CENTER FOR ADVANCED RESEARCH IN FINANCE

CARF Working Paper

CARF-F-104

Block Sampler and Posterior Mode Estimation for A Nonlinear and Non-Gaussian State-Space Model with Correlated Errors

Yasuhiro Omori
University of Tokyo
Toshiaki Watanabe
Hitotsubashi University

August 2007

CARF is presently supported by Bank of Tokyo-Mitsubishi UFJ, Ltd., Dai-ichi Mutual Life Insurance Company, Meiji Yasuda Life Insurance Company, Mizuho Financial Group, Inc., Nippon Life Insurance Company, Nomura Holdings, Inc. and Sumitomo Mitsui Banking Corporation (in alphabetical order). This financial support enables us to issue CARF Working Papers.

Block sampler and posterior mode estimation for a nonlinear and non-Gaussian state-space model with correlated errors

May 2003: Revised in August 2007

Yasuhiro Omori
Faculty of Economics, University of Tokyo
omori@e.u-tokyo.ac.jp

Toshiaki Watanabe
Institute of Economic Research, Hitotsubashi University
Watanabe@ier.hit-u.ac.jp

Abstract

This article introduces a new efficient simulation smoother and disturbance smoother for general state-space models where there exists a correlation between error terms of the measurement and state equations. The state vector is divided into several blocks where each block consists of many state variables. For each block, corresponding disturbances are sampled simultaneously from their conditional posterior distribution. The algorithm is based on the multivariate normal approximation of the conditional posterior density and exploits a conventional simulation smoother for a linear and Gaussian state space model. The performance of our method is illustrated using two examples (1) stochastic volatility models with leverage effects and (2) stochastic volatility models with leverage effects and state-dependent variances. The popular single move sampler which samples a state variable at a time is also conducted for comparison in the first example. It is shown that our proposed sampler produces considerable improvement in the mixing property of the Markov chain Monte Carlo chain.

Key words: Bayesian analysis; Disturbance smoother; Kalman filter; Leverage effects; Markov chain Monte Carlo; Metropolis-Hastings algorithm; Simulation smoother; State space model; Stochastic volatility.

1 Introduction

A state-space approach for time series provides various dynamic models for a wide range of applications such as in financial time series. Using the flexible structure of latent variables, the state-space model can express a variety of statistical models. Recently, for example, stochastic volatility models with leverage effects (Jacquier *et al.* (2004), Yu (2005), Omori *et al.* (2006)) and nonlinear state-space models with state-dependent variances (Stroud *et al.* (2003)) and a Cox-Ingersoll-Ross model with a nonlinear state equation (Frühwirth-Schnatter and Geyer (1998), Sanford and Martin (2003), Watanabe *et al.* (2006)) have come to receive widespread attention. However, because of many latent state variables, it has been difficult to obtain estimates of parameters for these models until recent development of Markov chain Monte Carlo (MCMC) method using Bayesian approach.

For a linear and Gaussian state-space model, the Kalman smoother is a well-known recursive procedure for computing smoothed estimates of states (see e.g. Anderson and Moore (1979)), and several efficient smoothers have been proposed for states (de Jong (1988)(1989), Kohn and Ansley (1989)) and disturbances (Koopman (1993)). To perform Bayesian analysis of unknown parameters, we need to sample from a highly multivariate posterior distribution of state variables. Simulation smoothers have been considered to sample state variables in a single block using a forward-filtering-backward sampling algorithm (Carter and Kohn (1994), Frühwirth-Schnatter (1994)). Further, more efficient simulation smoothers have been recently developed by sampling disturbances and subsequently sampling states (de Jong and Shephard (1995), Durbin and Koopman (2002)).

For nonlinear or non-Gaussian state-space models, several smoothing procedures have been introduced. Examples are Sage and Melsa (1971), Anderson and Moore (1979) for a nonlinear and Gaussian model and Fahrmeir (1992), Fahrmeir and Wagenpfeil (1997) for a dynamic generalized linear model or an exponential family state-space model. Monte Carlo filters and smoothers are also developed to approximate posterior distributions of states given system parameters using a few discrete points or particles where an observation vector and a state vector are assumed to be conditionally independent (Kitagawa (1996), Kitagawa (1998), Hürzeler and Künsch (1998), Tanizaki and Mariano (1998), Tanizaki (2001)). For such a smoother, an appropriate proposal density needs to be taken carefully to have a good Monte Carlo approximation by particles. The predictive density of the state variable at time t given past observations would usually be a candidate for such a proposal density, but its approximation becomes poor when its posterior density becomes quite different from the predictive density.

On the other hand, Carlin *et al.* (1992) used Gibbs sampler with rejection sampling to sample states for nonlinear and Gaussian models. Their method is a single move sampler that generates a single state at a time given the rest of the states and other parameters. It is usually easy to construct such a sampler, but the obtained samples are known to be highly autocorrelated. This implies we need to generate a huge number of samples to conduct a statistical inference and hence the sampler is inefficient.

To reduce sample autocorrelations effectively, block samplers (also called multi-move samplers) which generate a block of state variables have been recently proposed (Shephard and Pitt (1997), Gamerman (1998), Durbin and Koopman (2000), Watanabe and Omori (2004)). However, these samplers assume that a state equation is linear and that an observation vector and a state vector are conditionally independent. Thus they cannot be applied to an important class of models with a correlation between errors of measurement and state equations, a variance of state variable depending on the past state variable, and a nonlinear state equation. The exception is a mixture sampler by Omori *et al.* (2006) for a stochastic volatility with leverage effects, which extended Kim *et al.* (1998). It approximates a non-Gaussian model using a mixture of bivariate normal distributions given the sign of observed dependent variables. The mixture sampler is fast and highly efficient, but its use is limited to a certain class of the stochastic volatility model they considered. Stroud *et al.* (2003) considered a block sampler for models with state-dependent variances (but without leverage effects) using an auxiliary mixture model to generate a state proposal for Metropolis-Hastings algorithms.

In this article, we propose a new efficient smoother for a general classes of state-space models. First, we derive a recursive algorithm to find a posterior mode of the state vector for a non-Gaussian measurement model with a linear state equation using Taylor expansion of the logarithm of the conditional posterior density for the disturbances. Second we define an approximating linear and Gaussian measurement equation based on the obtained posterior mode. For a model with a nonlinear state equation, we construct an auxiliary linear state equation to derive an approximating linear and Gaussian state space model. Then we generate a candidate for a state variable in Metropolis-Hastings algorithm using this approximating linear and Gaussian state space model.

The organization of the paper is as follows. In Section 2, we introduce a general state-space model with examples. Section 3 describes a simulation smoother and a disturbance smoother for these models. In Section 4, we illustrate our method using simulated data and stock returns data. Section 5 concludes the paper.

2 General state-space model

We consider a general state-space model

$$y_t = h_t(\alpha_t, \epsilon_t), \quad \epsilon_t \sim p_{\epsilon|\eta}(\epsilon_t), \quad t = 1, \ldots, n, \tag{1}$$

$$\alpha_{t+1} = g_t(\alpha_t, \eta_t), \quad \eta_t \sim p_\eta(\eta_t), \quad t = 0, 1, \ldots, n, \tag{2}$$

where we assume that $\alpha_0 = 0$. The dependent variable at time t is y_t and a measurement equation is (1). The state variable at time $t+1$ is α_{t+1} and a state equation is given by (2). The error terms of the measurement and state equations are ϵ_t and η_t respectively, and may be correlated. The functions $g_t, h_t, p_\eta, p_{\epsilon|\eta}$ are assumed to be twice continuously differentiable. When α_{t+1} is a linear function of (α_t, η_t), the state equation (2) reduces to

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \quad \eta_t \sim p_\eta(\eta_t),$$

where T_t, R_t are constant matrices.

In this paper, we consider two examples (1) stochastic volatility models with leverage effects and (2) stochastic volatility models with leverage effects and state-dependent variances, to illustrate and investigate the performance of our method.

Example 1. Stochastic volatility model with leverage effects. We first consider a model given by

$$\begin{aligned} y_t &= \epsilon_t \sigma_\epsilon \exp(\alpha_t/2), \quad t = 1, \ldots, n, \\ \alpha_{t+1} &= \phi \alpha_t + \eta_t \sigma_\eta, \quad t = 1, \ldots, n-1, \\ \alpha_1 &= \eta_0 \sigma_0 \sim N(0, \sigma_\eta^2/(1-\phi^2)), \end{aligned} \tag{3}$$

where $|\phi| < 1$ and

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

The $\sigma_\epsilon \exp(\alpha_t/2)$ stands for the volatility of the response, y_t, and $\rho, \sigma_\epsilon, \sigma_\eta, \phi$ are parameters. The state equation is linear and Gaussian, while the measurement equation is nonlinear. A correlation between errors is considered to explain leverage effects. The stochastic volatility model without leverage effects ($\rho = 0$) has been widely used to explain time varying variances of the response in the analysis of financial time series data such as stock returns and foreign exchange rate data. However, it is well known that the fall of the stock return is followed by the high volatility and this is called a "leverage effect" (Black (1976), Nelson (1991)). Thus we expect a negative correlation, $\rho < 0$, between ϵ_t and η_t rather than $\rho = 0$. Jacquier *et al.* (2004) proposed a single move sampler for a stochastic volatility model with leverage effects using MCMC method. Since the single move sampler samples one α_t at a time given other state variables, the MCMC samples are highly autocorrelated, and we need to iterate MCMC runs a huge number of times to obtain a stable and reliable estimation results. Omori *et al.* (2006) proposed a fast and highly efficient mixture sampler to estimate such models and showed its high performance in estimation efficiency compared with a single move sampler. But their method is only applicable to the class of models (3). In this article, we give an efficient block sampler for more general state-space models which includes this stochastic volatility model .

Example 2. Stochastic volatility model with leverage effects and state-dependent variances. Stroud *et al.* (2003) considered state-dependent variance models (but without leverage effects) to explain such fat-tailed errors using a square-root stochastic volatility model with jumps in the analysis of Hong Kong interest rates. We may instead consider a simple extension of (3) such as

$$\begin{aligned} y_t &= \epsilon_t \sigma_\epsilon \exp(\alpha_t/2), \quad t = 1, \dots, n, \\ \alpha_{t+1} &= \phi\alpha_t + \eta_t\sigma_\eta \left\{ 1 + \frac{1}{1+\exp(-\alpha_t)} \right\}, \quad t = 1, \dots, n-1, \\ \alpha_1 &= \sigma_0\eta_0 \sim N(0, \sigma_0^2), \quad (\sigma_0^2\text{: known}), \end{aligned} \tag{4}$$

where $|\phi| < 1$. The variance of the error in the state equation depends on the level of the state variable. Thus the conditional variance tends to be larger for the large positive value of the state variable, α_t, while it becomes small for the negative value. We use this model to illustrate a state equation which is a nonlinear function of α_t and η_t.

We focus on above two examples to illustrate our method in this article, but there are still other important examples such as the Cox-Ingersoll -Ross (CIR) model. The CIR model is widely used to describe a term structure models of interest rates and in its state space formulation, it has a nonlinear and non-Gaussian state equation. The single move samplers are proposed to estimate of econometric multi-factor CIR model (Frühwirth-Schnatter and Geyer (1998)) and Affine term structure model (Sanford and Martin (2003)). Our proposed method would provide a block sampler for such models (Watanabe *et al.* (2006)).

3 Block sampler and posterior mode estimation

This section proposes a new efficient block sampler and a posterior mode estimation method for a conditional posterior density of state variables. Assuming a non-Gaussian measurement equation, we first consider a linear Gaussian state equation model and then discuss a nonlinear Gaussian state equation model. Finally, we describe a procedure for a model with a nonlinear and non-Gaussian state equation.

