, we assume we can evaluate $p(x_{j-1} \mid x_j, \theta)$ given in (16).

Remark 3. In the simple rolling-window sampler, we reweighted the particles according to the likelihood $g_\theta(y_{s-1}|x_{s-1}, x_s)$ in Step 2, while the unbiased estimate of the conditional likelihood $\hat{p}(y_{s-1} \mid x_{s+K}^n, y_{s:t}, \theta^n)$ is used in the practical double-block sampler. Algorithm 3 substantially improves the weight degeneracy since we condition on $y_{s:t}$ and integrate out $(x_{s-1}, \ldots, x_{s+K-1})$.

3.3 Sequential MCMC estimation without rolling the window

In the above discussion, it is implicitly assumed that the initial particles approximating $\pi(x_{1:L+1}, \theta \mid y_{1:L+1})$ 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 from the initial posterior distribution. However, based on our proposed method for the rolling estimation, we can obtain samples of $x_{1:L+1}$ and θ sequentially, simply by skipping Step 2. 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 B.). This initializing algorithm can be used for the ordinary sequential learning of $\pi(x_{1:t}, \theta \mid y_{1:t})$ $(t = 1, \ldots, T)$. We note that this approach is derived from the particle Gibbs scheme in Andrieu et al. (2010), and hence 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 Illustrative examples

This section demonstrates the efficiencies of our proposed algorithm using two illustrative examples. The simple rolling-window sampler suffers from the serious weight degeneracy problem, while (the idealised and the practical) double-block samplers overcome such difficulties. To evaluate the weight degeneracy in each of Steps 1 and 2, we define two ratios:

$$R_{1t} = \frac{\text{ESS}_{[s-1:t]}}{\text{ESS}_{[s-1:t-1]}}, \quad R_{2t} = \frac{\text{ESS}_{[s:t]}}{\text{ESS}_{[s-1:t]}}. \tag{38}$$

The ratio R_{1t} measures the relative change of ESS in Step 1 after adding y_t 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 change of ESS in Step 2 after removing y_{s-1} compared with that of the previous step.

4.1 Linear Gaussian state space model

We first consider the following univariate linear Gaussian state space model:

$$
\begin{aligned}
y_t &= x_t + \epsilon_t, \ \epsilon_t \sim \mathcal{N}(0, \sigma^2), \ t = 1, \ldots, 2000 \\
x_{t+1} &= \mu + 0.25(x_t - \mu) + \eta_t, \ \eta_t \sim \mathcal{N}(0, 2\sigma^2), \ t = 1, \ldots, 2000, \\
x_1 &= \mu + \frac{\eta_0}{\sqrt{1 - 0.25^2}}, \ \eta_0 \sim \mathcal{N}(0, 2\sigma^2),
\end{aligned}
$$

where $\theta = (\mu, \sigma^2)'$ is a parameter vector. We adopt weak conjugate priors, $\mu \mid \sigma^2 \sim \mathcal{N}(0, 10\sigma^2)$ and $\sigma^2 \sim \mathcal{IG}(5/2, 0.05/2)$ where $\mathcal{IG}(a, b)$ denotes an inverse gamma distribution with shape parameter a and scale parameter b. The rolling estimation is conducted with a window $[t - 999, t]$, $t = 1001, \ldots, 2000$ and $N = 1000$ using the particle rolling MCMC with and without the double-block sampling. We choose $K = 1, 2, 3, 5$ and 10 to investigate the effect of the block size. Since an idealised double-block sampler is feasible in the linear Gaussian state space model, we compare the following three samplers:

1. Simple rolling-window sampler (as a benchmark).

2. Idealised double-block sampler.

3. Practical double-block sampler with $M = 100, 300$ and 500.

Table 1 shows the number of resampling steps for three samplers. For the simple rolling-window sampler, the resampling steps are triggered 1027 times, while they are drastically reduced for the double-block samplers. They decrease as we increase K where the magnitude of the reduction is largest at $K = 2$. For $K = 2$, they are around 0.8% and 7.2% of the simple rolling-window sampler for the idealised and practical double-block samplers respectively. Additionally, the number of resampling steps of the practical double-block sampler decreases to that of the idealised double-block sampler as M increases.

Table 1: The number of resampling steps of three samplers.

Simple		Idealised	Practical		
	K		$M : 100$	300	500
	1	48	104	71	61
	2	8	74	33	23
1027	3	7	74	31	22
	5	6	72	32	22
	10	5	69	31	23

Figure 3 shows histograms of R_{1t} and R_{2t} for the simple rolling-window sampler and the practical double-block sampler with $K = 2$ and $M = 100$. The ratios R_{1t} and R_{2t} measure the relative magnitude of the effective sample size in Step 1 and Step 2 after adding y_t and removing y_{s-1} respectively when compared with that of the previous step at time t. The R_{1t} values for the practical double-block sampler are larger and less dispersed compared with those for the simple rolling-window sampler, suggesting that the forward block sampling is more efficient. Additionally, the R_{2t} values for the practical double-block sampler are much larger and much less dispersed than those for the simple rolling-window sampler, which implies that the backward block sampling is much more efficient.

Figure 3: The histograms of R_{1t} (left) and R_{2t} (right) ($t = 1001, \ldots, 2000$) for the simple rolling-window sampler (dotted blue) and the practical double-block sampler with $K = 2$ and $M = 100$ (solid red).
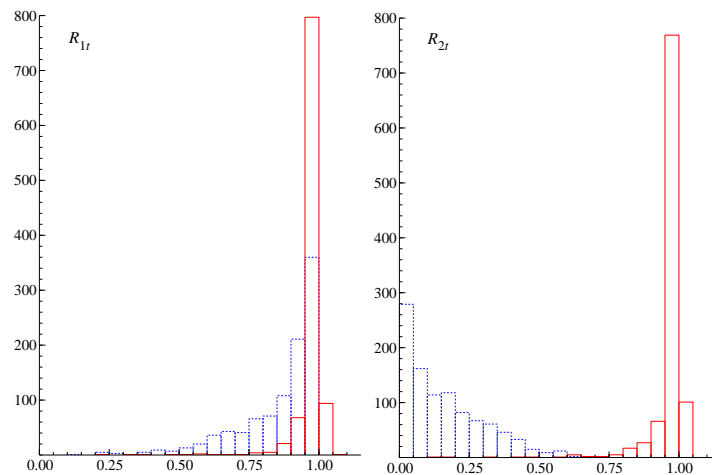
Further, the scatter plots of R_{1t} and R_{2t} are shown in Figure 4 for two sampling methods. These results demonstrate that our practical double-block sampler is more efficient at both Steps 1 and 2 of each rolling step.

17

Figure 4: The scatter plot of R_{2t} versus R_{1t} ($t = 1001, \ldots, 2000$) for the simple rolling-window sampler (blue plus) and the practical double-block sampler with $K = 2$ and $M = 100$ (red circle).
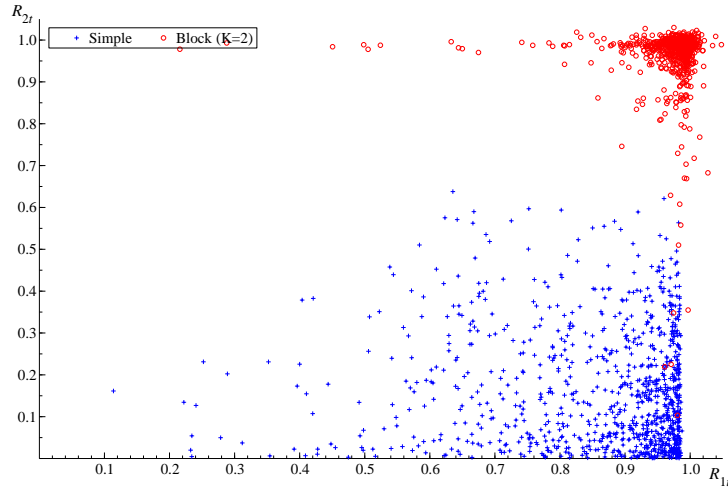

Table 2 shows the summary statistics of R_{1t} and R_{2t}. The average of R_{1t} for the practical double-block sampler is slightly larger than that for the simple rolling-window sampler, but the standard deviation for the former is less than half of that for the latter. Moreover, the average of R_{2t} for the double-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 practical double-block sampler drastically alleviate the weight degeneracy compared with the simple rolling-window sampler.

Table 2: Summary statistics of R_{1t} and R_{2t} for the simple rolling-window sampler and the practical double-block sampler ($K = 2, M = 100$)

	Method	Mean	Std. dev.
R_{1t}	Simple	0.862	0.145
	Practical	**0.975**	**0.057**
R_{2t}	Simple	0.161	0.139
	Practical	**0.970**	**0.068**

Finally, to assess the accuracy of the practical double-block sampler (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. In Figure 5, the algorithm seems to correctly capture both means and 95% credible intervals of the target posterior distribution. In Figure 6, true log marginal likelihoods and their estimates are shown in with errors. The estimation errors are very small overall, implying that the proposed

18

algorithm estimates the marginal likelihood $p(y_{t-999:t})$ accurately for $t = 1001, \ldots, 2000$.