3.1 Linear and Gaussian state equation

Consider the following state space model with linear Gaussian state equation given by

$$y_t = h_t(\alpha_t, \epsilon_t), \quad \epsilon_t \sim p_{\epsilon|\eta}(\epsilon_t), \quad t = 1, \ldots, n, \tag{5}$$

$$\alpha_{t+1} = T_t \alpha_t + R_t \eta_t, \quad \eta_t \sim N(0, I), \quad t = 0, 1, \ldots, n, \tag{6}$$

where $\alpha_0 = 0$ and $W_t \equiv R_t R_t'$ is assumed to be nonsingular for simplicity, but we may drop this assumption as we discuss at the end of this subsection. Suppose that we sample from the conditional posterior distribution of $\alpha = (\alpha_{s+1}, \ldots, \alpha_{s+m})$ given $\alpha_s, \alpha_{s+m+1}, y_s, \ldots, y_{s+m}$ where $m \geq 2$. We construct a proposal density based on disturbance terms $\eta = (\eta_s', \ldots, \eta_{s+m-1}')'$.

Normal approximation of the posterior density of η. We first expand the logarithm of the conditional posterior density of η around the mode $\hat{\eta}$ to obtain a normal approximation (given α_s, α_{s+m+1}). Let l_s denote the logarithm of conditional likelihood of y_s given α and

$$L = \begin{cases} \sum_{t=s}^{s+m} l_t - \frac{1}{2}(\alpha_{s+m+1} - T_{s+m}\alpha_{s+m})' W_{s+m}^{-1} (\alpha_{s+m+1} - T_{s+m}\alpha_{s+m}), \\ \qquad \text{if} \quad s+m < n, \\ \sum_{t=s}^{n} l_t, \qquad \text{if} \quad s+m = n. \end{cases} \tag{7}$$

Further define

$$d = (d_{s+1}', \ldots, d_{s+m}')', \quad d_t = \frac{\partial L}{\partial \alpha_t}, \quad t = s+1, \ldots, s+m, \tag{8}$$

$$Q = -E\left[\frac{\partial^2 L}{\partial \alpha \partial \alpha'}\right] = \begin{pmatrix} A_{s+1} & B_{s+2}' & O & \cdots & O \\ B_{s+2} & A_{s+2} & B_{s+3}' & \cdots & O \\ O & B_{s+3} & A_{s+3} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & B_{s+m}' \\ O & \cdots & O & B_{s+m} & A_{s+m} \end{pmatrix},$$

$$A_t = -E\left[\frac{\partial^2 L}{\partial \alpha_t \partial \alpha_t'}\right], \quad t = s+1, \ldots, s+m, \tag{9}$$

$$B_t = -E\left[\frac{\partial^2 L}{\partial \alpha_t \partial \alpha_{t-1}'}\right], \quad t = s+2, \ldots, s+m, \quad B_{s+1} = O. \tag{10}$$

Then the approximating normal density f^* is as follows (see Appendix A1).

$$
\begin{aligned}
&\log f(\eta_s, \ldots, \eta_{s+m-1}|\alpha_s, \alpha_{s+m+1}, y_s, \ldots, y_{s+m}) \\
\approx\ & \text{const} - \frac{1}{2}\sum_{t=s}^{s+m-1} \eta_t'\eta_t + \hat{L} + \left.\frac{\partial L}{\partial \eta'}\right|_{\eta=\hat{\eta}} (\eta - \hat{\eta}) + \frac{1}{2}(\eta-\hat{\eta})'\, E\left(\frac{\partial L^2}{\partial\eta\partial\eta'}\right)\Big|_{\eta=\hat{\eta}} (\eta-\hat{\eta}) \\
=\ & \text{const} - \frac{1}{2}\sum_{t=s}^{s+m-1} \eta_t'\eta_t + \hat{L} + \hat{d}'(\alpha-\hat{\alpha}) - \frac{1}{2}(\alpha-\hat{\alpha})'\hat{Q}(\alpha-\hat{\alpha}) \qquad (11) \\
=\ & \text{const} + \log f^*(\eta_s, \ldots, \eta_{s+m-1}|\alpha_s, \alpha_{s+m+1}, y_s, \ldots, y_{s+m}) \qquad (12)
\end{aligned}
$$

where $\hat{d}, \hat{L}, \hat{Q}$ denote d, L, Q evaluated at $\alpha = \hat{\alpha}$ (or, equivalently, at $\eta = \hat{\eta}$). The expectations are taken with respect to y_t's conditional on α_t's We use an information matrix for Q because we require that Q is everywhere strictly positive definite. However, other matrices such as a numerical negative Hessian matrix may be used to construct a positive definite matrix Q.

Posterior mode estimation. Next we describe how to find a mode, $\hat{\eta}$, of the conditional posterior density of η (see Appendix A2 for a derivation of Algorithm 1.1). We repeat the following algorithm until $\hat{\eta}$ converges to to the posterior mode.

Algorithm 1.1 (Disturbance smoother):

1. Initialize $\hat{\eta}$ and compute $\hat{\alpha}$ at $\eta = \hat{\eta}$ using (6) recursively.
2. Evaluate $\hat{d}_t$'s, $\hat{A}_t$'s, and $\hat{B}_t$'s using (8)–(10) at $\alpha = \hat{\alpha}$.
3. Compute the following D_t, J_t and b_t for $t = s+2, \ldots, s+m$ recursively.

$$
\begin{aligned}
D_t &= \hat{A}_t - \hat{B}_t D_{t-1}^{-1} \hat{B}_t', \quad D_{s+1} = \hat{A}_{s+1}, \\
J_t &= K_{t-1}^{-1\prime} \hat{B}_t, \quad J_{s+1} = O, \quad J_{s+m+1} = O, \\
b_t &= \hat{d}_t - J_t K_{t-1}^{-1} b_{t-1}, \qquad b_{s+1} = \hat{d}_{s+1},
\end{aligned}
$$

   where K_t denotes a Choleski decomposition of D_t such that $D_t = K_t K_t'$.

4. Define auxiliary variables $\hat{y}_t = \hat{\gamma}_t + D_t^{-1} b_t$ where

$$
\hat{\gamma}_t = \hat{\alpha}_t + K_t^{\prime -1} J_{t+1}' \hat{\alpha}_{t+1}, \quad t = s+1, \ldots, s+m,
$$

5. Consider the linear Gaussian state-space model given by

$$
\begin{aligned}
\hat{y}_t &= Z_t\alpha_t + G_t\xi_t, \quad t = s+1, \ldots, s+m, \qquad (13) \\
\alpha_{t+1} &= T_t\alpha_t + H_t\xi_t, \quad t = s, s+1, \ldots, s+m, \qquad (14) \\
&\xi_t = (\epsilon_t', \eta_t')' \sim N(0, I),
\end{aligned}
$$

where

$$Z_t = I + K_t'^{-1} J_{t+1}' T_t, \quad G_t = K_t'^{-1}[I, J_{t+1}' R_t], \quad H_t = [O, R_t].$$

Apply the Kalman filter and the disturbance smoother (e.g. de Jong and Shephard (1995)) to the linear Gaussian system (13) and (14) and obtain the posterior mode $\hat{\eta}$ and $\hat{\alpha}$.

6. Goto 2.

In the MCMC implementation, the current sample of η may be taken as an initial value of the $\hat{\eta}$. It can be shown that the posterior density of η_t^*'s obtained from (13) and (14) is the same as f^* in (12). Thus, applying the Kalman filter and the disturbance smoother to the linear Gaussian system (13) and (14), we first obtain a smoothed estimate of η_t and then substitute it recursively to the linear state equation (6) to obtain a smoothed estimate of α_t. Then we replace $\hat{\eta}_t, \hat{\alpha}_t$ by obtained smoothed estimates. By repeating the procedure until the smoothed estimates converge, we obtain the posterior mode of η_t, α_t. This is equivalent to the method of scoring to maximise the logarithm of the conditional posterior density.

Sampling from the posterior density of η. To sample η from the conditional posterior density, we propose a candidate sample from the density $q(\eta)$ which is proportional to $\min(f(\eta_y), cf^*(\eta_y))$ and conduct the Metropolis-Hastings algorithm (see e.g. Tierney (1994), Chib and Greenberg (1995)).

Algorithm 1.2 (Simulation smoother):

1. Given the current value η_x, find the mode $\hat{\eta}$ using Algorithm 1.1. Since it is enough to find an approximate value of the mode for a purpose of generating a candidate, we usually need to repeat Algorithm 1.1 only several times.

2. Proceed Step 2–4 of Algorithm 1.1 to obtain the approximate linear Gaussian system (13)–(14).

3. Propose a candidate η_y by sampling from $q(\eta_y) \propto \min(f(\eta_y), cf^*(\eta_y))$ using the Acceptance-Rejection algorithm where the logarithm of c can be constructed from a constant term and $\hat{L}$ in (11).

   (i) Generate $\eta_y \sim f^*$ using we may use a simulation smoother (e.g. de Jong and Shephard (1995), Durbin and Koopman (2002)) for the approximating linear Gaussian state-space model (13)–(14).

(ii) Accept η_y with probability

$$\frac{\min(f(\eta_y), cf^*(\eta_y))}{cf^*(\eta_y)}.$$

If it is rejected, go back to (i).

4. Conduct the MH algorithm using the candidate η_y. Given the current value η_x, we accept η_y with probability

$$\min\left\{1, \frac{f(\eta_y)\min(f(\eta_x), cf^*(\eta_x))}{f(\eta_x)\min(f(\eta_y), cf^*(\eta_y))}\right\}.$$

where a proposal density proportional to $\min(f(\eta_y), cf^*(\eta_y))$. If it is rejected, accept η_x as a sample.

Note that the independence between ϵ_t and η_t implies $B_t = O$ for all t, and equations (13) (14) reduce to

$$\begin{aligned}
\hat{y}_t &= \alpha_t + K_t^{\prime -1}\epsilon_t, \quad \epsilon_t \sim N(0, I), \quad t = s+1, \ldots, s+m, \\
\alpha_{t+1} &= T_t\alpha_t + R_t\eta_t, \quad \eta_t \sim N(0, I). \quad t = s, s+1, \ldots, s+m,
\end{aligned}$$

where $\hat{y}_t = \hat{\alpha}_t + \hat{A}_t^{-1}\hat{d}_t$ for $t = s+1, \ldots, s+m-1$ and $\hat{y}_{s+m} = \hat{\alpha}_{s+m}$.

We have derived an approximate linear Gaussian system and made use of a simulation smoother to generate error terms of the state equation to generate a sample of η. However, we could generate η directly without using such smoothers (see Appendix A3 for details).

Example. Stochastic volatility model with leverage effects. The state equation is linear and error term η_t follows normal distribution such that

$$\begin{aligned}
\alpha_{t+1} &= \phi\alpha_t + \sigma_\eta\eta_t, \quad \eta_t \sim N(0, 1), \quad t = 1, \ldots, n-1, \\
\alpha_1 &\sim N\left(0, \frac{\sigma_\eta^2}{1-\phi^2}\right).
\end{aligned}$$

This implies

$$\begin{aligned}
T_t &= \frac{\partial\alpha_{t+1}}{\partial\alpha_t} = \phi, \quad R_t = \frac{\partial\alpha_{t+1}}{\partial\eta_t} = \sigma_\eta, \quad t = 1, \ldots, n-1, \\
R_0 &= \frac{\partial\alpha_1}{\partial\eta_0} = \frac{\sigma_\eta}{\sqrt{1-\phi^2}}.
\end{aligned}$$

The measurement equation, on the other hand, is nonlinear and the conditional distribution of y_t given $\alpha_1, \dots, \alpha_n$ is normal with mean μ_t and variance σ_t^2 where

$$\mu_t = \begin{cases} \rho\sigma_\epsilon\sigma_\eta^{-1}(\alpha_{t+1} - \phi\alpha_t)\exp(\alpha_t/2), & t = 1, \dots, n-1, \\ 0, & t = n, \end{cases} \tag{15}$$

$$\sigma_t^2 = \begin{cases} (1-\rho^2)\sigma_\epsilon^2 \exp(\alpha_t), & t = 1, \dots, n-1, \\ \sigma_\epsilon^2 \exp(\alpha_n), & t = n. \end{cases} \tag{16}$$

The logarithm of conditional likelihood of y_t (excluding constant term) is given by

$$l_t = -\frac{\alpha_t}{2} - \frac{(y_t - \mu_t)^2}{2\sigma_t^2}. \tag{17}$$

Suppose that $k_{i-1} = s$ and $k_i = s + m$ for the i-th block. Then the log conditional posterior for η_t $(t = s, s+1, \dots, s+m-1)$ is $-\sum_{t=s}^{s+m-1} \eta_t^2/2 + L$ (excluding a constant term) where

$$L = \begin{cases} \sum_{t=s}^{s+m} l_s - \frac{(\alpha_{s+m+1} - \phi\alpha_{s+m})^2}{2\sigma_\eta^2}, & s+m < n, \\ \sum_{t=s}^{s+m} l_s, & s+m = n. \end{cases}$$

The first derivative of the logarithm of the conditional likelihood with respect to α_t is given by

$$d_t = \frac{\partial L}{\partial \alpha_t} = \begin{cases} -\frac{1}{2} + \frac{(y_t - \mu_t)^2}{2\sigma_t^2} + \frac{(y_t - \mu_t)}{\sigma_t^2}\frac{\partial \mu_t}{\partial \alpha_t} + \frac{(y_{t-1} - \mu_{t-1})}{\sigma_{t-1}^2}\frac{\partial \mu_{t-1}}{\partial \alpha_t}, \\ \qquad t = s+1, \dots, s+m-1, \quad \text{or} \quad t = s+m = n, \\ -\frac{1}{2} + \frac{(y_t - \mu_t)^2}{2\sigma_t^2} + \frac{(y_t - \mu_t)}{\sigma_t^2}\frac{\partial \mu_t}{\partial \alpha_t} + \frac{(y_{t-1} - \mu_{t-1})}{\sigma_{t-1}^2}\frac{\partial \mu_{t-1}}{\partial \alpha_t} + \frac{\phi(\alpha_{t+1} - \phi\alpha_t)}{\sigma_\eta^2}, \\ \qquad t = s+m < n, \end{cases} \tag{18}$$

where

$$\frac{\partial \mu_t}{\partial \alpha_t} = \begin{cases} \frac{\rho\sigma_\epsilon}{\sigma_\eta}\left\{-\phi + \frac{(\alpha_{t+1} - \phi\alpha_t)}{2}\right\}\exp\left(\frac{\alpha_t}{2}\right), & t = 1, \dots, n-1, \\ 0, & t = n, \end{cases} \tag{19}$$

$$\frac{\partial \mu_{t-1}}{\partial \alpha_t} = \begin{cases} 0, & t = 1, \\ \frac{\rho\sigma_\epsilon}{\sigma_\eta}\exp\left(\frac{\alpha_{t-1}}{2}\right), & t = 2, \dots, n. \end{cases} \tag{20}$$

Taking expectations of second derivatives multiplied by -1 with respect to y_t's, we obtain the A_t's and B_t's as follows.

$$
\begin{aligned}
A_t &= -E\left(\frac{\partial^2 L}{\partial \alpha_t^2}\right) = \begin{cases} \dfrac{1}{2} + \sigma_t^{-2}\left(\dfrac{\partial \mu_t}{\partial \alpha_t}\right)^2 + \sigma_{t-1}^{-2}\left(\dfrac{\partial \mu_{t-1}}{\partial \alpha_t}\right)^2, \\ \qquad\qquad t = s+1,\ldots,s+m-1, \quad \text{or} \quad t = s+m = n, \\ \dfrac{1}{2} + \sigma_t^{-2}\left(\dfrac{\partial \mu_t}{\partial \alpha_t}\right)^2 + \sigma_{t-1}^{-2}\left(\dfrac{\partial \mu_{t-1}}{\partial \alpha_t}\right)^2 + \phi^2\sigma_\eta^{-2}, \\ \qquad\qquad t = s+m < n, \end{cases} \\
B_t &= -E\left(\frac{\partial^2 L}{\partial \alpha_t \partial \alpha_{t-1}}\right) = \sigma_{t-1}^{-2}\frac{\partial \mu_{t-1}}{\partial \alpha_{t-1}}\frac{\partial \mu_{t-1}}{\partial \alpha_t}, \quad t = 2,\ldots,n.
\end{aligned}
$$

Thus, using Algorithm 1.1 and 1.2, we can generate $(\alpha_{s+1},\ldots,\alpha_{s+m})$ given α_s, α_{s+m+1} (α_s when $s+m=n$) and other parameters.

In this subsection, we assumed that W_t is non-singular. When it is singular, we can still conduct a disturbance smoother and simulation smoother with a slight modification. We first express η_{s+m} using α_{s+m} and α_{s+m+1} such that $\eta_{s+m} = M_{s+m+1}(\alpha_{s+m+1} - T_{s+m}\alpha_{s+m})$ for $s+m<n$. Then noting that

$$
\begin{aligned}
\eta_{s+m}'\eta_{s+m} &= \{M_{s+m+1}(\alpha_{s+m+1} - T_t\alpha_{s+m})\}'\{M_{s+m+1}(\alpha_{s+m+1} - T_t\alpha_{s+m})\} \\
&\equiv (\hat{y}_{s+m}^* - M_{s+m+1}T_{s+m}\alpha_{s+m})'(\hat{y}_{s+m}^* - M_{s+m+1}T_{s+m}\alpha_{s+m})
\end{aligned}
$$

where $\hat{y}_{s+m}^* = M_{s+m+1}\alpha_{s+m+1}$ $(s+m<n)$, we construct an approximating linear and Gaussian state space model as follows. We replace (7) by $L = \sum_{t=s}^{s+m} l_t$ and compute $\hat{d}_t, \hat{A}_t, \hat{B}_t$. Further, to include the term $-\eta_{s+m}'\eta_{s+m}/2$ in the logarithm of the likelihood function, we modify the measurement equation (13) for $t = s+m$ such that

$$
\begin{pmatrix} \hat{y}_{s+m} \\ \hat{y}_{s+m}^* \end{pmatrix} = \begin{pmatrix} I \\ M_{s+m+1}T_{s+m} \end{pmatrix} \alpha_t + [K_{s+m}^{-1}, I, O] \begin{pmatrix} \epsilon_{s+m} \\ \epsilon_{s+m}^* \\ \eta_{s+m} \end{pmatrix} \tag{21}
$$

where ϵ_{s+m}^* follows $N(0,I)$ and is independent of ϵ_{s+m} and η_{s+m}. Using this linear and Gaussian state space model, we can apply Algorithm 1.1 and Algorithm 1.2 when W_t is singular.

3.2 Nonlinear state equation

We extend Algorithm 1.1 and 1.2 to the model with a nonlinear and Gaussian state equation given by

$$
\begin{aligned}
y_t &= h_t(\alpha_t, \epsilon_t), \quad \epsilon_t \sim p_{\epsilon|\eta}(\epsilon_t), \quad t = 1, \ldots, n, && (22)\\
\alpha_{t+1} &= g_t(\alpha_t, \eta_t), \quad \eta_t \sim N(0, I), \quad t = 1, \ldots, n-1, && (23)
\end{aligned}
$$

where $\alpha_0 = 0$.

Normal approximation of the conditional posterior density. To construct a proposal density, we expand the logarithm of the conditional posterior density of η around $\hat{\eta}$ given α_s, α_{s+m+1}, as in the previous section, but further introduce the following auxiliary linear state equation

$$
\begin{aligned}
\beta_{t+1} &= \hat{T}_t \beta_t + \hat{R}_t \eta_t, \quad \eta_t \sim N(0, I), && (24)\\
\hat{T}_t &= \left.\frac{\partial \alpha_{t+1}}{\partial \alpha_t'}\right|_{\eta=\hat{\eta}}, \quad \hat{R}_t = \left.\frac{\partial \alpha_{t+1}}{\partial \eta_t'}\right|_{\eta=\hat{\eta}}, && (25)
\end{aligned}
$$

for $t = s, \ldots, s+m-1$ with an initial condition $\beta_s = \hat{\beta}_s$. When the state equation is linear and Gaussian, we have $\beta_t = \alpha_t$ for $t = s+1, \ldots, s+m$ and $\beta_s = \alpha_s$. Otherwise, we shall take $\beta_s = \hat{\beta}_s = 0$ for convenience sake. Let $L = \sum_{t=s}^{s+m} l_t$ and $\eta = (\eta_s', \ldots, \eta_{s+m-1}')'$. Then

$$
\begin{aligned}
&\log f(\eta|\alpha_s, \alpha_{s+m+1}, y_s, \ldots, y_{s+m})\\
&= \text{const} - \frac{1}{2}\sum_{t=s}^{s+m-1} \eta_t'\eta_t + L + \log p(\alpha_{s+m+1}|\alpha_{s+m})\\
&\approx \text{const} - \frac{1}{2}\sum_{t=s}^{s+m-1} \eta_t'\eta_t + \hat{L} + \hat{d}'(\beta - \hat{\beta}) - \frac{1}{2}(\beta - \hat{\beta})'\hat{Q}(\beta - \hat{\beta}) + \log p(\alpha_{s+m+1}|\hat{\alpha}_{s+m})\\
&= \text{const} + \log f^*(\eta|\alpha_s, \alpha_{s+m+1}, y_s, \ldots, y_{s+m}) + \log p(\alpha_{s+m+1}|\hat{\alpha}_{s+m}), \qquad (26)
\end{aligned}
$$

We separate the term $\log p(\alpha_{s+m+1}|\alpha_{s+m})$ to construct the approximating normal proposal density since its Hessian matrix $\partial^2 \log p(\alpha_{s+m+1}|\alpha_{s+m})/\partial\alpha_{s+m}\partial\alpha_{s+m}'$ may not be negative definite. However, when it is negative definite, we would include this term in L as in Algorithm 1.1.

Mode estimation. Algorithm 2.1 describes how to find a mode, $\hat{\eta}$, of $L - 1/2\sum_{t=s}^{s+m-1} \eta_t'\eta_t$ by repeating it until $\hat{\eta}$ converges (see Appendix A2 for the derivation).

Algorithm 2.1:

1. Initialize $\hat{\eta}$.

2. Evaluate $\hat{T}_t$'s, $\hat{R}_t$'s in (25) at $\eta = \hat{\eta}$ and compute $\hat{\alpha}_t$'s and $\hat{\beta}_t$'s recursively.

$$\begin{aligned} \hat{\alpha}_{t+1} &= g_t(\hat{\alpha}_t, \hat{\eta}_t), \\ \hat{\beta}_{t+1} &= \hat{T}_t \hat{\beta}_t + \hat{R}_t \hat{\eta}_t, \end{aligned}$$

for $t = s, s+1, \ldots, s+m-1$.

3. Evaluate $\hat{d}_t$'s, $\hat{A}_t$'s, and $\hat{B}_t$'s using (8)–(10) at $\alpha = \hat{\alpha}$.

4. Compute the following D_t, J_t and b_t for $t = s+2, \ldots, s+m$ recursively.

$$\begin{aligned} D_t &= \hat{A}_t - \hat{B}_t D_{t-1}^{-1} \hat{B}_t', \quad D_{s+1} = \hat{A}_{s+1}, \\ J_t &= K_{t-1}^{-1\prime} \hat{B}_t, \quad J_{s+1} = O, \quad J_{s+m+1} = O, \\ b_t &= \hat{d}_t - J_t K_{t-1}^{-1} b_{t-1}, \quad b_{s+1} = \hat{d}_{s+1}, \end{aligned}$$

where K_t denotes a Choleski decomposition of D_t such that $D_t = K_t K_t'$.

5. Define auxiliary variables $\hat{y}_t = \hat{\gamma}_t + D_t^{-1} b_t$, where

$$\hat{\gamma}_t = \hat{\beta}_t + K_t^{\prime -1} J_{t+1}' \hat{\beta}_{t+1}, \quad t = s+1, \ldots, s+m,$$

6. Consider the linear Gaussian state-space model with the auxiliary state equation given by

$$\begin{aligned} \hat{y}_t &= Z_t \beta_t + G_t \xi_t, \quad t = s+1, \ldots, s+m, && (27) \\ \beta_{t+1} &= \hat{T}_t \beta_t + H_t \xi_t, \quad t = s, s+1, \ldots, s+m-1, && (28) \\ & \xi_t = (\epsilon_t', \eta_t')' \sim N(0, I), \end{aligned}$$

where

$$Z_t = I + K_t^{\prime -1} J_{t+1}' \hat{T}_t, \quad G_t = K_t^{\prime -1} [I, J_{t+1}' \hat{R}_t], \quad H_t = [O, \hat{R}_t].$$

Apply the Kalman filter and the disturbance smoother to the linear Gaussian system (27) and (28) and obtain the posterior mode $\hat{\eta}$.