Figure 5: True posterior means and 95% credible intervals (dotted black) with their estimates (solid red) for μ and σ^2 using the practical double-block sampler.
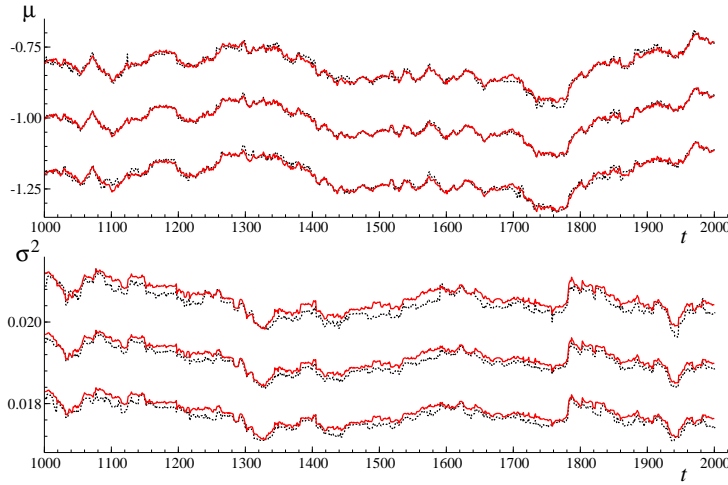
Figure 6: Top: true log marginal likelihoods $\log p(y_{t-999:t})$ (dotted black) and their estimates (solid red) by the practical double block sampler. Bottom: estimation errors $\log \hat{p}(y_{t-999:t}) - \log p(y_{t-999:t})$ for $t = 1001, \ldots, 2000$.
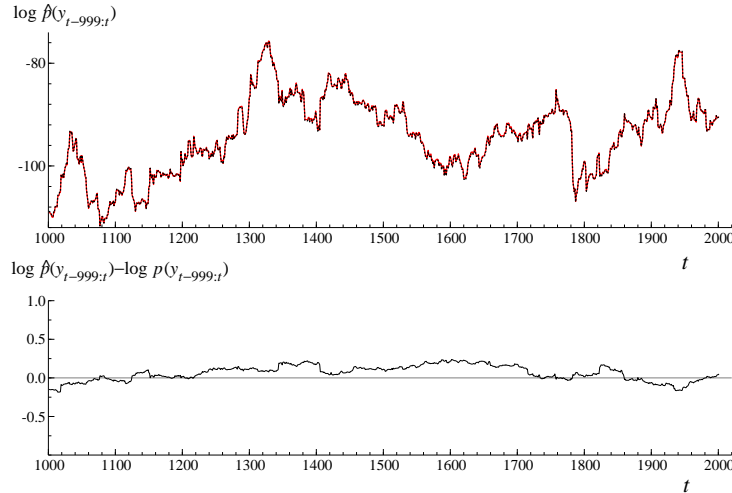
4.2 Realized stochastic volatility model

This subsection considers the RSV model given by (3)-(7) where the idealised double-block sampler is not feasible. For the static parameter $\theta = (\mu, \phi, \sigma_\eta^2, \xi, \sigma_u^2, \rho)'$, we assume the prior distributions

as in Takahashi et al. (2009):

$$\frac{\phi + 1}{2} \sim \mathcal{B}(20, 1.5), \ c \sim \mathcal{N}(0, 10), \ \sigma_u^2 \sim \mathcal{IG}(5/2, 0.05/2), \tag{39}$$

$$\Sigma \sim \mathcal{IW}(5, \Sigma_0), \ \Sigma_0 = \left(5 \begin{bmatrix} 1 & -0.3 \times 0.1 \\ -0.3 \times 0.1 & 0.01 \end{bmatrix} \right)^{-1}. \tag{40}$$

using the transformation

$$\sigma_\epsilon = \exp(\mu/2), \ c = \xi + \mu, \ \Sigma = \begin{bmatrix} \sigma_\epsilon^2 & \rho \sigma_\epsilon \sigma_\eta \\ \rho \sigma_\epsilon \sigma_\eta & \sigma_\eta^2 \end{bmatrix}, \tag{41}$$

where $\mathcal{B}(a, b)$ and $\mathcal{IW}(r, S)$ denote a beta distribution with parameters (a, b), and an inverse Wishart distribution with r degrees-of-freedom and the scale matrix S respectively. 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[5] created by Heber et al. (2009) (see Shephard and Sheppard (2010) for details). The initial estimation period is from January 1, 2000 ($t = 1$) to December 31, 2007 ($t = 1988$) with $L + 1 = 1988$. The rolling estimation started after 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.

We first implement the simple rolling-window sampler. 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 the MCMC sampling). We set $N = 1000$ and construct the proposal density $q_{t,\theta}(x_t \mid x_{t-1}, y_{s-1:t})$ based on the normal mixture approximation (see Omori et al. (2007)), which is expected to improve the weight degeneracy. Table 3 presents a summary of R_{1t} and R_{2t}. As expected, R_{2t}'s are low, so the update with MCMC kernel should be implemented in almost every step. The results for R_{1t}'s also indicate 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.

Table 3: Summary statistics for R_{1t} and R_{2t} ($t = 1988, \ldots, 2248$)
for the simple rolling-window sampler.

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

Next, we implement the practical double-block sampler with[6] $M = 300$ and $N = 1000$. Further we always implement 10 MCMC iterations below unless otherwise stated. As a proposal density, we

---

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

[6]We also tried using other values of M but the computation time is the shortest with $M = 300$.

simply use a prior density $q_{t,\theta}(x_t \mid x_{t-1}, y_{s-1:t}) = f_\theta(x_t \mid x_{t-1})$ to demonstrate that the practical double-block sampler improves even when using the simple 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 rolling-window algorithm, both means are close to 1 demonstrating that our proposed 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 $K = 15$.

Table 4: Summary statistics for R_{1t} and R_{2t} ($t = 1988, \ldots, 4248$)
using the practical double-block sampler.

	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

Figure 7 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 ($t = 4248$). From the rolling estimation results, we are able to observe the transition of the economic structure and the effect of the financial crisis ($t = 2150, \ldots, 2213$ correspond to September, October and November in 2008) . The posterior distribution of μ seems to be stable before $t = 4000$ (January 7, 2016), but its mean and 95% intervals decrease after $t = 4000$. The average level of log volatility started to decrease sharply 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 7: Trace plot of estimated posterior means and 95% credible intervals for parameters for S&P500 return in RSV model (from December 31, 2007 to December 30, 2016) using the practical double-block sampler with 5 (dotted blue) and 10 (solid red) iterations for the MCMC step.
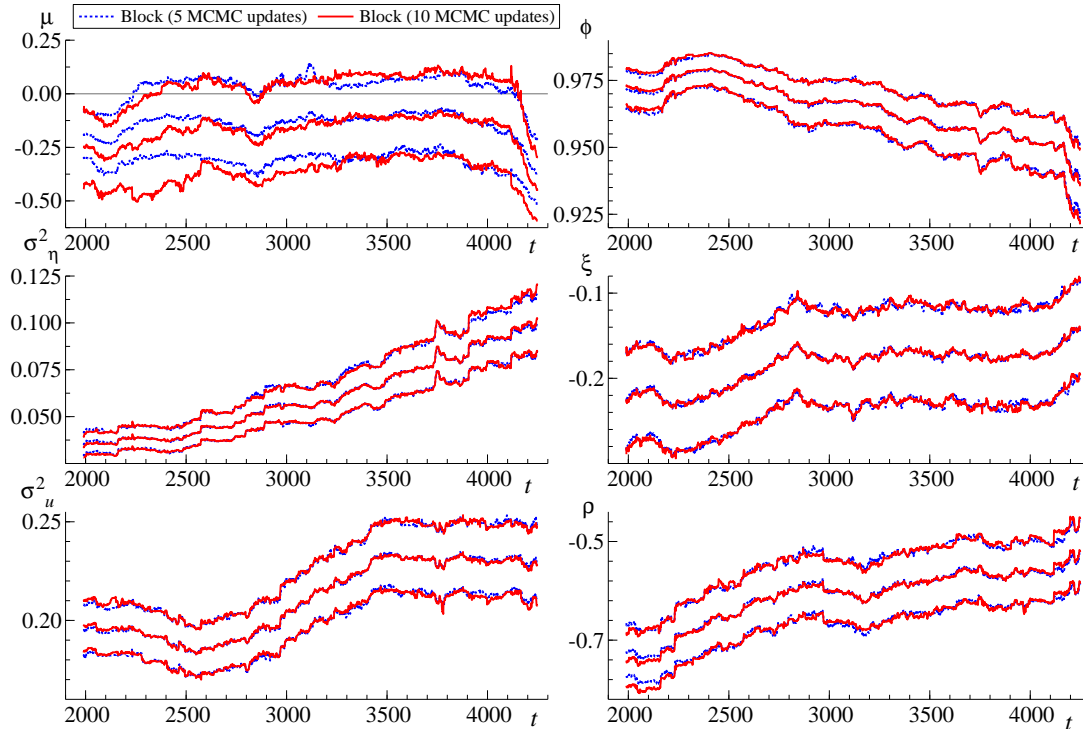
Figure 8 shows three cumulative computation times (wall time) for the same period corresponding to $K = 5, 10$ and 15. The computation times with $K = 5$ and $K = 15$ are longer than that with $K = 10$. This finding implies that, when $K = 5$, the effect of the blocking is not sufficient to reduce the path dependence between x_t and x_{t-K-1} (similarly, x_{s-1} and x_{s+K}). When $K = 15$, the Monte Carlo error in the local conditional SMC increased the variance of the importance weights even though there is a certain decrease in the variance due to the increase in K (we shall see more details in Section 5)[7].