7. Goto 2.

Note that the above algorithm produces the posterior mode of η when we include the term $\log p(\alpha_{s+m+1} | \alpha_{s+m})$ in L. If ϵ_t and η_t are independent, the approximating linear Gaussian state-space model reduces to

$$\begin{aligned} \hat{y}_t &= \beta_t + K_t^{\prime -1} \epsilon_t, \quad \epsilon_t \sim N(0, I), \\ \beta_{t+1} &= \hat{T}_t \beta_t + \hat{R}_t \eta_t, \quad \eta_t \sim N(0, I). \end{aligned}$$

To generate η from the conditional posterior density, we conduct the Metropolis-Hastings algorithm using a proposal density $f^*(\eta_y)$.

Algorithm 2.2 (Simulation smoother):

1. Given the current value η_x, find the approximate value of mode, $\hat{\eta}$, using Algorithm 2.1.

2. Proceed Step 2–5 of Algorithm 2.1 to obtain the approximate linear Gaussian system (27)–(28).

3. Generate a candidate η_y from $f^*(\eta_y)$ using a simulation smoother for the approximating linear Gaussian state-space model (27)–(28). Given the current value η_x, we accept η_y with probability

$$\min\left\{1, \frac{f(\eta_y)f^*(\eta_x)}{f(\eta_x)f^*(\eta_y)}\right\}.$$

   If it is rejected, accept η_x as a sample.

Example. Stochastic volatility model with leverage effects and state-dependent variances. In the model (4), the state equation in nonlinear such that

$$\alpha_{t+1} = \phi\alpha_t + \eta_t\sigma_\eta\left\{1 + \frac{1}{1+\exp(-\alpha_t)}\right\}.$$

Then, $\hat{T}_t$ and $\hat{R}_t$ in the auxiliary state equation (24) are

$$\begin{aligned}\hat{T}_t &= \phi + \eta_t\sigma_\eta\frac{\exp(-\hat{\alpha}_t)}{\{1+\exp(-\hat{\alpha}_t)\}^2},\\ \hat{R}_t &= \sigma_\eta\left\{1 + \frac{1}{1+\exp(-\hat{\alpha}_t)}\right\},\\ & \quad t = 1,\ldots,n-1, \quad R_0 = \sigma_0,\end{aligned}$$

respectively. Given α_t's, y_t follows normal distribution with mean μ_t and variance σ_t^2 ($y_t|\alpha \sim N(\mu_t, \sigma_t^2)$) where

$$\mu_t = \rho\sigma_\epsilon\sigma_\eta^{-1}(\alpha_{t+1} - \phi\alpha_t)\left\{1 + \frac{1}{1+\exp(-\alpha_t)}\right\}^{-1}\exp(\alpha_t/2), \qquad (29)$$

and σ_t^2 given by (16). The logarithm of conditional likelihood of y_t (excluding constant term) is the same as in (17).

To sample a block $(\alpha_{s+1},\ldots,\alpha_{s+m})$ given α_s, α_{s+m+1} and other parameters, we consider the log conditional posterior for η_t ($t = s, s+1, \ldots, s+m-1$) given by

$-\sum_{t=s}^{s+m-1}\eta_t^2/2+L$ (excluding a constant term) where

$$L = \begin{cases} \sum_{t=s}^{s+m} l_s - \log\left\{1+\frac{1}{1+\exp(-\alpha_{s+m})}\right\} - \frac{(\alpha_{s+m+1}-\phi\alpha_{s+m})^2}{2\sigma_\eta^2\left\{1+\frac{1}{1+\exp(-\alpha_{s+m})}\right\}^2}, & \text{if } s+m<n, \\ \sum_{t=s}^{s+m} l_s & \text{if } s+m=n, \end{cases}$$

The d_t, first derivative of the L, is the same as in (18) but replacing (19) (20) by

$$\frac{\partial \mu_t}{\partial \alpha_t} = \begin{cases} \frac{\partial \mu_t}{\partial \alpha_{t+1}}\left[-\phi+(\alpha_{t+1}-\phi\alpha_t)\left\{\frac{1}{2}-\frac{1}{3+2\exp(\alpha_t)+\exp(-\alpha_t)}\right\}\right], \\ \qquad t=1,\ldots,n-1, \\ 0, \qquad t=n, \end{cases} \tag{30}$$

$$\frac{\partial \mu_{t-1}}{\partial \alpha_t} = \begin{cases} 0, & t=1, \\ \frac{\rho\sigma_\epsilon}{\sigma_\eta}\left\{1+\frac{1}{1+\exp(-\alpha_{t-1})}\right\}^{-1}\exp\left(\frac{\alpha_t}{2}\right), & t=2,\ldots,n, \end{cases} \tag{31}$$

and the A_t's and B_t's are given by

$$\begin{aligned} A_t &= \frac{1}{2}+\sigma_t^{-2}\left(\frac{\partial \mu_t}{\partial \alpha_t}\right)^2+\sigma_{t-1}^{-2}\left(\frac{\partial \mu_{t-1}}{\partial \alpha_t}\right)^2, \quad t=1,\ldots,n, \\ B_t &= \sigma_{t-1}^{-2}\frac{\partial \mu_{t-1}}{\partial \alpha_{t-1}}\frac{\partial \mu_{t-1}}{\partial \alpha_t}, \quad t=2,\ldots,n. \end{aligned}$$

Using Algorithm 2.1 and 2.2, we generate $(\alpha_{s+1},\ldots,\alpha_{s+m})$ given α_s,α_{s+m+1} (α_s when $s+m=n$) and other parameters.

3.3 Nonlinear and non-Gaussian state equation

When the η_t's are not normally distributed, we may try to transform η_t to follow normal distribution so that we can apply the previous results. If an appropriate transformation cannot be found, we use Taylor expansion around the mode $\hat{\eta}_t$ to obtain the approximate normality as follows. Let $r(\eta_t)$ denote the logarithm of the density for η. By Taylor expansion,

$$\begin{aligned} r(\eta_t) &\approx \text{const}+r(\hat{\eta}_t)+\left.\frac{\partial r}{\partial \eta_t'}\right|_{\eta_t=\hat{\eta}_t}(\eta_t-\hat{\eta}_t)-\frac{1}{2}(\eta_t-\hat{\eta}_t)'\Omega_t(\eta_t-\hat{\eta}_t) \\ &\equiv s(\eta_t), \end{aligned}$$

where $\Omega_t = -\left.\partial^2 r/\partial\eta_t\partial\eta_t'\right|_{\eta_t=\hat{\eta}_t}$. Assuming that Ω_t is positive definite, we approximate it by the normal distribution

$$N(\hat{\mu}_t,\Omega_t^{-1}), \quad \hat{\mu}_t=\hat{\eta}_t+\Omega_t^{-1}\left.\frac{\partial r}{\partial \eta_t}\right|_{\eta_t=\hat{\eta}_t}.$$

Define

$$\tilde{\eta}_t \equiv \Omega_t^{1/2}(\eta_t - \hat{\mu}_t) \sim N(0, I),$$

for $t = s+1, \ldots, s+m$. Then we substitute $\eta_t = \mu + \Omega^{-1/2}\tilde{\eta}_t$ into (27) and (28) to obtain approximating equations (32) and (33) with the auxiliary state equation. Thus we approximate the logarithm of the posterior density $\log f(\eta)$ (where we may exclude the term $\log p(\alpha_{s+m+1}|\alpha_{s+m})$ for the approximation) by $\log f^*(\eta)$ where

$$\begin{aligned}
\log f(\eta) &= \text{constant} + \sum_{t=s+1}^{s+m} r(\eta_t) + L + \log p(\alpha_{s+m+1}|\alpha_{s+m}), \\
\log f^*(\eta) &= \text{constant} + \sum_{t=s+1}^{s+m} s(\eta_t) + \hat{L} + d'(\beta - \hat{\beta}) - \frac{1}{2}(\beta - \hat{\beta})' Q (\beta - \hat{\beta}).
\end{aligned}$$

Algorithm 3.1:

1. Proceed Step 1–5 of Algorithm 2.1.
2. Consider the linear Gaussian state-space model with the auxiliary state equation given by

$$\begin{aligned}
\hat{y}_t &= F_t \hat{\mu}_t + Z_t \beta_t + \tilde{G}_t \tilde{\xi}_t, && (32) \\
\beta_{t+1} &= \hat{R}_t \hat{\mu}_t + \hat{T}_t \beta_t + \tilde{H}_t \tilde{\xi}_t, && (33) \\
& \tilde{\xi}_t = (\epsilon_t', \tilde{\eta}_t') \sim N(0, I),
\end{aligned}$$

   where

$$\begin{aligned}
Z_t &= I + K_t^{'-1} J_{t+1}' \hat{T}_t, \quad F_t = K_t^{'-1} J_{t+1}' \hat{R}_t, \\
\tilde{G}_t &= K_t^{'-1}[I, J_{t+1}' \hat{R}_t \Omega_t^{-1/2}], \quad \tilde{H}_t = [O, \hat{R}_t \Omega_t^{-1/2}].
\end{aligned}$$

   Using the Kalman filter and the disturbance smoother, we obtain a smoothed estimate $\tilde{\eta}_t^*$ of $\tilde{\eta}_t$ and calculate $\hat{\eta}_t = \hat{\mu}_t + \Omega_t^{-1/2} \tilde{\eta}_t^*$.

3. Goto 1.

When ϵ_t and η_t are independent, we have $J_t = O$ for all t and equations (32) and (33) reduce to

$$\begin{aligned}
\hat{y}_t &= Z_t \beta_t + K_t^{'-1} \epsilon_t, \quad \epsilon_t \sim N(0, I), \\
\beta_{t+1} &= \hat{R}_t \hat{\mu}_t + \hat{T}_t \beta_t + \hat{R}_t \Omega_t^{-1/2} \tilde{\eta}_t, \quad \tilde{\eta}_t \sim N(0, I).
\end{aligned}$$

To sample α_t's and η_t's using a block sampler, we implement MCMC draws as follows.

Algorithm 3.2:(Simulation smoother)

1. Proceed Step 1–2 of Algorithm 4.1 to obtain an approximate mode $\hat{\eta}$ of the logarithm of $f(\eta)$ (where we may exclude the term $\log p(\alpha_{s+m+1}|\alpha_{s+m})$ for the approximation).

2. Propose a candidate of $\tilde{\eta}_t$ (and compute $\eta_t = \hat{\mu}_t + \Omega_t^{-1/2}\tilde{\eta}_t$) by simulation smoother for the auxiliary linear Gaussian state-space model given by (32) and (33). Given the current value η_x, we accept η_y with probability

$$\min\left\{1, \frac{f(\eta_y)f^*(\eta_x)}{f(\eta_x)f^*(\eta_y)}\right\}.$$

If it is rejected, accept η_x as a sample.

4 Illustrative examples

We illustrate how to implement our block sampler of state variables α_t's using simulated data and stock returns data. We show that our method attains a considerable improvement in the estimation efficiency compared with results from using a single move sampler (which samples one α_t at a time given $\alpha_{-t} = (\alpha_1, \ldots, \alpha_{t-1}, \alpha_{t+1}, \ldots, \alpha_n)$).

4.1 Stochastic volatility model with leverage effects

4.1.1 MCMC algorithm

Let y, Σ denote $y = (y_1, \ldots, y_n)'$ and

$$\Sigma = \begin{pmatrix} \sigma_\epsilon^2 & \rho\sigma_\epsilon\sigma_\eta \\ \rho\sigma_\epsilon\sigma_\eta & \sigma_\eta^2 \end{pmatrix},$$

respectively. We first initialize $\{\alpha_t\}_{t=1}^n, \phi, \Sigma$ and proceed an MCMC implementation in 3 steps.

1. Sample $\{\alpha_t\}_{t=1}^n|\phi, \Sigma, y$.
   (a) Generate K stochastic knots $(k_1, \ldots, k_K)$ and set $k_0 = 0$, $k_{K+1} = n$.
   (b) Sample $\{\alpha_t\}_{t=k_{i-1}+1}^{k_i}|\{\alpha_t|t \leq k_{i-1}, t > k_i\}, \phi, \Sigma, y$ for $i = 1, \ldots, K+1$.