---

[7]Also see Supplementary Material C for the comparison of the computation time of the practical double-block sampler with those of the MCMC and the particle MCMC.

Figure 8: Cumulative computation times (wall time, unit time = second) ($t = 1988, \ldots, 4248$).
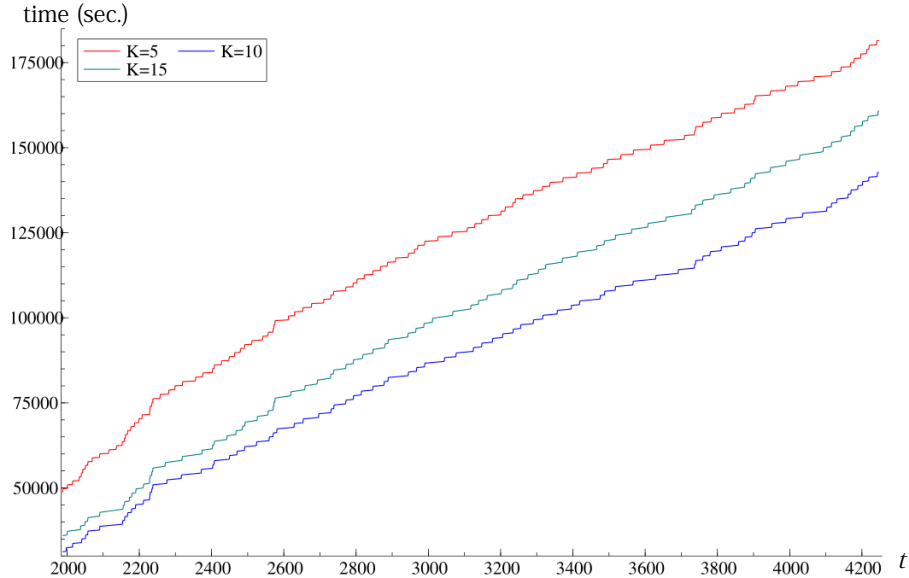
Finally, in Figure 9, we investigate the effect of the number of iterations in the MCMC steps on the estimation accuracy of the posterior distribution function of θ for the proposed sampling algorithm. The estimation period is from January 1, 2000 to December 31, 2007 ($t = 1, \ldots, 1988$). First, the MCMC sampling is conducted to obtain the accurate estimates of the distribution functions (solid gray). Then we apply our practical double-block sampler with $K = 10$, $M = 300$ and $N = 1000$ for three cases: one, five and ten MCMC updates. Among three cases, the estimates obtained by 5 or 10 iterations are close to those obtained by the exact MCMC sampling. 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 $x_{s-1:t}^n$. 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.

Figure 9: The estimated posterior cumulative distribution functions of θ.

MCMC (solid gray) and practical double-block sampler:

1 (dashed green), 5 (dotted blue) and 10 (solid red) iterations.
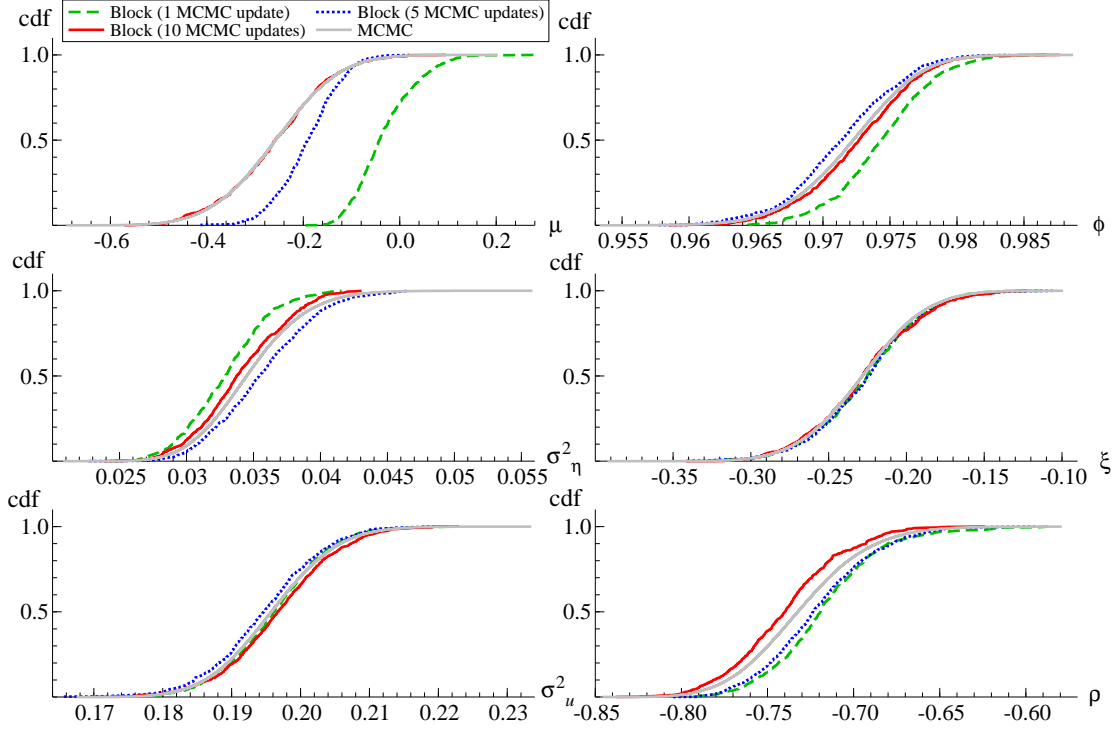
5 Theoretical justification

In this section, we provide theoretical justifications of our proposed algorithm in Section 3. We prove that our posterior density is obtained as a marginal density of the artificial target density.

5.1 Forward block sampling

The artificial target density and its marginal density. We prove that our posterior density of $(x^n_{s-1:t}, \theta^n)$ 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(a) of Section 3.2.1, the probability density function of $(x^{k_{t-K}}_{t-K}, \ldots, x^{k_{t-1}}_{t-1}) = x_{t-K:t-1}$ and $(a^{k_{t-K+1}}_{t-K}, \ldots, a^{k_t}_{t-1})$ given (x_{t-K-1}, θ) and $y_{t-K:t-1}$ is

$$p(x_{t-K:t-1}, a^{k_{t-K+1}}_{t-K}, \ldots, a^{k_t}_{t-1} \mid x_{t-K-1}, y_{t-K:t-1}, \theta) = \frac{\pi(x_{t-K:t-1} \mid x_{t-K-1}, y_{t-K:t-1}, \theta)}{M^K}. \tag{42}$$

Let $a^{1:M}_j = (a^1_j, \ldots, a^M_j)$, $x^{1:M}_j = (x^1_j, \ldots, x^M_j)$ and $a^{-k_{j+1}}_j \equiv a^{1:M}_j \setminus a^{k_{j+1}}_j = a^{1:M}_j \setminus k_j$ for $j = t-K, \ldots, t-1$ where we note $a^{k_{j+1}}_j = k_j$ and $k_t = 1$ in (24). Further, let $a^{1:M}_{t-K:t-1} = \{a^{1:M}_{t-K}, \ldots, a^{1:M}_{t-1}\}$,

and $x_j^{-k_j} = \{x_j^{a_j^1}, \ldots, x_j^{a_j^M}\}\backslash x_j^{k_j}$. Then, in 1(b), 1(c) and 1(d) of Step 1, given $x_{t-K-1}, (x_{t-K}^{k_{t-K}}, \ldots, x_{t-1}^{k_{t-1}}) = x_{t-K:t-1}$ 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

$$
\psi_\theta\left(x_{t-K}^{-k_{t-K}}, \ldots, x_{t-1}^{-k_{t-1}}, x_t^{1:M}, a_{t-K}^{-k_{t-K+1}}, \ldots, a_{t-1}^{-k_t}, k_t^* \mid x_{t-K-1:t-1}, a_{t-K}^{k_{t-K+1}}, \ldots, a_{t-1}^{k_t}, y_{t-K:t}\right)
$$

$$
= \prod_{\substack{m=1 \\ m \neq k_{t-K}}}^{M} q_{t-K,\theta}(x_{t-K}^m \mid x_{t-K-1}, y_{t-K}) \times \prod_{j=t-K+1}^{t-1} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j-1,\theta}^{a_{j-1}^m} q_{j,\theta}(x_j^m \mid x_{j-1}^{a_{j-1}^m}, y_j)
$$

$$
\times q_{t,\theta}(x_t^1 \mid x_{t-1}^{k_{t-1}}, y_t) \times \prod_{m=2}^{M} V_{t-1,\theta}^{a_{t-1}^m} q_{t,\theta}(x_t^m \mid x_{t-1}^{a_{t-1}^m}, y_t) \times V_{t,\theta}^{k_t^*}. \tag{43}
$$

In Step 1-2, we multiply $W_{[s-1,t-1]}$ by $\hat{p}(y_t \mid x_{t-K-1}^n, y_{s-1:t-1}, \theta^n)$ to adjust the importance weight for $W_{[s-1,t]}$. Let $x_{t-K:t}^{1:M} = (x_{t-K}^{1:M}, \ldots, x_t^{1:M})$ and $a_{t-K:t-1}^{1:M} = (a_{t-K}^{1:M}, \ldots, a_{t-1}^{1:M})$. Our artificial target density (before the particle smoother step) is written as

$$
\hat{\pi}(x_{s-1:t-K-1}, x_{t-K:t}^{1:M}, a_{t-K:t-1}^{1:M}, k_t^*, \theta \mid y_{s-1:t})
$$

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

$$
\times \psi_\theta(x_{t-K}^{-k_{t-K}}, \ldots, x_{t-1}^{-k_{t-1}}, x_t^{1:M}, a_{t-K}^{-k_{t-K+1}}, \ldots, a_{t-1}^{-k_t}, k_t^* \mid x_{t-K-1:t-1}, a_{t-K}^{k_{t-K+1}}, \ldots, a_{t-1}^{k_t}, y_{t-K:t})
$$

$$
\times \frac{\hat{p}(y_t \mid x_{t-K-1}, y_{s-1:t-1}, \theta)}{p(y_t \mid y_{s-1:t-1})}
$$

$$
= \frac{\pi(x_{s-1:t-1}, \theta \mid y_{s-1:t-1})}{M^K}
$$

$$
\times \prod_{\substack{m=1 \\ m \neq k_{t-K}}}^{M} q_{t-K,\theta}(x_{t-K}^m \mid x_{t-K-1}, y_{t-K}) \times \prod_{j=t-K+1}^{t-1} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j-1,\theta}^{a_{j-1}^m} q_{j,\theta}(x_j^m \mid x_{j-1}^{a_{j-1}^m}, y_j)
$$

$$
\times \quad q_{t,\theta}(x_t^1 \mid x_{t-1}^{k_{t-1}}, y_t) \times \prod_{m=2}^{M} V_{t-1,\theta}^{a_{t-1}^m} q_{t,\theta}(x_t^m \mid x_{t-1}^{a_{t-1}^m}, y_t) \times V_{t,\theta}^{k_t^*}
$$

$$
\times \frac{\hat{p}(y_t \mid x_{t-K-1}, y_{s-1:t-1}, \theta)}{p(y_t \mid y_{s-1:t-1})}. \tag{44}
$$

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 5.2. The proposed forward block sampling is justified by proving that the marginal density of $(x_{s-1}, \ldots, x_{t-K-1}, x_{t-K}^{k_{t-K}^*}, \ldots, x_t^{k_t^*}, \theta)$ in the above artificial target density $\hat{\pi}$ is $\pi(x_{s-1}, \ldots, x_{t-K-1}, x_{t-K}^{k_{t-K}^*}, \ldots, x_t^{k_t^*}, \theta \mid y_{s-1:t})$.

Proposition 5.1. *The artificial target density $\hat{\pi}$ for the forward block sampling can be written as*

$$\hat{\pi}(x_{s-1:t-K-1}, x_{t-K:t}^{1:M}, a_{t-K:t-1}^{1:M}, k_t^*, \theta \mid y_{s-1:t})$$

$$= \frac{\pi(x_{s-1:t-K-1}, x_{t-K}^{k_{t-K}^*}, \ldots, x_t^{k_t^*}, \theta \mid y_{s-1:t})}{M^{K+1}} \times \prod_{\substack{m=1 \\ m \neq k_{t-K}^*}}^{M} q_{t-K,\theta}(x_{t-K}^m \mid x_{t-K-1}, y_{t-K})$$

$$\times \prod_{j=t-K+1}^{t} \prod_{\substack{m=1 \\ m \neq k_j^*}}^{M} V_{j-1,\theta}^{a_{j-1}^m} q_{j,\theta}(x_j^m \mid x_{j-1}^{a_{j-1}^m}, y_j), \tag{45}$$

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

Proof. See Supplementary Material A.

Proposition 5.1 implies that we can obtain a posterior random sample $(x_{s-1:t}, \theta)$ 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.

Properties of the incremental weight. We consider the mean and variance of the (unnormalized) incremental weight, $\hat{p}(y_t \mid y_{s-1:t-1}, x_{t-K-1}, \theta)$. Proposition 5.2 shows that this weight can be considered an unbiased estimator.

Proposition 5.2. *If*

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

and

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

where ψ_θ is given in (43), then

$$E[\hat{p}(y_t \mid x_{t-K-1}, y_{s-1:t-1}, \theta) \mid y_{s-1:t}] = E[p(y_t \mid x_{t-K-1}, y_{s-1:t-1}, \theta) \mid x_{t-K-1}, y_{s-1:t}, \theta]$$

$$= p(y_t \mid y_{s-1:t-1}).$$

Proof. See Supplementary Material A.

This shows that the incremental weight $\hat{p}(y_t \mid x_{t-K-1}, y_{s-1:t-1}, \theta)$ is an unbiased estimator of the conditional likelihood $p(y_t \mid x_{t-K-1}, y_{s-1:t-1}, \theta)$ given (x_{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 $\hat{\pi}$.

26

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

$$\text{Var}[\hat{p}(y_t \mid x_{t-K-1}, y_{s-1:t-1}, \theta) \mid y_{s-1:t}]$$
$$= \text{Var}[p(y_t \mid x_{t-K-1}, y_{s-1:t-1}, \theta) \mid y_{s-1:t}]$$
$$+ E\left[\text{Var}[\hat{p}(y_t \mid x_{t-K-1}, y_{s-1:t-1}, \theta) \mid x_{s-1:t-K-1}, y_{s-1:t}, \theta]\right].$$

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:

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

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

$$E\left[p(y_t \mid x_{t-K_1-1}, y_{s-1:t-1}, \theta) \mid x_{s-1:t-K_2-1}, \theta\right] = p(y_t \mid x_{t-K_2-1}, y_{s-1:t-1}, \theta). \tag{46}$$

On the other hand, the second component may become large as K increases, but it is expected to be controlled by changing the number of particles M.

5.2 Backward block sampling

The artificial target density and its marginal density. This subsection proves that our posterior density of $(x_{s:t}^n, \theta^n)$ 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(a) of Section 3.2.2, the probability density function of $(x_{s-1}^{k_{s-1}}, \ldots, x_{s+K-1}^{k_{s+K-1}}) = x_{s-1:s+K-1}$ and $(a_{s-1}^{k_{s-2}}, \ldots, a_{s+K-1}^{k_{s+K-2}})$ given (x_{s+K}, θ) and $y_{s-1:t}$ is

$$p(x_{s-1:s+K-1}, a_{s-1}^{k_{s-2}}, \ldots, a_{s+K-1}^{k_{s+K-2}} \mid x_{s+K}, y_{s-1:t}, \theta) = \frac{\pi(x_{s-1:s+K-1} \mid x_{s+K}, y_{s-1:t}, \theta)}{M^{K+1}}. \tag{47}$$

In 1(b), 1(c) and 1(d) of Steps 2, given x_{s+K}, $(x_{s-1}^{k_{s-1}}, \ldots, x_{s+K-1}^{k_{s+K-1}}) = x_{s-1:s+K-1}$, $(a_{s-1}^{k_{s-2}}, \ldots, a_{s+K-1}^{k_{s+K-2}}) = (k_{s-1}, \ldots, k_{s+K-1})$ and $y_{s-1:s+K-1}$, the probability density function of all variables is defined as

$$\bar{\psi}_\theta(x_{s-1}^{-k_{s-1}}, \ldots, x_{s+K-1}^{-k_{s+K-1}}, a_s^{-k_{s-1}}, \ldots, a_{s+K-1}^{-k_{s+K-2}}, k_s^* \mid x_{s-1:s+K}, a_{s-1}^{k_{s-2}}, \ldots, a_{s+K-1}^{k_{s+K-2}}, y_{s-1:s+K-1})$$
$$= \prod_{\substack{m=1 \\ m \neq k_{s+K-1}}}^{M} q_{s+K-1,\theta}(x_{s+K-1}^m \mid x_{s+K}, y_{s+K-1}) \times \prod_{j=s-1}^{s+K-2} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(x_j^m \mid x_{j+1}^{a_{j+1}^m}, y_j) \times V_{s,\theta}^{k_s^*}.$$
$$\tag{48}$$

In Step 2-2, we divide $W_{[s-1,t]}$ by $\hat{p}(y_{s-1} \mid x_{s+K}^n, y_{s:t}, \theta^n)$ to adjust the importance weight for $W_{[s,t]}$. Similarly to the discussion in Section 5.1, we consider an extended space with the artificial target

density written as

$$\check{\pi}(x_{s-1:s+K-1}^{1:M}, x_{s+K:t}, a_{s:s+K-1}^{1:M}, k_{s-1}, k_s^*, \theta \mid y_{s-1:t})$$

$$\equiv \frac{\pi(x_{s-1:t}, \theta \mid y_{s-1:t})}{M^{K+1}}$$

$$\times \prod_{\substack{m=1 \\ m \neq k_{s+K-1}}}^{M} q_{s+K-1,\theta}(x_{s+K-1}^m \mid x_{s+K}, y_{s+K-1}) \times \prod_{j=s-1}^{s+K-2} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(x_j^m \mid x_{j+1}^{a_{j+1}^m}, y_j)$$

$$\times \quad V_{s,\theta}^{k_s^*} \times \frac{p(y_{s-1} \mid y_{s:t})}{\hat{p}(y_{s-1} \mid x_{s+K}, y_{s:t}, \theta)}, \tag{49}$$

where $p(y_{s-1} \mid y_{s:t})^{-1}$ is the normalizing constant of this target density as shown in Proposition 5.4. Below we state Proposition 5.3 for the backward block sampling, which correspond to Proposition 5.1 for the forward block sampling.