2. Sample $\phi|\{\alpha_t\}_{t=1}^n, \Sigma, y$.

3. Sample $\Sigma|\{\alpha_t\}_{t=1}^n, \phi, y$.

Step 1. We construct blocks by dividing $(\alpha_1, \ldots, \alpha_n)$ into $K+1$ blocks, $(\alpha_{k_{i-1}+1}, \ldots, \alpha_{k_i})'$ using $(k_1, \ldots, k_K)$ with $k_0 = 0$, $k_{K+1} = n$, $k_i - k_{i-1} \geq 2$ for $i = 1, \ldots, K+1$. The K

knots, $(k_1, \ldots, k_K)$, are generated randomly using

$$k_i = \text{int}[n \times (i + U_i)/(K+2)], \quad i = 1, \ldots, K,$$

where U_i's are independent uniform random variables on $(0,1)$ (see e.g. Shephard and Pitt (1997), Watanabe and Omori (2004)) . For each block, use Algorithm 1.1 and 2.1 to generate state variables $(\alpha_{k_{i-1}+1}, \ldots, \alpha_{k_i})$ $i = 1, \ldots, K+1$.

Step 2. Let $\pi(\phi)$ denote a prior probability density for ϕ. The logarithm of the conditional posterior density for ϕ (excluding a constant term) is given by

$$\log \pi(\phi) + \frac{1}{2}\log(1-\phi^2) - \frac{\alpha_1^2(1-\phi^2)}{2\sigma_\eta^2} - \frac{\sum_{t=1}^{n-1}\left\{\alpha_{t+1} - \phi\alpha_t - \rho\sigma_\eta\sigma_\epsilon^{-1}\exp(-\alpha_t/2)y_t\right\}^2}{2(1-\rho^2)\sigma_\eta^2}.$$

We propose a candidate for the MH algorithm using a truncated normal distribution on $(-1,1)$, with mean μ_ϕ and variance σ_ϕ^2 (which we denote by $\phi \sim TN_{(-1,1)}(\mu_\phi, \sigma_\phi^2)$) where

$$\mu_\phi = \frac{\sum_{t=1}^{n-1}\alpha_t\left(\alpha_{t+1} - \rho\sigma_\eta\sigma_\epsilon^{-1}e^{-\alpha_t/2}y_t\right)}{\rho^2\alpha_1^2 + \sum_{t=2}^{n-1}\alpha_t^2}, \quad \sigma_\phi^2 = \frac{(1-\rho^2)\sigma_\eta^2}{\rho^2\alpha_1^2 + \sum_{t=2}^{n-1}\alpha_t^2}.$$

Given the current sample ϕ_x, generate $\phi_y \sim TN_{(-1,1)}(\mu_\phi, \sigma_\phi^2)$ and accept it with probability

$$\min\left\{\frac{\pi(\phi_y)\sqrt{1-\phi_y^2}}{\pi(\phi_x)\sqrt{1-\phi_x^2}}, 1\right\}.$$

Step 3. We assume that a prior distribution of Σ^{-1} follows Wishart distribution (which we denote by $\Sigma^{-1} \sim W(\nu_0, \Sigma_0)$). Then the logarithm of the conditional posterior density of Σ (excluding a constant term) is

$$-\log\sigma_\eta - \frac{\alpha_1^2(1-\phi^2)}{2\sigma_\eta^2} - \frac{\nu_1}{2}\log|\Sigma| - \frac{1}{2}\text{tr}\left(\Sigma_1^{-1}\Sigma^{-1}\right),$$

where

$$\nu_1 = \nu_0 + n - 1, \quad \Sigma_1^{-1} = \Sigma_0^{-1} + \sum_{t=1}^{n-1} x_t x_t', \quad x_t = (y_t\exp(-\alpha_t/2), \alpha_{t+1} - \phi\alpha_t).$$

We sample Σ using MH algorithm with a proposal $\Sigma^{-1} \sim W(\nu_1, \Sigma_1)$. Given the current value Σ_x^{-1}, generate $\Sigma_y^{-1} \sim W(\nu_1, \Sigma_1)$ and accept it with probability

$$\min \left\{ \frac{\sigma_{\eta,y}^{-1} \exp - \dfrac{\alpha_1^2(1-\phi^2)}{2\sigma_{\eta,y}^2}}{\sigma_{\eta,x}^{-1} \exp - \dfrac{\alpha_1^2(1-\phi^2)}{2\sigma_{\eta,x}^2}}, 1 \right\}.$$

4.1.2 Illustration using simulated data

To simulate the daily financial data, we set $\phi = 0.97, \sigma_\epsilon = 1, \sigma_\eta = 0.1, \rho = -0.5$ and generate $n = 1,000$ observations. We take a beta distribution with parameters 20 and 1.5 for the $(1+\phi)/2$ and hence the prior mean and standard deviation of ϕ are 0.86 and 0.11 respectively. For a prior distribution of Σ^{-1}, we assume a less informative distribution and take a Wishart distribution with $\nu_0 = 0.01$ and Σ_0^{-1} equal to the true value of $0.01 \times \Sigma$. The computational results were generated using Ox version 4.04 (Doornik (2002)) throughout.

Summary statistics

Number of blocks = 40						
Parameter	True	Mean	Stdev	95% interval	Inefficiency	CD
ϕ	0.97	0.984	0.011	[0.957, 0.997]	260.1	0.94
σ_ϵ	1.0	0.930	0.084	[0.756, 1.105]	279.0	0.13
σ_η	0.1	0.080	0.026	[0.040, 0.140]	432.7	0.83
ρ	−0.5	−0.387	0.206	[−0.729, 0.058]	68.7	0.42

Table 1: Summary statistics. The number of MCMC iterations is 50,000, and sample size is 1,000. The bandwidth 5,000 is used to compute the inefficiency factors and CD (p value of convergence diagnostic test).

Estimation results. We set $K = 40$ so that each block contains 25 α_t's on the average. The initial 5,000 iterations are discarded as burn-in period and the following 50,000 iterations are recorded. Table 1 summarises the posterior means, standard deviations, 95% credible intervals, inefficiency factors and p values of convergence diagnostic tests by Geweke (1992) for the parameters $\phi, \sigma_\epsilon, \sigma_\eta$ and ρ. The posterior means are close to true values and true values of all parameters are covered in 95% credible intervals. All p values of convergence diagnostic (CD) tests are greater than 0.05, suggesting that there is no significant evidence against the convergence of the distribution of MCMC samples to the posterior distribution.

The inefficiency factor is defined as $1 + 2\sum_{s=1}^{\infty} \rho_s$ where ρ_s is the sample autocorrelation at lag s, and are computed to measure how well the MCMC chain mixes (see e.g. Chib (2001)). It is the ratio of the numerical variance of the posterior sample mean to the variance of the sample mean from uncorrelated draws. The inverse of inefficiency factor is also known as relative numerical efficiency (Geweke (1992)). When the inefficiency factor is equal to m, we need to draw MCMC samples m times as many as uncorrelated samples.

Summary statistics

Single move sampler						
Parameter	True	Mean	Stdev	95% interval	Inefficiency	CD
ϕ	0.97	0.973	0.015	[0.937, 0.994]	2199.2	0.30
σ_ϵ	1.0	0.918	0.078	[0.763, 1.058]	103.1	0.39
σ_η	0.1	0.099	0.025	[0.060, 0.420]	3506.6	0.09
ρ	-0.5	-0.324	0.172	$[-0.595, 0.064]$	1038.0	0.47

Table 2: Summary statistics for the single move sampler. The number of MCMC iteration is 250,000 and sample size is 1,000. The bandwidth 25,000 is used to compute the inefficiency factors and CD (p value of convergence diagnostic test).

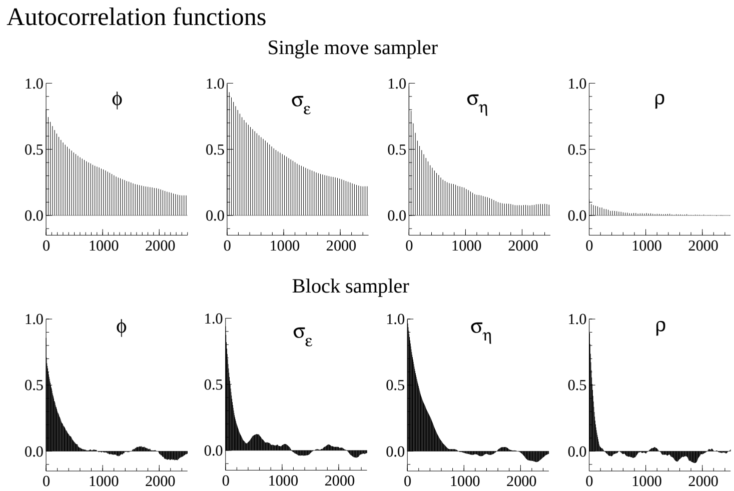

Figure 1: Sample autocorrelation functions of MCMC samples.

Comparison with a single move sampler. To show the efficiency of our proposed block sampler using inefficiency factors, we also conducted a single move sampler which samples one α_t at a time. We employ the algorithm of the single move sampler proposed by Jacquier *et al.* (2004) with a slight modification since they modeled the leverage effects in a different manner (where they considered the correlation between ϵ_t and η_{t-1}). The initial 25,000 iterations are discarded as burn-in period and the following $250,000$ iterations are recorded since obtained MCMC samples are highly autocorrelated and a large number of draws need to be taken to obtain stable and reliable estimation results. Table 2 shows summary statistics of the experiment using a single move sampler. The inefficiency factors of the sampler are between 100 and 3510, while those of the block sampler are between 60 and 440. This implies that our

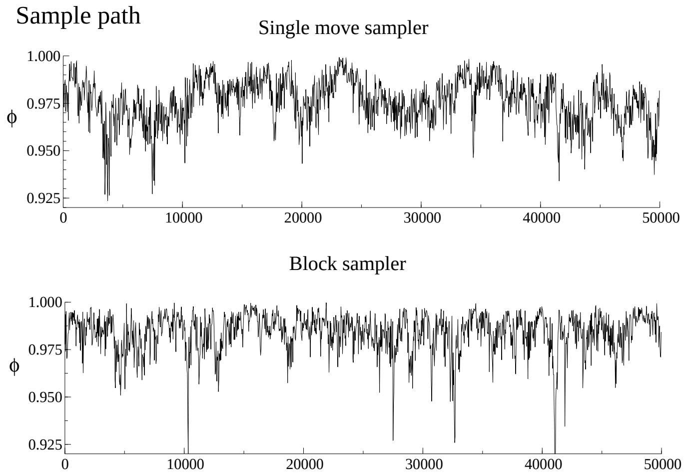

Figure 2: Sample path of ϕ's using first 50,000 MCMC samples.

proposed sampler reduces sample autocorrelations considerably and that it produces more accurate estimation results than the single move sampler. In Figure 1, we can see clear reductions in the sample autocorrelation functions for the block sampler in all parameters. Figure 2 shows sample paths of ϕ's using first 50,000 MCMC draws. The sample path of the single move sampler does not move as fast as the block sampler in the state space. These results clearly show that our method produces great improvement in the mixing property of MCMC chains.

Selection of a number of blocks. To investigate the effect of block sizes on the speed of

Inefficiency factors

Parameter	Number of blocks							
	5	10	20	30	40	50	100	200
ϕ	314.1	329.0	220.8	254.4	260.1	**185.6**	347.0	599.4
σ_ϵ	526.7	**153.8**	312.3	449.4	279.0	680.9	684.4	1897.7
σ_η	465.3	538.9	**394.6**	452.6	432.7	322.5	524.7	687.4
ρ	172.8	178.7	266.6	251.5	**68.7**	301.7	235.3	193.4
α_{500}	264.2	**134.4**	142.5	237.3	138.7	305.3	394.4	1183.3

Table 3: Inefficiency factors of MCMC samples using various number of blocks.

convergence to the posterior distribution, we repeated our experiments using different number of blocks varying from 5 blocks to 200 blocks. The inefficiency factors of MCMC samples are shown in Table 3. They tend to be larger as the number of blocks increases from 40 to 200, while the small number of blocks such as 5 blocks would also lead to high inefficiency factors. The latter is a result of low acceptance rates in MH algorithm for the α_t's in the block sampler as shown in Table 4.