Proposition 5.3. *The artificial target density $\check{\pi}$ for the backward block sampling can be rewritten as*

$$\check{\pi}(x_{s-1:s+K-1}^{1:M}, x_{s+K:t}, a_{s:s+K-1}^{1:M}, k_{s-1}, k_s^*, \theta \mid y_{s-1:t})$$

$$= \frac{\pi(x_s^{k_s^*}, \ldots, x_{s+K-1}^{k_{s+K-1}^*}, x_{s+K:t}, \theta \mid y_{s:t})}{M^K} \times \prod_{\substack{m=1 \\ m \neq k_{s+K-1}^*}}^{M} q_{s+K-1,\theta}(x_{s+K-1}^m \mid x_{s+K}, y_{s+K-1:t})$$

$$\times \prod_{j=s}^{s+K-2} \prod_{\substack{m=1 \\ m \neq k_j^*}}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(x_j^m \mid x_{j+1}^{a_{j+1}^m}, y_j) \times \prod_{m=1}^{M} V_{s,\theta}^{a_s^m} q_{s-1,\theta}(x_{s-1}^m \mid x_s^{a_s^m}, y_{s-1}) \times V_{s-1,\theta}^{k_{s-1}}, \tag{50}$$

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

Proof. See Supplementary Material A.

Although the probability density (50) in Proposition 5.3 has a bit different form from that of (45) in Proposition 5.1, its marginal probability density is found to be the target posterior density $\pi(x_{s:t}, \theta \mid y_{s:t})$.

Properties of the incremental weight. Similar results to Proposition 5.2 hold for the backward block sampling, and are summarized in Proposition 5.4.

Proposition 5.4. *If*

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

and

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

where $\bar{\psi}_\theta$ is given in (48), then

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

Proof. See Supplementary Material A.

5.3 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 the particle Gibbs (see e.g. Chopin and Singh (2015), Lee et al. (2020)). We also incorporate such a particle simulation smoother into the double block sampling based on the following proposition.

Proposition 5.5. *The joint conditional density of* $(k^*_{t-K}, \ldots, k^*_t)$ *is given by*

$$
\begin{aligned}
&\hat{\pi}(k^*_{t-K}, \ldots, k^*_t | x_{s-1:t-K-1}, x^{1:M}_{t-K:t}, a^{1:M}_{t-K:t-1}, y_{s-1:t}, \theta) \\
&= \hat{\pi}(k^*_t | x_{s-1:t-K-1}, x^{1:M}_{t-K:t}, a^{1:M}_{t-K:t-1}, y_{s-1:t}, \theta) \\
&\times \prod_{t_0 = t-1}^{t-K} \hat{\pi}(k^*_{t_0} | x_{s-1:t-K-1}, x^{1:M}_{t-K:t_0}, a^{1:M}_{t-K:t_0-1}, x^{k^*_{t_0+1}}_{t_0+1}, \ldots, x^{k^*_t}_t, k^*_{t_0+1:t}, y_{s-1:t}, \theta), \qquad (51)
\end{aligned}
$$

where

$$
\begin{aligned}
&\hat{\pi}(k^*_{t_0} | x_{s-1:t-K-1}, x^{1:M}_{t-K:t_0}, a^{1:M}_{t-K:t_0-1}, x^{k^*_{t_0+1}}_{t_0+1}, \ldots, x^{k^*_t}_t, k^*_{t_0+1:t}, y_{s-1:t}, \theta) \\
&= \bar{V}^{k^*_{t_0}}_{t_0,\theta}, \quad \bar{V}^m_{j,\theta} \equiv \frac{V^m_{j,\theta} f_\theta(x^{k^*_{j+1}}_{j+1} \mid x^m_j, y_{j+1})}{\sum_{i=1}^M V^i_{j,\theta} f_\theta(x^{k^*_{j+1}}_{j+1} \mid x^i_j, y_{j+1})}. \qquad (52)
\end{aligned}
$$

Proof. See Supplementary Material A.

Suppose we have $(x_{s-1:t-K-1}, x^{1:M}_{t-K:t}, a^{1:M}_{t-K:t-1}, k^*_t, \theta) \sim \hat{\pi}$ where $\hat{\pi}$ is defined in (44). In Step 1-1(d), 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 $k^*_j \sim \mathcal{M}(\bar{V}^{1:M}_{j,\theta})$, $j = t-1, \ldots, t-K$, recursively.

6 Conclusion

In this paper, we propose a novel efficient estimation method to implement the rolling-window particle MCMC simulation using a SMC 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. The illustrative examples show that our proposed sampler outperforms the simple rolling-window sampler.

Acknowledgement

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.

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

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

Lee, A., S. S. Singh, and M. Vihola (2020). Coupled conditional backward sampling particle filter. *Annals of Statistics 48*(5), 3066–3089.

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.

Supplementary Material

A Proofs

A.1 Proof of Proposition 5.1

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

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

$$
\frac{\pi(x_{s-1:t-K-1}, x_{t-K}^{k_{t-K}}, \ldots, x_{t_0}^{k_{t_0}}, \theta \mid y_{s-1:t_0})}{M^{t_0 - (t-K) + 1}}
$$

$$
\times \prod_{\substack{m=1 \\ m \neq k_{t-K}}}^{M} q_{t-K,\theta}(x_{t-K}^m \mid x_{t-K-1}, y_{t-K}) \times \prod_{j=t-K+1}^{t_0} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j-1,\theta}^{a_{j-1}^m} q_{j,\theta}(x_j^m \mid x_{j-1}^{a_{j-1}^m}, y_j) \tag{53}
$$

$$
= \pi(x_{s-1:t-K-1}, \theta \mid y_{s-1:t-K-1}) \times \prod_{m=1}^{M} q_{t-K,\theta}(x_{t-K}^m \mid x_{t-K-1}, y_{t-K})
$$

$$
\times \prod_{j=t-K+1}^{t_0} \prod_{m=1}^{M} V_{j-1,\theta}^{a_{j-1}^m} q_{j,\theta}(x_j^m \mid x_{j-1}^{a_{j-1}^m}, y_j) \times V_{t_0,\theta}^{k_{t_0}} \times \prod_{j=t-K}^{t_0} \frac{\hat{p}(y_j \mid x_{t-K-1}, y_{s-1:j-1}, \theta)}{p(y_j \mid y_{s-1:j-1})}, \tag{54}
$$

where

$$
\hat{p}(y_j \mid x_{t-K-1}, y_{s-1:j-1}, \theta) = \frac{1}{M} \sum_{m=1}^{M} v_{j,\theta}(x_{j-1}^{a_{j-1}^m}, x_j^m), \quad j = t - K, \ldots, t_0, \tag{55}
$$

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

The probability density (53) 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 $x_{t-K:t-1}^{1:M}$ conditional on $(x_{s-1:t-K-1}, \theta)$, while SMC^2 considers that of $x_{1:t}^{1:M}$ conditional on θ. Further, the former updates the importance weight for $(x_{s-1:t}, \theta)$ and the latter updates that for θ sequentially.

Proof of Lemma A.1. Using Bayes' theorem and

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

the numerator of the first term in (53) is

$$
\pi(x_{s-1:t-K-1}, x_{t-K}^{k_{t-K}}, \ldots, x_{t_0}^{k_{t_0}}, \theta \mid y_{s-1:t_0})
$$

$$
= \frac{\pi(x_{s-1:t-K-1}, \theta \mid y_{s-1:t-K-1})}{p(y_{t-K:t_0} \mid y_{s-1:t-K-1})} \prod_{j=t-K}^{t_0} f_\theta(x_j^{k_j} \mid x_{j-1}^{k_{j-1}}, y_{j-1}) g_\theta(y_j \mid x_j^{k_j})
$$

$$
= \frac{\pi(x_{s-1:t-K-1}, \theta \mid y_{s-1:t-K-1})}{p(y_{t-K:t_0} \mid y_{s-1:t-K-1})} \prod_{j=t-K}^{t_0} v_{j,\theta}(x_{j-1}^{k_{j-1}}, x_j^{k_j}) \prod_{j=t-K}^{t_0} q_{j,\theta}(x_j^{k_j} \mid x_{j-1}^{k_{j-1}}, y_j). \tag{56}
$$

1

Thus we obtain

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

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

Using Lemma A.1, we obtain Proposition 5.1 as follows.

Proof of Proposition 5.1.

By applying Lemma A.1 with $t_0 = t-1$ to the first three terms of (44), we obtain

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

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. $\qquad\square$

2

A.2 Proof of Proposition 5.2

We first define the probability density function

$$
\psi_{\theta,0}(x_{t-K:t}^{1:M}, a_{t-K:t-1}^{1:M}, k_t^* \mid x_{t-K-1}, y_{s-1:t})
$$
$$
\equiv \frac{\pi(x_{t-K:t-1} \mid x_{t-K-1}, y_{s-1:t-1}, \theta)}{M^K}
$$
$$
\times \psi_\theta(x_{t-K}^{-k_{t-K}}, \ldots, x_{t-1}^{-k_{t-1}}, x_t^{1:M}, a_{t-K}^{-k_{t-K+1}}, \ldots, a_{t-1}^{-k_t}, k_t^* \mid x_{t-K-1:t-1}, a_{t-K}^{k_{t-K+1}}, \ldots, a_{t-1}^{k_t}, y_{t-K:t}),
$$

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

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

Noting that

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

where we used the definition of $\hat{\pi}$ in (44),

$$
E_{\psi_{\theta,0}}[\hat{p}(y_t \mid x_{t-K-1}, y_{s-1:t-1}, \theta) \mid x_{t-K-1}, y_{s-1:t}, \theta]
$$
$$
= \int \hat{p}(y_t \mid x_{t-K-1}, y_{s-1:t-1}, \theta) \psi_{\theta,0}(x_{t-K:t}^{1:M}, a_{t-K:t-1}^{1:M}, k_t^* \mid x_{t-K-1}, y_{s-1:t}) dx_{t-K:t}^{1:M} da_{t-K:t-1}^{1:M} dk_t^*
$$
$$
= \int \hat{\pi}(x_{s-1:t-K-1}, x_{t-K:t}^{1:M}, a_{t-K:t-1}^{1:M}, k_t^*, \theta \mid y_{s-1:t}) dx_{t-K:t}^{1:M} da_{t-K:t-1}^{1:M} dk_t^* \frac{p(y_t \mid y_{s-1:t-1})}{\pi(x_{s-1:t-K-1}, \theta \mid y_{s-1:t-1})}
$$
$$
= \frac{\pi(x_{s-1:t-K-1}, \theta \mid y_{s-1:t}) p(y_t \mid y_{s-1:t-1})}{\pi(x_{s-1:t-K-1}, \theta \mid y_{s-1:t-1})}
$$
$$
= p(y_t \mid x_{s-1:t-K-1}, y_{s-1:t-1}, \theta) = p(y_t \mid x_{t-K-1}, y_{s-1:t-1}, \theta).
$$

Also it is easy to see

$$
E[p(y_t \mid x_{t-K-1}, y_{s-1:t-1}, \theta) \mid y_{s-1:t}]
$$
$$
= \int p(y_t \mid x_{t-K-1}, y_{s-1:t-1}, \theta) \pi(x_{t-K-1}, \theta \mid y_{s-1:t-1}) d\theta dx_{t-K-1} = p(y_t \mid y_{s-1:t-1}).
$$

$\square$

A.3 Proof of Proposition 5.3

We first establish the following lemma as in the proof of Proposition 5.1.

Lemma A.2. *For any t, s_0, and s $(s-1 \leq s_0 \leq s+K-1)$,*

$$\frac{\pi(x_{s_0}^{k_{s_0}}, \ldots, x_{s+K-1}^{k_{s+K-1}}, x_{s+K:t}, \theta \mid y_{s_0:t})}{M^{(s+K-1)-s_0+1}}$$

$$\times \prod_{\substack{m=1 \\ m \neq k_{s+K-1}}}^{M} q_{s+K-1,\theta}(x_{s+K-1}^m \mid x_{s+K}, y_{s+K-1}) \times \prod_{j=s_0}^{s+K-2} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(x_j^m \mid x_{j+1}^{a_{j+1}^m}, y_j)$$

$$= \pi(x_{s+K:t}, \theta \mid y_{s+K:t}) \times \prod_{m=1}^{M} q_{s+K-1,\theta}(x_{s+K-1}^m \mid x_{s+K}, y_{s+K-1})$$

$$\times \prod_{j=s_0}^{s+K-2} \prod_{m=1}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(x_j^m \mid x_{j+1}^{a_{j+1}^m}, y_j) \times V_{s_0,\theta}^{k_{s_0}} \times \prod_{j=s_0}^{s+K-1} \frac{\hat{p}(y_j \mid x_{s+K}, y_{j+1:t}, \theta)}{p(y_j \mid y_{j+1:t})}$$

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

$$\hat{p}(y_j \mid x_{s+K}, y_{j+1:t}, \theta) = \frac{1}{M} \sum_{m=1}^{M} v_{j,\theta}(x_j^m, x_{j+1}^{a_{j+1}^m}). \tag{57}$$

Proof of Lemma A.2.

Since

$$\pi(x_{s_0}^{k_{s_0}}, \ldots, x_{s+K-1}^{k_{s+K-1}}, x_{s+K:t}, \theta \mid y_{s_0:t})$$

$$= \frac{\pi(x_{s+K:t}, \theta \mid y_{s+K:t})}{p(y_{s_0:s+K-1} \mid y_{s+K:t})} \prod_{j=s_0}^{s+K-1} p(x_j^{k_j} \mid x_{j+1}^{k_{j+1}}, \theta) g_\theta(y_j \mid x_j^{k_j}, x_{j+1}^{k_{j+1}})$$

$$= \frac{\pi(x_{s+K:t}, \theta \mid y_{s+K:t})}{p(y_{s_0:s+K-1} \mid y_{s+K:t})} \prod_{j=s_0}^{s+K-1} v_{j,\theta}(x_j^{k_j}, x_{j+1}^{k_{j+1}}) \times \prod_{j=s_0}^{s+K-1} q_{j,\theta}(x_j^{k_j} \mid x_{j+1}^{k_{j+1}}, y_j),$$

we obtain

$$\frac{\pi(x_{s_0}^{k_{s_0}}, \ldots, x_{s+K-1}^{k_{s+K-1}}, x_{s+K:t}, \theta \mid y_{s_0:t})}{M^{(s+K-1)-s_0+1}}$$

$$\times \prod_{\substack{m=1 \\ m \neq k_{s+K-1}}}^{M} q_{s+K-1,\theta}(x_{s+K-1}^m \mid x_{s+K}, y_{s+K-1}) \times \prod_{j=s_0}^{s+K-2} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(x_j^m \mid x_{j+1}^{a_{j+1}^m}, y_j)$$

4

$$
= \frac{\pi(x_{s+K:t}, \theta \mid y_{s+K:t})}{M^{(s+K-1)-s_0+1} p(y_{s_0:s+K-1} \mid y_{s+K:t})} \times \prod_{j=s_0}^{s+K-1} \prod_{m=1}^{M} q_{j,\theta}(x_j^m \mid x_{j+1}^{a_{j+1}^m}, y_j)
$$

$$
\times \prod_{j=s_0-1}^{s+K-2} v_{j+1,\theta}(x_{j+1}^{k_{j+1}}, x_{j+2}^{k_{j+2}}) \times \prod_{j=s_0}^{s+K-2} \prod_{\substack{m=1 \\ m \neq k_j}}^{M} V_{j+1,\theta}^{a_{j+1}^m}
$$

$$
= \frac{\pi(x_{s+K:t}, \theta \mid y_{s+K:t})}{p(y_{s_0:s+K-1} \mid y_{s+K:t})} \times \prod_{j=s_0}^{s+K-1} \prod_{m=1}^{M} q_{j,\theta}(x_j^m \mid x_{j+1}^{a_{j+1}^m}, y_j)
$$

$$
\times \prod_{j=s_0}^{s+K-1} \left\{ \frac{1}{M} \sum_{i=1}^{M} v_{j,\theta}(x_j^i, x_{j+1}^{a_{j+1}^i}) \right\} \times \prod_{j=s_0}^{s+K-2} \prod_{m=1}^{M} V_{j+1,\theta}^{a_{j+1}^m} \times V_{s_0,\theta}^{k_{s_0}}
$$

$$
= \frac{\pi(x_{s+K:t}, \theta \mid y_{s+K:t})}{\prod_{j=s_0}^{s+K-1} p(y_j \mid y_{j+1:t})} \times \prod_{j=s_0}^{s+K-1} \prod_{m=1}^{M} q_{j,\theta}(x_j^m \mid x_{j+1}^{a_{j+1}^m}, y_j)
$$

$$
\times \prod_{j=s_0}^{s+K-2} \prod_{m=1}^{M} V_{j+1,\theta}^{a_{j+1}^m} \times V_{s_0,\theta}^{k_{s_0}} \times \prod_{j=s_0}^{s+K-1} \hat{p}(y_j \mid x_{s+K}, y_{j+1:t}, \theta)
$$

$\square$

Proof of Proposition 5.3.