Acceptance rates in MH algorithm

Parameter	Number of blocks							
	5	10	20	30	40	50	100	200
α (AR)	0.820	0.878	0.926	0.946	0.954	0.964	0.981	0.990
α	0.817	0.886	0.935	0.955	0.962	0.972	0.986	0.993
ϕ	0.793	0.798	0.792	0.793	0.813	0.797	0.800	0.794
Σ	0.984	0.983	0.985	0.985	0.985	0.984	0.984	0.985

Table 4: Acceptance rates in MH algorithm. α(AR) corresponds to the acceptance rate in acceptance-rejection algorithm.

When the number of blocks is equal to 5, the acceptance rate of α_t's is 81.7%. This is relatively smaller than those obtained with larger number of blocks since high dimensional probability density of α_t would be more difficult to be approximated by multivariate normal density. In this example, the optimal number of blocks with small inefficiency factors would be between 20 and 40 where average block sizes are between 25 and 50.

4.1.3 Stock returns data

We next apply our method to the daily Japanese stock returns. Using TOPIX (Tokyo Stock Price Index) from 1 August 1997 to 31 July 2002, the stock returns are computed as 100 times the difference of the logarithm of the series. The times series plot is shown in Figure 3 where the number of observations is 1,230.

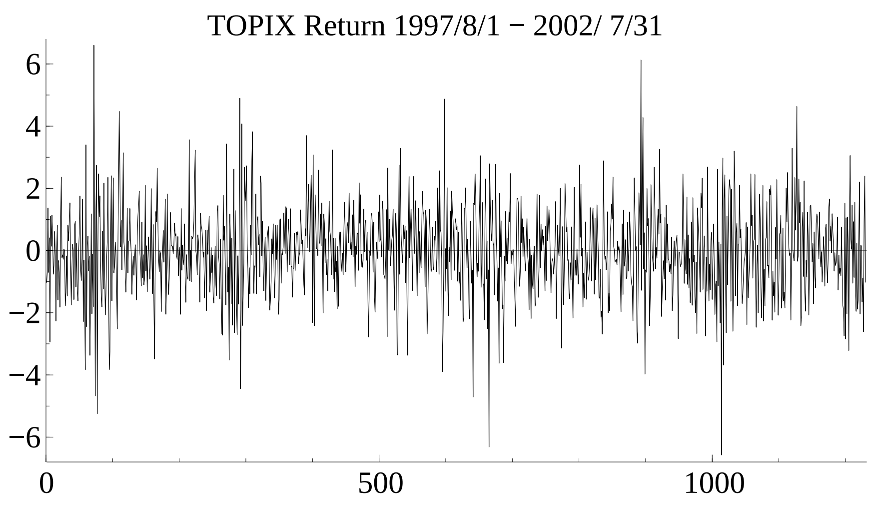

Figure 3: TOPIX return data. 1997/8/1 – 2002/7/31.

Summary statistics

			Number of blocks = 40		
Parameter	Mean	Stdev	95% interval	Inefficiency	CD
ϕ	0.945	0.019	[0.902, 0.974]	118.2	0.24
σ_ϵ	1.259	0.070	[1.121, 1.398]	20.8	0.06
σ_η	0.193	0.033	[0.138, 0.267]	206.7	0.32
ρ	-0.442	0.103	$[-0.630, -0.231]$	92.7	0.89

Table 5: Summary statistics. The number of MCMC iteration is 50,000. The bandwidth 5,000 is used to compute the inefficiency factors and CD.

The prior distribution of parameters, the number of blocks, the number of iterations and the burn-in period are taken as in the simulated data example. Table 5 shows summary statistics of MCMC samples. The results are similar to those obtained in the previous subsection. Since 95% credible interval for ρ is $(-0.630, -0.231)$ with the posterior mean -0.442, the posterior probability that ρ is negative is greater than 0.95. It shows the importance of leverage effects in the stochastic volatility model as we expected. Although the acceptance rates of α_t's in Metropolis-Hastings algorithm are relatively small as shown in Table 6, inefficiency factors of obtained samples are found to be small. This is because the sample size is larger than that of previous examples and the average block size becomes larger accordingly.

Acceptance rates in MH algorithm

Parameter	Acceptance rates
α (AR)	0.852
α	0.856
ϕ	0.955
Σ	0.990

Table 6: TOPIX data. Acceptance rates in MH algorithm. α(AR) corresponds to the acceptance rate in acceptance-rejection algorithm.

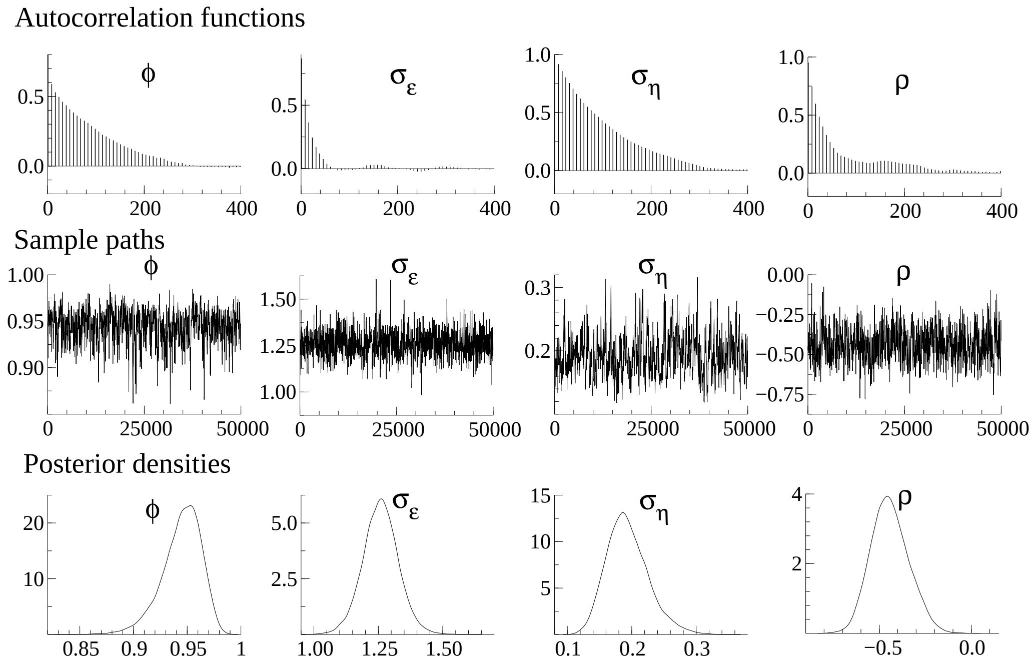

Figure 4: Sample autocorrelation functions of MCMC samples.

Figure 4 shows sample autocorrelation functions, sample paths and the posterior densities. The sample autocorrelations decay quickly and MCMC samples move fast over the state space.

4.2 Stochastic volatility model with state-dependent variances

This subsection illustrates our method using simulated data generated by the stochastic volatility model in (4). The MCMC algorithm proceed in 3 steps as in Section 4.1. We use Algorithm 2.1 and 2.2 to generate $(\alpha_{s+1}, \ldots, \alpha_{s+m})$ given α_s, α_{s+m+1} (α_s when $s+m = n$) and other parameters. Then, given α_t's, we sample from conditional posterior distribution of ϕ and Σ as in previous subsection.

We set $\phi = 0.95, \sigma_\epsilon = 1, \sigma_\eta = 0.1, \rho = -0.5$ and generate $n = 1,000$ observations. The distribution of the initial state α_1 is assumed to be $N(0, 0.1)$. The prior distribution of other parameters are taken as in the previous example.

Summary statistics

Number of blocks = 30						
Parameter	True	Mean	Stdev	95% interval	Inefficiency	CD
ϕ	0.95	0.944	0.019	[0.900, 0.975]	192.9	0.55
σ_ϵ	1.0	0.994	0.056	[0.887, 1.111]	86.2	0.32
σ_η	0.1	0.129	0.025	[0.088, 0.184]	332.4	0.53
ρ	−0.5	−0.415	0.117	[−0.624, −0.172]	116.3	0.15

Table 7: Summary statistics. The number of MCMC iterations is 50,000 and sample size is 1,000. The bandwidth 5,000 is used to compute the inefficiency factors and CD.

We set $K = 30$ and the initial 20,000 iterations are discarded as burn-in period and the following $50,000$ iterations are recorded. Table 7 summarises the posterior means, standard deviations, 95% credible intervals, inefficiency factors and p values of convergence diagnostic tests for the parameters $\phi, \sigma_\epsilon, \sigma_\eta$ and ρ. The posterior means are close to true values and true values of all parameters are covered in 95% credible intervals. All p values of convergence diagnostic tests are greater than 0.05, suggesting that there is no significant evidence against the convergence of the distribution of MCMC samples to the posterior distributions.

Inefficiency factors

Parameter	Number of blocks							
	5	10	20	30	40	50	100	200
ϕ	207.1	396.4	199.2	192.9	252.4	273.0	243.6	**191.5**
σ_ϵ	94.6	**47.0**	80.3	86.2	60.2	132.3	71.5	267.8
σ_η	372.4	618.4	347.1	**332.4**	427.0	433.2	434.8	403.3
ρ	224.9	93.4	171.1	116.3	**91.1**	96.1	145.8	126.4
α_{500}	15.1	10.0	15.0	12.1	14.2	20.5	**8.7**	36.0

Table 8: Inefficiency factors of MCMC samples using various number of blocks.

Table 8 shows the the effect of block sizes on the mixing property of chains. As shown in Section 4.1, the larger the number of blocks becomes (from 40 to 200), the larger the inefficiency factors become. On the other hand, very small number of blocks such as 5 blocks resulted in high inefficiency factors. In Table 9, acceptance rates of the Metropolis-Hastings algorithm are shown. The acceptance rates of α are much smaller than those in the previous section due to dropping the terms $\log p(\alpha_{s+m+1}|\hat{\alpha}_{s+m})$ in (26). The appropriate number of blocks for this particular example would be from 20 to 40.

Acceptance rates in MH algorithm

Parameter	Number of blocks							
	5	10	20	30	40	50	100	200
α	0.307	0.383	0.428	0.450	0.460	0.470	0.501	0.534
ϕ	0.986	0.985	0.987	0.987	0.986	0.985	0.985	0.987
Σ	0.993	0.993	0.993	0.992	0.993	0.993	0.992	0.993

Table 9: Acceptance rates in MH algorithm. α(AR) corresponds to the acceptance rate in acceptance-rejection algorithm.

5 Conclusion

In this article, we described a disturbance smoother and a simulation smoother for a general state-space model with a non-Gaussian measurement equation and a nonlinear and non-Gaussian state equation. The dependent variable and the state variable are allowed to be correlated. The high performance of our proposed method is shown in estimation efficiencies using illustrative numerical examples in comparison with a single move sampler.

Acknowledgement

The authors thank Herman K. van Dijk and Hisashi Tanizaki for helpful discussions. This work is partially supported by Seimeikai Foundation, The Japan Economic Research Foundation, and Grants-in-Aid for Scientific Research 15500181, 15530221, 18330039, 18203901 from the Japanese Ministry of Education, Science, Sports, Culture and Technology.

Appendix A1

Suppose that a state equation is nonlinear such that

$$\alpha_{t+1} = g_t(\alpha_t, \eta_t), \quad \eta_t \sim N(0, I), \quad t = s, \ldots, s+m-1,$$

(α_s : given). Consider an auxiliary state equation given by

$$\beta_{t+1} = \hat{T}_t \beta_t + \hat{R}_t \eta_t, \quad t = s, \ldots, s+m-1,$$

with $\beta_s = \hat{\beta}_s$, where

$$\hat{T}_t = \left. \frac{\partial \alpha_{t+1}}{\partial \alpha_t'} \right|_{\eta=\hat{\eta}}, \quad \hat{R}_t = \left. \frac{\partial \alpha_{t+1}}{\partial \eta_t'} \right|_{\eta=\hat{\eta}}.$$

For a linear Gaussian state equation, we replace β_t by α_t and set $\alpha_{t+1} = T_t\alpha_t + R_t\eta_t$. Using

$$\frac{\partial L}{\partial \eta_j'} = \sum_{t=j+1}^{s+m} \frac{\partial L}{\partial \alpha_t'} \frac{\partial \alpha_t}{\partial \eta_j'}, \qquad \frac{\partial \alpha_t}{\partial \eta_j'} = \begin{cases} \frac{\partial \alpha_t}{\partial \alpha_{t-1}'} \cdots \frac{\partial \alpha_{j+2}}{\partial \alpha_{j+1}'} \frac{\partial \alpha_{j+1}}{\partial \eta_j'}, & t \geq j+1, \\ 0 & t \leq j, \end{cases}$$

and

$$\begin{aligned} \beta_t &= \hat{T}_{t-1}\beta_{t-1} + \hat{R}_{t-1}\eta_{t-1} \\ &= \sum_{j=s}^{t-1} \hat{T}_{t-1} \cdots \hat{T}_{j+1}\hat{R}_j\eta_j + \hat{T}_{t-1} \cdots \hat{T}_s\beta_s \\ &= \sum_{j=s}^{t-1} \left.\frac{\partial \alpha_t}{\partial \eta_j'}\right|_{\eta=\hat{\eta}} \eta_j + \hat{T}_{t-1} \cdots \hat{T}_s\beta_s, \end{aligned}$$

we obtain

$$\begin{aligned} \left.\frac{\partial L}{\partial \eta}\right|_{\eta=\hat{\eta}}' (\eta - \hat{\eta}) &= \sum_{j=s}^{s+m-1} \left.\frac{\partial L}{\partial \eta_j}\right|_{\eta=\hat{\eta}}' (\eta_j - \hat{\eta}_j) \\ &= \sum_{j=s}^{s+m-1} \sum_{t=j+1}^{s+m} \left.\frac{\partial L}{\partial \alpha_t'}\right|_{\alpha=\hat{\alpha}} \left.\frac{\partial \alpha_t}{\partial \eta_j'}\right|_{\eta=\hat{\eta}} (\eta_j - \hat{\eta}_j) \\ &= \sum_{t=s+1}^{s+m} \hat{d}_t' \sum_{j=s}^{t-1} \left.\frac{\partial \alpha_t}{\partial \eta_j'}\right|_{\eta=\hat{\eta}} (\eta_j - \hat{\eta}_j) \\ &= \sum_{t=s+1}^{s+m} \hat{d}_t'(\beta_t - \hat{\beta}_t) = \hat{d}'(\beta - \hat{\beta}). \end{aligned} \tag{34}$$

where $\alpha = (\alpha_{s+1}', \ldots, \alpha_{s+m}')'$, $\beta = (\beta_{s+1}', \ldots, \beta_{s+m}')'$, On the other hand, the second derivative of log likelihood is given by

$$\begin{aligned} \frac{\partial L^2}{\partial \eta_{il} \partial \eta_{jm}} &= \frac{\partial}{\partial \eta_{il}} \sum_{t_2=j+1}^{s+m} \sum_{k_2=1}^{p} \frac{\partial L}{\partial \alpha_{t_2k_2}} \frac{\partial \alpha_{t_2k_2}}{\partial \eta_{jm}} \\ &= \sum_{t_2=j+1}^{s+m} \sum_{k_2=1}^{p} \frac{\partial^2 L}{\partial \eta_{il} \partial \alpha_{t_2k_2}} \frac{\partial \alpha_{t_2k_2}}{\partial \eta_{jm}} + \frac{\partial L}{\partial \alpha_{t_2k_2}} \frac{\partial^2 \alpha_{t_2k_2}}{\partial \eta_{il} \partial \eta_{jm}} \\ &= \sum_{t_2=j+1}^{s+m} \sum_{k_2=1}^{p} \left(\sum_{t_1=i+1}^{s+m} \sum_{k_1=1}^{p} \frac{\partial L^2}{\partial \alpha_{t_1k_1} \partial \alpha_{t_2k_2}} \frac{\partial \alpha_{t_1k_1}}{\partial \eta_{il}} \frac{\partial \alpha_{t_2k_2}}{\partial \eta_{jm}} \right) + \frac{\partial L}{\partial \alpha_{t_2k_2}} \frac{\partial^2 \alpha_{t_2k_2}}{\partial \eta_{il} \partial \eta_{jm}}. \end{aligned}$$

Its expected value is

$$E\left(\frac{\partial L^2}{\partial \eta_{il} \partial \eta_{jm}}\right) = \sum_{t_1=i+1}^{s+m} \sum_{t_2=j+1}^{s+m} \sum_{k_1=1}^{p} \sum_{k_2=1}^{p} E\left(\frac{\partial L^2}{\partial \alpha_{t_1k_1} \partial \alpha_{t_2k_2}}\right) \frac{\partial \alpha_{t_1k_1}}{\partial \eta_{il}} \frac{\partial \alpha_{t_2k_2}}{\partial \eta_{jm}}.$$

Thus the (i,j) block of the information matrix is

$$E\left(\frac{\partial L^2}{\partial \eta_i \partial \eta_j'}\right) = \sum_{t_1=i+1}^{s+m}\sum_{t_2=j+1}^{s+m} \frac{\partial \alpha_{t_1}}{\partial \eta_i} E\left(\frac{\partial L^2}{\partial \alpha_{t_1}\partial \alpha_{t_2}'}\right)\frac{\partial \alpha_{t_2}}{\partial \eta_j'}.$$

Therefore, we obtain

$$\begin{aligned}
&(\eta-\hat{\eta})' \, E\left(\frac{\partial L^2}{\partial \eta \partial \eta'}\right)\bigg|_{\eta=\hat{\eta}} (\eta-\hat{\eta}) \\
&= \sum_{i=s}^{s+m-1}\sum_{j=s}^{s+m-1} (\eta_i-\hat{\eta}_i)' \, E\left(\frac{\partial L^2}{\partial \eta_i \partial \eta_j'}\right)\bigg|_{\eta=\hat{\eta}} (\eta_j-\hat{\eta}_j) \\
&= \sum_{i=s}^{s+m-1}\sum_{j=s}^{s+m-1}\sum_{t_1=i+1}^{s+m}\sum_{t_2=j+1}^{s+m} (\eta_i-\hat{\eta}_i)' \, \frac{\partial \alpha_{t_1}}{\partial \eta_i}\bigg|_{\eta=\hat{\eta}} E\left(\frac{\partial L^2}{\partial \alpha_{t_1}\partial \alpha_{t_2}'}\right)\bigg|_{\eta=\hat{\eta}} \frac{\partial \alpha_{t_2}}{\partial \eta_j'}\bigg|_{\eta=\hat{\eta}} (\eta_j-\hat{\eta}_j) \\
&= \sum_{t_1=s+1}^{s+m}\sum_{t_2=s+1}^{s+m}\sum_{i=s}^{t_1-1}\sum_{j=s}^{t_2-1} (\eta_i-\hat{\eta}_i)' \, \frac{\partial \alpha_{t_1}}{\partial \eta_i}\bigg|_{\eta=\hat{\eta}} E\left(\frac{\partial L^2}{\partial \alpha_{t_1}\partial \alpha_{t_2}'}\right)\bigg|_{\eta=\hat{\eta}} \frac{\partial \alpha_{t_2}}{\partial \eta_j'}\bigg|_{\eta=\hat{\eta}} (\eta_j-\hat{\eta}_j) \\
&= \sum_{t_1=s+1}^{s+m}\sum_{t_2=s+1}^{s+m} (\beta_{t_1}-\hat{\beta}_{t_1})' \, E\left(\frac{\partial L^2}{\partial \alpha_{t_1}\partial \alpha_{t_2}'}\right)\bigg|_{\eta=\hat{\eta}} (\beta_{t_2}-\hat{\beta}_{t_2}) \\
&= (\beta-\hat{\beta})' \, E\left(\frac{\partial L^2}{\partial \alpha \partial \alpha'}\right)\bigg|_{\eta=\hat{\eta}} (\beta-\hat{\beta}) = -(\beta-\hat{\beta})'\hat{Q}(\beta-\hat{\beta}). \qquad (35)
\end{aligned}$$

The results are obtained from equations (34) and (35).

Appendix A2

Since $\hat{Q}$ is assumed to be a positive definite matrix, there exists a lower triangular matrix U such that $\hat{Q} = UU'$ using a Choleski decomposition where

$$U = \begin{pmatrix} K_{s+1} & O & O & \dots & O \\ J_{s+2} & K_{s+2} & O & \dots & O \\ O & J_{s+3} & K_{s+3} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & O \\ O & \dots & O & J_{s+m} & K_{s+m} \end{pmatrix},$$

so that

$$\begin{aligned}
\hat{A}_t &= J_t J_t' + K_t K_t', \quad t = s+1, \dots, s+m, \\
\hat{B}_t &= J_t K_{t-1}', \quad t = s+2, \dots, s+m, \\
&\quad B_{s+1} = J_{s+1} = O.
\end{aligned}$$

Denote $C_t = J_t J_t', D_t = K_t K_t'$ and we obtain

$$
\begin{aligned}
C_t &= \hat{B}_t (K_{t-1} K_{t-1}')^{-1} \hat{B}_t' = \hat{B}_t D_{t-1}^{-1} \hat{B}_t', \\
D_t &= \hat{A}_t - C_t = \hat{A}_t - \hat{B}_t D_{t-1}^{-1} \hat{B}_t',
\end{aligned}
$$

for $t = s+2, \ldots, s+m$, and $D_{s+1} = \hat{A}_{s+1}$. The matrix K_t is a Choleski decomposition of D_t and $J_t = K_{t-1}^{-1'} \hat{B}_t$. Let $K = \mathrm{diag}(K_{s+1}, \ldots, K_{s+m})$, $D = \mathrm{diag}(D_{s+1}, \ldots, D_{s+m})$, $b = KU^{-1}\hat{d}$, $\gamma = K^{'-1}U'\beta$, and $\hat{\gamma} = K^{'-1}U'\hat{\beta}$. Then

$$
\begin{aligned}
&\hat{d}'(\beta - \hat{\beta}) - \frac{1}{2}(\beta - \hat{\beta})'\hat{Q}(\beta - \hat{\beta}) \\
&= b'K^{'-1}U'(\beta - \hat{\beta}) - \frac{1}{2}(\beta - \hat{\beta})'UK^{-1}DK^{'-1}U'(\beta - \hat{\beta}) \\
&= b'(\gamma - \hat{\gamma}) - \frac{1}{2}(\gamma - \hat{\gamma})'D(\gamma - \hat{\gamma}) \\
&= -\frac{1}{2}(\hat{y} - \gamma)'D(\hat{y} - \gamma) \qquad (36)
\end{aligned}
$$

where $\hat{y} = \hat{\gamma} + D^{-1}b$, $\hat{y}_t = \hat{\gamma}_t + D_t^{-1}b_t$. On the other hand, since $\hat{d} = UK^{-1}b$, and $\gamma = K^{'-1}U'\beta$,

$$
\begin{aligned}
\gamma_t &= \beta_t + K_t^{'-1}J_{t+1}'\beta_{t+1}, t = s+1, \ldots, s+m, \quad J_{s+m+1} = O, \\
b_t &= \hat{d}_t - J_t K_{t-1}^{-1} b_{t-1}, \quad t = s+2, \ldots, s+m, \quad b_{s+1} = d_{s+1}.
\end{aligned}
$$

Thus, given β_t ($t = s, s+1, \ldots, s+m$), the equation (36) is a likelihood function for

$$
\begin{aligned}
\hat{y}_t &= \beta_t + K_t^{'-1}J_{t+1}'\beta_{t+1} + K_t^{'-1}\epsilon_t \\
&= \beta_t + K_t^{'-1}J_{t+1}'(\hat{T}_t\beta_t + \hat{R}_t\eta_t) + K_t^{'-1}\epsilon_t \\
&= Z_t\beta_t + G_t\xi_t, \qquad (37) \\
&\quad \xi_t = (\epsilon_t', \eta_t')' \sim N(0, I).
\end{aligned}
$$

where $Z_t = I + K_t^{'-1}J_{t+1}'\hat{T}_t$ and $G_t = K_t^{'-1}[I, J_{t+1}'\hat{R}_t]$.