By applying Lemma A.2 with $s_0 = s - 1$ to the first three terms of the target distribution in (49), we have

$$
\check{\pi}(x_{s-1:s+K-1}^{1:M}, x_{s+K:t}, a_{s:s+K-1}^{1:M}, k_{s-1}, k_s^*, \theta \mid y_{s-1:t})
$$

$$
= \pi(x_{s+K:t}, \theta \mid y_{s+K:t}) \times \prod_{m=1}^{M} q_{s+K-1,\theta}(x_{s+K-1}^m \mid x_{s+K}, y_{s+K-1:t})
$$

$$
\times \prod_{j=s-1}^{s+K-2} \prod_{m=1}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(x_j^m \mid x_{j+1}^{a_{j+1}^m}, y_j) \times V_{s-1,\theta}^{k_{s-1}} \times \prod_{j=s-1}^{s+K-1} \frac{\hat{p}(y_j \mid x_{s+K}, y_{j+1:t}, \theta)}{p(y_j \mid y_{j+1:t})}
$$

$$
\times V_s^{k_s^*} \times \frac{p(y_{s-1} \mid y_{s:t})}{\hat{p}(y_{s-1} \mid x_{s+K}, y_{s:t}, \theta)}
$$

$$
= \pi(x_{s+K:t}, \theta \mid y_{s+K:t}) \times \prod_{m=1}^{M} q_{s+K-1,\theta}(x_{s+K-1}^m \mid x_{s+K}, y_{s+K-1:t})
$$

$$
\times \prod_{j=s}^{s+K-2} \prod_{m=1}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(x_j^m \mid x_{j+1}^{a_{j+1}^m}, y_j) \times V_s^{k_s^*} \times \prod_{j=s}^{s+K-1} \frac{\hat{p}(y_j \mid x_{s+K}, y_{j+1:t}, \theta)}{p(y_j \mid y_{j+1:t})}
$$

$$
\times \prod_{m=1}^{M} V_{s,\theta}^{a_s^m} q_{s-1,\theta}(x_{s-1}^m \mid x_s^{a_s^m}, y_{s-1}) \times V_{s-1,\theta}^{k_{s-1}}
$$

$$
= \quad \frac{\pi(x_s^{k_s^*}, \dots, x_{s+K-1}^{k_{s+K-1}^*}, x_{s+K:t}, \theta \mid y_{s:t})}{M^K} \times \prod_{\substack{m=1 \\ m \neq k_{s+K-1}^*}}^{M} q_{s+K-1,\theta}(x_{s+K-1}^m \mid x_{s+K}, y_{s+K-1:t})
$$

$$
\times \prod_{j=s}^{s+K-2} \prod_{\substack{m=1 \\ m \neq k_j^*}}^{M} V_{j+1,\theta}^{a_{j+1}^m} q_{j,\theta}(x_j^m \mid x_{j+1}^{a_{j+1}^m}, y_j) \times \prod_{m=1}^{M} V_{s,\theta}^{a_s^m} q_{s-1,\theta}(x_{s-1}^m \mid x_s^{a_s^m}, y_{s-1}) \times V_{s-1,\theta}^{k_{s-1}}
$$

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

A.4 Proof of Proposition 5.4

Proof of Proposition 5.4.

We first define the probability density function

$$
\bar{\psi}_{\theta,0}(x_{s-1:s+K-1}^{1:M}, a_{s:s+K-1}^{1:M}, k_{s-1}, k_s^* \mid x_{s+K}, y_{s-1:s+K-1})
$$

$$
\equiv \quad \frac{\pi(x_{s-1:s+K-1} \mid x_{s+K}, y_{s-1:t}, \theta)}{M^{K+1}}
$$

$$
\times \quad \bar{\psi}_\theta(x_{s-1}^{-k_{s-1}}, \dots, x_{s+K-1}^{-k_{s+K-1}}, a_s^{-k_{s-1}}, \dots, a_{s+K-1}^{-k_{s+K-2}}, k_s^* \mid x_{s-1:s+K}, a_{s-1}^{k_{s-2}}, \dots, a_{s+K-1}^{k_{s+K-2}}, y_{s-1:s+K-1}),
$$

and note that

$$
\pi(x_{s-1:s+K-1} \mid x_{s+K}, y_{s-1:t}, \theta) = \frac{\pi(x_{s-1:s+K-1}, x_{s+K:t}, \theta \mid y_{s-1:t})}{\pi(x_{s+K:t}, \theta \mid y_{s-1:t})}. \tag{58}
$$

Since

$$
\frac{1}{\hat{p}(y_{s-1} \mid x_{s+K}, y_{s:t}, \theta)} \bar{\psi}_{\theta,0}(x_{s-1:s+K-1}^{1:M}, a_{s:s+K-1}^{1:M}, k_{s-1}, k_s^* \mid x_{s+K}, y_{s-1:s+K-1})
$$

$$
= \quad \check{\pi}(x_{s-1:s+K-1}^{1:M}, x_{s+K:t}, a_{s:s+K-1}^{1:M}, k_{s-1}, k_s^*, \theta \mid y_{s-1:t}) \frac{1}{p(y_{s-1} \mid y_{s:t}) \pi(x_{s+K:t}, \theta \mid y_{s-1:t})},
$$

where we used the definition of $\check{\pi}$ in (49), we obtain

$$
E_{\bar{\psi}_{\theta,0}}\left[\hat{p}(y_{s-1} \mid x_{s+K}, y_{s:t}, \theta)^{-1} \mid x_{s+K}, y_{s-1:t}, \theta \right]
$$

$$
= \quad \int \frac{1}{\hat{p}(y_{s-1} \mid x_{s+K}, y_{s:t}, \theta)} \bar{\psi}_{\theta,0}(x_{s-1:s+K-1}^{1:M}, a_{s:s+K-1}^{1:M}, k_{s-1}, k_s^* \mid x_{s+K}, y_{s-1:s+K-1})
$$

$$
dx_{s-1:s+K-1}^{1:M} da_{s:s+K-1}^{1:M} dk_{s-1} dk_s^*
$$

$$
= \quad \int \check{\pi}(x_{s-1:s+K-1}^{1:M}, x_{s+K:t}, a_{s:s+K-1}^{1:M}, k_{s-1}, k_s^*, \theta \mid y_{s-1:t}) dx_{s-1:s+K-1}^{1:M} da_{s:s+K-1}^{1:M} dk_{s-1} dk_s^*
$$

$$
\times \frac{1}{p(y_{s-1} \mid y_{s:t}) \pi(x_{s+K:t}, \theta \mid y_{s-1:t})}
$$

$$
= \quad \frac{\pi(x_{s+K:t}, \theta \mid y_{s:t})}{p(y_{s-1} \mid y_{s:t}) \pi(x_{s+K:t}, \theta \mid y_{s-1:t})} = \frac{1}{p(y_{s-1} \mid x_{s+K:t}, y_{s:t}, \theta)} = \frac{1}{p(y_{s-1} \mid x_{s+K}, y_{s:t}, \theta)}
$$

where we use Proposition 5.3 in the third equality. Further,

$$
E[p(y_{s-1} \mid x_{s+K}, y_{s:t}, \theta)^{-1} \mid y_{s-1:t}] = \int \frac{\pi(x_{s+K}, \theta \mid y_{s-1:t})}{p(y_{s-1} \mid x_{s+K}, y_{s:t}, \theta)} dx_{s+K} d\theta
$$

$$
= \int \frac{\pi(x_{s+K}, \theta \mid y_{s:t})}{p(y_{s-1} \mid y_{s:t})} dx_{s+K} d\theta = p(y_{s-1} \mid y_{s:t})^{-1}.
$$

$\square$

Proof of Proposition 5.5

Proof of Proposition 5.5.

Consider the joint marginal density of (45):

$$
\hat{\pi}(x_{s-1:t-K-1}, x_{t-K:t_0}^{1:M}, a_{t-K:t_0-1}^{1:M}, x_{t_0+1}^{k_{t_0+1}^*}, \ldots, x_t^{k_t^*}, k_{t_0:t}^*, \theta \mid y_{s-1:t})
$$

$$
= \frac{\pi(x_{s-1:t-K-1}, x_{t-K}^{k_{t-K}^*}, \ldots, x_t^{k_t^*}, \theta \mid y_{s-1:t})}{M^{K+1}} \times \prod_{\substack{m=1 \\ m \neq k_{t-K}^*}}^{M} q_{t-K,\theta}(x_{t-K}^m \mid x_{t-K-1}, y_{t-K})
$$

$$
\times \prod_{j=t-K+1}^{t_0} \prod_{\substack{m=1 \\ m \neq k_j^*}}^{M} V_{j-1}^{a_j^m} q_{j,\theta}(x_j^m \mid x_{j-1}^{a_j^m}, y_j), \tag{59}
$$

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

$$
\hat{\pi}(x_{s-1:t-K-1}, x_{t-K}^{1:M}, x_{t-K+1}^{k_{t-K+1}^*}, \ldots, x_t^{k_t^*}, k_{t-K:t}^*, \theta \mid y_{s-1:t})
$$

$$
= \frac{\pi(x_{s-1:t-K-1}, x_{t-K}^{k_{t-K}^*}, \ldots, x_t^{k_t^*}, \theta \mid y_{s-1:t})}{M^{K+1}} \times \prod_{\substack{m=1 \\ m \neq k_{t-K}^*}}^{M} q_{t-K,\theta}(x_{t-K}^m \mid x_{t-K-1}, y_{t-K}). \tag{60}
$$