Appendix A3

Consider the log conditional posterior density of the approximating model given in (11). Equation (11) is equal to (excluding constant term),

$$\begin{aligned}
&-\frac{1}{2}\sum_{t=s}^{s+m-1}(\alpha_{t+1}-T_t\alpha_t)'V_t(\alpha_{t+1}-T_t\alpha_t)+\hat{L}+\hat{d}'(\alpha-\hat{\alpha})-\frac{1}{2}(\alpha-\hat{\alpha})'\hat{Q}(\alpha-\hat{\alpha})\\
&= -\frac{1}{2}\alpha'T'VT\alpha+\hat{L}+\hat{d}'(\alpha-\hat{\alpha})-\frac{1}{2}(\alpha-\hat{\alpha})'\hat{Q}(\alpha-\hat{\alpha})\\
&= \text{const}+\tilde{d}'(\alpha-\hat{\alpha})-\frac{1}{2}(\alpha-\hat{\alpha})'\tilde{Q}(\alpha-\hat{\alpha})
\end{aligned}$$

where $V_t = W_t^{-1}, V = \text{diag}(V_s, \ldots, V_{s+m-1})$, $\tilde{d} = \hat{d} - T'VT\hat{\alpha}, \tilde{Q} = \hat{Q} + T'VT$ and

$$T = \begin{pmatrix} I & O & O & \ldots & O \\ -T_{s+1} & I & O & \ldots & O \\ O & -T_{s+2} & I & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & O \\ O & \ldots & O & -T_{s+m-1} & I \end{pmatrix}.$$

The $\tilde{d}_t$, the element of $\tilde{d} = (\tilde{d}'_{s+1}, \ldots, \tilde{d}'_{s+m})'$, and corresponding block matrices $\tilde{A}_t$, $\tilde{B}_t$ of $\tilde{Q}$ are

$$\begin{aligned}
\tilde{d}_t &= \hat{d}_t - V_t\hat{\eta}_t + T'_{t+1}V_{t+1}\hat{\eta}_{t+1}, && (38)\\
\tilde{A}_t &= \hat{A}_t + V_t + T'_tV_{t+1}T_t, && (39)\\
\tilde{B}_t &= \hat{B}_t - V_tT_t. && (40)
\end{aligned}$$

We replace $\hat{d}, \hat{Q}$ by $\tilde{d}, \tilde{Q}$ in Appendix A2, and we obtain

$$\tilde{d}'(\alpha-\hat{\alpha})-\frac{1}{2}(\alpha-\hat{\alpha})'\tilde{Q}(\alpha-\hat{\alpha}) = \text{const}-\frac{1}{2}\sum_t(\hat{y}_t-\gamma_t)'D_t(\hat{y}_t-\gamma_t).$$

To find a mode of η, we set $\hat{\gamma}_t = \hat{y}_t$ and solve in terms of ($\hat{\alpha}_t$ and) $\hat{\eta}_t$ recursively. Similarly, to sample η, generate $\gamma_t \sim N(\hat{y}, D_t^{-1})$ and obtain η_t recursively.

Algorithm 4.1:

1. Initialize $\hat{\eta}$ and compute $\hat{\alpha}$ at $\eta = \hat{\eta}$ using (6) recursively.

2. Evaluate $\tilde{d}_t$'s, $\tilde{A}_t$'s, and $\tilde{B}_t$'s in (38)–(40) at $\alpha = \hat{\alpha}$.

3. Compute the following D_t, J_t and b_t for $t = s+2, \ldots, s+m$ recursively.

$$\begin{aligned} D_t &= \tilde{A}_t - \tilde{B}_t D_{t-1}^{-1} \tilde{B}_t', \quad D_{s+1} = \tilde{A}_{s+1}, \\ J_t &= K_{t-1}^{-1\prime} \tilde{B}_t, \quad J_{s+1} = O, \quad J_{s+m+1} = O, \\ b_t &= \tilde{d}_t - J_t K_{t-1}^{-1} b_{t-1}, \quad b_{s+1} = \tilde{d}_{s+1}, \end{aligned}$$

where K_t denotes a Choleski decomposition of D_t such that $D_t = K_t K_t'$.

4. Define auxiliary variables $\hat{y}_t = \hat{\gamma}_t + D_t^{-1} b_t$ where

$$\hat{\gamma}_t = \hat{\alpha}_t + K_t^{\prime -1} J_{t+1}' \hat{\alpha}_{t+1}, \quad t = s+1, \ldots, s+m-1,$$

5. Update the posterior mode recursively

$$\hat{\alpha}_t = \hat{y}_t - K_t^{\prime -1} J_{t+1} \hat{\alpha}_{t+1}, \quad t = s+m, s+m-1, \ldots, 1,$$

6. Goto 2.

Algorithm 4.2: (Simulation smoother):

1. Given the current value η_x, find the mode $\hat{\alpha}$ using Algorithm 4.1.

2. Proceed Step 2–4 of Algorithm 4.1.

3. Propose a candidate η_y by sampling from $q(\eta_y) \propto \min(f(\eta_y), cf^*(\eta_y))$ using the Acceptance-Rejection algorithm where the logarithm of c can be constructed from a constant term and $\hat{L}$ in (11).

   (i) Generate $\eta_y \sim f^*$ by sampling $\gamma_t \sim N(\hat{y}_t, D_t^{-1})$ and compute

   $$\alpha_t = \gamma_t - K_t^{\prime -1} J_{t+1} \alpha_{t+1}, \quad t = s+m, s+m-1, \ldots, 1,$$

   recursively and obtain α_y, η_y.

   (ii) Accept η_y with probability

   $$\frac{\min(f(\eta_y), cf^*(\eta_y))}{cf^*(\eta_y)}.$$

   If it is rejected, go back to (i).

4. Conduct the MH algorithm using the candidate η_y. Given the current value η_x, we accept η_y with probability

$$\min\left\{1, \frac{f(\eta_y)\min(f(\eta_x), cf^*(\eta_x))}{f(\eta_x)\min(f(\eta_y), cf^*(\eta_y))}\right\}.$$

where a proposal density proportional to $\min(f(\eta_y), cf^*(\eta_y))$. If it is rejected, accept η_x as a sample.

References

Anderson, B. D. D. and Moore, J. B. (1979), *Optimal filtering.* Englewood Cliffs, NJ:Prentice-Hall.

Black, F. (1976), "Studies of stock market volatility changes," *1976 Proceedings of the American Statistical Association, Business and Economic Statistics Section,* 177-181.

Carlin, B. P., Nicholas, G. P. and Stoffer, D. S. (1992), "A Monte Carlo approach to non-normal and nonlinear state-space modeling," *Journal of the American Statistical Association,* 87, 493-500.

Carter, C. K. and Kohn, R. (1994), "On Gibbs sampling for state space models," *Biometrika*, 81, 541-553.

Chib, S. (2001), "Markov Chain Monte Carlo Methods: Computation and Inference," in J.J. Heckman and E. Leamer (eds.), *Handbook of Econometrics*, 5. North Holland, Amsterdam, 3569-3649.

Chib, S. and Greenberg, E. (1995), "Understanding the Metropolis-Hastings algorithm," *The American Statistician,* 49, 327-335.

de Jong, P. (1988), "A cross-validation filter for times series models, *Biometrika*, 75, 594-600.

de Jong, P. (1989), "Smoothing and interpolation with the state-space model", *Journal of the American Statistical Association,* 84 , 1085-1088.

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

Doornik, J.A. (2002). *Object-oriented matrix programming using Ox, 3rd ed.* London: Timberlake Consultants Press and Oxford.

Durbin, J. and Koopman, S. J. (2000), "Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives," *Journal of the Royal Statistical Society, Ser. B,* 62, 1-27.

Durbin, J. and Koopman, S. J. (2002), "A simple and efficient simulation smoother for state space time series analysis," *Biometrika,* 89, 603-615.

Fahrmeir, L. (1992), "Posterior mode estimation by extended Kalman filtering for multivariate dynamic generalized linear models," *Journal of the American Statistical Association,* 87, 501-509.

Fahrmeir, L. and Wagenpfeil, S. (1997), "Penalized likelihood estimation and iterative Kalman filtering for non-Gaussian dynamic regression models," *Computational Statistics and Data Analysis,* 24, 295-320.

Frühwirth-Schnatter, S. (1994), "Data augmentation and dynamic linear models," *Journal of Time Series Analysis,* 15, 183-202.

Frühwirth-Schnatter, S. and Geyer, A. L. J. (1998), "Bayesian estimation of econometric multi-factor Cox-Ingersoll-Ross models of the term structure of interest rates via MCMC methods," Working paper, Department of Statistics, Vienna University of Economics and Business Administration.

Gamerman, D. (1998), "Markov chain Monte Carlo for dynamic generalised linear models," *Biometrika,* 85, 215-227.

Geweke, J. (1992), "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments," in J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds.), *Bayesian Statistics*, 4, New York: Oxford University Press, 169-188.

Harvey, A. C. and Shephard, N. (1996), "Estimation of an asymmetric stochastic volatility model for asset returns," *Journal of Business and Economic Statistics,* 14, 429-434.

Hürzeler, M. and Künsch, H. R.(1998), "Monte Carlo approximations for general state-space models," *Journal of Computational and Graphical Statistics,* 7, 175-193.

Jacquier, E., Polson, N. G. and Rossi, P. E. (2004), "Bayesian analysis of stochastic volatility with fat-tails and correlated errors," *Journal of Econometrics*, 122, 185-212.

Kim, S., Shephard, N. and S. Chib (1998), "Stochastic volatility: likelihood inference and comparison with ARCH models," *Review of Economic Studies,* 65, 361-393.

Kitagawa, G. (1987), "Non-Gaussian state-space modeling of nonstationary time series," (with discussion) *Journal of the American Statistical Association,* 82, 1032-1063.

Kitagawa, G. (1996), "Monte Carlo filter and smoother for non-Gaussian nonlinear state space models," *Journal of Computational and Graphical Statistics,* 5, 1-25.

Kitagawa, G. (1998), "A self-organizing state-space model," *Journal of the American Statistical Association,* 93, 1203-1215.

Kohn, R. and Ansley, C. F. (1989), "A fast algorithm for signal extraction, influence and cross-validation in state space models," *Biometrika,* 76, 65-79.

Kim, S., Shephard, N. and Chib, S. (1998), "Stochastic volatility: likelihood inference and comparison with ARCH models," *Review of Economic Studies,* 65, 361-394.

Nelson, D. B. (1991), "Conditional heteroskedasticity in asset returns: a new approach," *Econometrica,* 62, 1-41.

Omori, Y., Chib, S., Shephard, N. and Nakajima, J. (2006), "Stochastic volatility model with leverage: fast likelihood inference," to appear in *Journal of Econometrics.*

Sage, A. P. and Melsa, J. L. (1971), *Estimation theory with applications to communications and control.* New York;McGraw-Hill.

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

Sanford, A.D. and Martin, G. M. (2003), "Simulation-based Bayesian estimation of Affine term structure models," Working paper, Department of Econometrics and Business Statistics, Monash University.

Stroud, J. R., Müller, P. and Polson, N. G. (2003), "Non-Gaussian state-space models with state-dependent variances," *Journal of the American Statistical Association,* 98, 377-386.

Tanizaki, H. (2001), "Nonlinear and non-Gaussian state space modeling using sampling techniques," *Annals of the Institute of Statistical Mathematics,* 53, 63-81.

Tanizaki, H. and Mariano (1998), "Nonlinear and non-Gaussian state space modeling using with Monte Carlo simulations," *Journal of Econometrics,* 83, 263-290.

Tierney, L. (1994), "Markov chains for exploring posterior distributions," *Annals of Statistics*, **22**, 1701–1728.

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

Watanabe, T., T. Yamada, and K. Tanaka (2006), "MCMC-based Bayesian analysis of Cox-Ingersoll-Ross models of the term structure of interest rates," mimeo.

Yu, J. (2005), "On a leverage in a stochastic volatility model," *Journal of Econometrics,* 127, 165-178.