Then we obtain

$$
\hat{\pi}(k_{t_0}^* \mid x_{s-1:t-K-1}^{1:M}, x_{t-K:t_0}^{1:M}, a_{t-K:t_0-1}^{1:M}, x_{t_0+1}^{k_{t_0+1}^*}, \ldots, x_t^{k_t^*}, k_{t_0+1:t}^*, y_{s-1:t}, \theta)
$$

$$
\propto \hat{\pi}(x_{s-1:t-K-1}^{1:M}, x_{t-K:t_0}^{1:M}, a_{t-K:t_0-1}^{1:M}, x_{t_0+1}^{k_{t_0+1}^*}, \ldots, x_t^{k_t^*}, k_{t_0:t}^*, \theta \mid y_{s-1:t})
$$

$$
\propto \frac{\pi(x_{s-1:t-K-1}, x_{t-K}^{k_{t-K}^*}, \ldots, x_{t_0}^{k_{t_0}^*}, \theta \mid y_{s-1:t_0})}{M^{t_0-(t-K)+1}} \times \prod_{\substack{m=1 \\ m \neq k_{t-K}^*}}^{M} q_{t-K,\theta}(x_{t-K}^m \mid x_{t-K-1}, y_{t-K})
$$

$$
\times \prod_{j=t-K+1}^{t_0} \prod_{\substack{m=1 \\ m \neq k_j^*}}^{M} V_{j-1}^{a_j^m} q_{j,\theta}(x_j^m \mid x_{j-1}^{a_j^m}, y_j) \times \prod_{j=t_0+1}^{t} f_\theta(x_j^{k_j^*} \mid x_{j-1}^{k_{j-1}^*}, y_{j-1}) g_\theta(y_j \mid x_j^{k_j^*})
$$

$$
= \pi(x_{s-1:t-K-1}, \theta \mid y_{s-1:t-K-1}) \times \prod_{m=1}^{M} q_{t-K,\theta}(x_{t-K}^m \mid x_{t-K-1}, y_{t-K})
$$

$$
\times \prod_{j=t-K+1}^{t_0} \prod_{m=1}^{M} V_{j-1,\theta}^{a_j^m} q_{j,\theta}(x_j^m \mid x_{j-1}^{a_j^m}, y_j) \times V_{t_0,\theta}^{k_{t_0}^*} \times \prod_{j=t-K}^{t_0} \frac{\hat{p}(y_j \mid x_{t-K-1}, y_{s-1:j-1}, \theta)}{p(y_j \mid y_{s-1:j-1})},
$$

$$
\times \prod_{j=t_0+1}^{t} f_\theta(x_j^{k_j^*} \mid x_{j-1}^{k_{j-1}^*}, y_{j-1}) g_\theta(y_j \mid x_j^{k_j^*})
$$

$$
\propto V_{t_0,\theta}^{k_{t_0}^*} \times f_\theta(x_{t_0+1}^{k_{t_0+1}^*} \mid x_{t_0}^{k_{t_0}^*}, y_{t_0}), \tag{61}
$$

where we use Lemma A.1 at the equality. $\square$

7

B   Sequential MCMC estimation without rolling the window

We first give the initializing algorithm which is obtained by skipping the discarding step (Step 2) in the particle rolling algorithm. Next, we describe how to estimate the marginal likelihood.

B.1   Algorithm

(1) At time $j = 1$, sample (x_1^n, θ^n) from $\pi(x_1, \theta \mid y_1)$ for $n = 1, \ldots, N$.

    1. Sample $\theta^n \sim p(\theta)$, and $x_1^{n,m} \sim q_{1,\theta^n}(\cdot \mid y_1)$ for each $m \in \{1, \ldots, M\}$.

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

$$V_{1,\theta^n}^{n,m} \;=\; \frac{v_{1,\theta^n}(x_1^{n,m})}{\sum_{i=1}^M v_{1,\theta^n}(x_1^{n,i})}, \quad v_{1,\theta^n}(x_1^{n,m}) = \frac{\mu_{\theta^n}(x_1^{n,m})g_{\theta^n}(y_1 \mid x_1^{n,m})}{q_{1,\theta^n}(x_1^{n,m} \mid y_1)}. \tag{62}$$

    3. Set $x_1^n = x_1^{n,k_1}$ and store (x_1^n, θ^n) with its importance weight

$$W_1^n \;\propto\; \hat{p}(y_1 \mid \theta^n), \quad \hat{p}(y_1 \mid \theta^n) = \sum_{m=1}^M v_{1,\theta^n}(x_1^{n,m}). \tag{63}$$

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

$$W_j^n \;\propto\; \hat{p}(y_j \mid x_{j-K-1}^n, y_{1:j-1}, \theta^n) \times W_{j-1}^n, \tag{64}$$

$$\hat{p}(y_j \mid x_{j-K-1}^n, y_{1:j-1}, \theta^n) = \frac{1}{M} \sum_{m=1}^M v_{j,\theta^n}(x_{j-1}^{n,a_{j-1}^{n,m}}, x_j^{n,m}). \tag{65}$$

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

Especially when j is small and the dimension of $x_{1:j}$ 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

$$p(y_{s:t}) \;=\; \int p(y_{s:t} \mid x_{s:t}, \theta)p(x_{s:t} \mid \theta)p(\theta)dx_{s:t}d\theta, \tag{66}$$

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

$$p(y_{s:t}) \;=\; \frac{p(y_t \mid y_{s-1:t-1})}{p(y_{s-1} \mid y_{s:t})}p(y_{s-1:t-1}), \tag{67}$$

<div align="center">8</div>

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

$$\hat{p}(y_{s:t}) = \frac{\hat{p}(y_t \mid y_{s-1:t-1})}{\hat{p}(y_{s-1} \mid y_{s:t})}\hat{p}(y_{s-1:t-1}), \tag{68}$$

where

$$\hat{p}(y_t \mid y_{s-1:t-1}) = \sum_{n=1}^{N} W_{[s-1,t-1]}^{n}\hat{p}(y_t \mid x_{t-K-1}^{n}, y_{s-1:t-1}, \theta^n), \tag{69}$$

$$\hat{p}(y_{s-1} \mid y_{s:t}) = \sum_{n=1}^{N} W_{[s-1,t]}^{n}\hat{p}(y_{s-1} \mid x_{s+K}^{n}, y_{s:t}, \theta^n), \tag{70}$$

using (28), (29), (34) and (35). The initial estimate $\hat{p}(y_{1:L+1})$, $L = t - s$ is given by

$$\hat{p}(y_{1:L+1}) = \hat{p}(y_1) \prod_{j=2}^{L+1} \hat{p}(y_j \mid y_{1:j-1}), \tag{71}$$

where we use (63), (64) and (65) to obtain

$$\hat{p}(y_1) = \sum_{n=1}^{N}\hat{p}(y_1 \mid \theta^n), \quad \hat{p}(y_j \mid y_{1:j-1}) = \sum_{n=1}^{N} W_{j-1}^{n}\hat{p}(y_j \mid x_{j-K-1}^{n}, y_{1:j-1}, \theta^n). \tag{72}$$

C   Additional comparison in the RSV model

We compare the computation time and the ESS of the practical double-block sampler with those of the MCMC and the particle MCMC. For the initial sample period (using $y_{1:1988}$), the MCMC sampling is implemented with 10,000 iteration (2,000 MCMC samples in the burn-in period are discarded). Table 5 shows the computation times[8] and ESSs[9] for three methods.

---

[8]The total computation time for the MCMC and the particle MCMC to complete the rolling-window estimation is obtained by multiplying the computation time for the initial sample period by 2261. Thus we obtain $1,293 \times 2,261 = 2,923,473$ and $3,189 \times 2,261 = 7,210,329$ respectively.

[9]The ESS is computed as the average of the ESSs during the rolling estimations for our double-block sampler, while that for each parameter is computed as the MCMC sample size (10,000) divided by the inefficiency factor (defined as $1 + 2\sum_{s=1}^{\infty} \rho_s$, where ρ_s is the MCMC sample autocorrelation at lag s).

Table 5: Computation times for the double-block sampler, MCMC and the particle MCMC.

	Time (seconds)	Param.	ESS
Double-block sampler	**142,709**	-	**729**
MCMC	$\mathbf{2,923,473}$	μ	**35**
		ϕ	1764
		σ_η^2	189
		ξ	3942
		σ_u^2	636
		ρ	257
Particle MCMC	**7,210,329**	μ	**40**
		ϕ	2184
		σ_η^2	221
		ξ	4878
		σ_u^2	656
		ρ	189

The recursive estimation using the standard MCMC or the particle MCMC takes 20-50 times longer than our proposed method. If we take account of the ESS, it would take 400-900 times longer. These results show that the computation time for our proposed method is much smaller compared with recursive estimations using the standard MCMC or the particle MCMC.