Block sampler and posterior mode estimation for asymmetric stochastic volatility models

Yasuhiro Omori[a, *] Toshiaki Watanabe[b]

[a] Faculty of Economics, University of Tokyo, Tokyo 113-0033, Japan.
[b] Institute of Economic Research, Hitotsubashi University, Tokyo 186-8603, Japan.

Abstract

A new efficient simulation smoother and disturbance smoother are introduced for asymmetric stochastic volatility models where there exists a correlation between today's return and tomorrow's volatility. 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) simple asymmetric stochastic volatility model and (2) asymmetric stochastic volatility model with 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: Asymmetric stochastic volatility model; Bayesian analysis; Disturbance smoother; Kalman filter; Markov chain Monte Carlo; Metropolis-Hastings algorithm; Simulation smoother.

1 Introduction

It is well known in financial markets that return volatility changes randomly with a high persistence. It has also long been recognized in stock markets that there is a negative correlation between today's return and tomorrow's volatility (Black (1976) and Christie (1982)). This phenomenon is called "leverage effect" or "asymmetry". We use the term "asymmetry" in this artcile since some researchers show that this phenomenon cannot be attributed to financial leverage (Avramov *et al.* (2006)).

*Corresponding author: Tel:+81-3-5841-5516, Fax:+81-3-5841-5521, E-maill:omori@e.u-tokyo.ac.jp.

The asymmetric stochastic volatility model is well-known to describe these phenomena for stock returns (alternative models are, e.g., GJR (Glosten *et al.* (1993)), EGARCH (Nelson (1991)) and APGARCH (Ding *et al.* (1993)) models). This article proposes an efficient Bayesian method using Markov chain Monte Carlo (MCMC) for the estimation of asymmetric stochastic volatility models. In the previous literature, simple estimation procedures are proposed. For example, Melino and Turnbull (1990) use the GMM (generalized methods of moments) and Harvey and Shephard (1996) use the QML (quasi-maximum likelihood method) via the Kalman filter for the estimation of asymmetric stochastic volatility models. However, they are less efficient than the MCMC-based Bayesian method (Jacquier *et al.* (1994)).

This method requires us to sample state variables as well as parameters from their joint posterior distribution, which is possible by using Gibbs sampler, i.e., sampling them from their full conditional distributions iteratively. The most important is how to sample the state variables from their full conditional distribution. A simple method is the single-move sampler that generates a single state variable at a time given the rest of the state variables and other parameters. It is usually easy to construct such a sampler, but the obtained samples are known to be highly autocorrelated. This implies that we need to generate a huge number of samples to conduct a statistical inference and hence the sampler is inefficient.

Two methods have been proposed to reduce sample autocorrelations effectively. One method is mixture samplers proposed by Kim *et al.* (1998) for a symmetric stochastic volatility model and extended by Omori *et al.* (2007) for an asymmetric stochastic volatility model. This method transforms the model into a linear state-space model and approximates the error distribution by a mixture of normal distributions (see e.g. Frühwirth-Schnatter and Frühwirth (2007) for the similar approximation in the logistic models). The mixture samplers is fast and highly efficient, but instead its use is limited to the models that can be transformed into a linear state-space form. For example, it is not applicable to the stochastic volatility model with risk premium because it cannot be represented by a linear state-space model.

The other methods are block samplers (also called multi-move samplers) proposed by Shephard and Pitt (1997) and Watanabe and Omori (2004), which generate a block of state variables. This method can be applied to the model directly without transforming into a linear state-space form. However, the block samplers proposed by Shephard and Pitt (1997) and Watanabe and Omori (2004) assume that an observation vector and a state vector are conditionally independent. Thus they cannot be applied to asymmetric stochastic volatility models.

In this article, we develop a block sampler for asymmetric stochastic volatility 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.

Since our method can be applied to more general models, we also apply our method to an extended asymmetric stochastic volatility model where the variance of the disturbance in the volatility equation is state-dependent. Stroud *et al.* (2003) considered a block sampler for models with state-dependent variances (but without asymmetry) using an auxiliary mixture model to generate a state proposal for Metropolis-Hastings algorithms. The block samplers proposed by Shephard and Pitt (1997) and Watanabe and Omori (2004) cannot be applied to such a model because they assume that a state equation is linear. For this model, 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.

We compare the performance of our method with the single move sampler which samples a state variable at a time using a simple asymmetric stochastic volatility model. We find that our proposed sampler produces considerable improvement in the mixing property of the Markov chain Monte Carlo chain. We also estimate the asymmetric stochastic volatility model with state-dependent variances using stock returns data.

The organization of the article is as follows. In Section 2, we introduce a simple asymmetric stochastic volatility model. Section 3 describes a simulation smoother and a disturbance smoother for this model. Section 4 extends our method for an asymmetric volatility model with state-dependent variance. In Section 5, we illustrate our method using simulated data and stock returns data. Section 6 concludes the article.

2 Asymmetric stochastic volatility model

In this article, we first consider the following asymmetric stochastic volatility model.

$$\begin{aligned}
y_t &= \epsilon_t \sigma_\epsilon \exp(\alpha_t/2), \quad t = 1, \dots, n, & (1)\\
\alpha_{t+1} &= \phi\alpha_t + \eta_t\sigma_\eta, \quad |\phi| < 1, \quad t = 1, \dots, n-1, & (2)\\
\alpha_1 &\sim N(0, \sigma_\eta^2/(1-\phi^2)), & (3)\\
\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} &\sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right), & (4)
\end{aligned}$$

where α_t is the unobserved state variable, $\sigma_\epsilon \exp(\alpha_t/2)$ stands for the volatility of the response, y_t, and $(\rho, \sigma_\epsilon, \sigma_\eta, \phi)$ are parameters. We assume $|\phi| < 1$ for the stationarity of α_t. The state equation (2) is linear and Gaussian, while the measurement equation (1) is nonlinear. (4) assumes that error terms ϵ_t and η_t follow a bivariate normal distribution. A correlation between these errors is considered to explain asymmetry. The symmetric stochastic volatility model ($\rho = 0$) 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 in stock markets that the fall of the stock return is followed by the high volatility (Black (1976) and Christie (1982)). Thus we expect a negative correlation, $\rho < 0$, between ϵ_t and η_t rather than $\rho = 0$ in stock markets.

Jacquier *et al.* (2004) considered a correlation between ϵ_t and η_{t-1}. Harvey and Shephard (1996) and Yu (2005) point out that y_t is a martingale difference sequence if ϵ_t and η_t are correlated whereas it is not so and inconsistent with the efficient market hypothesis if ϵ_t and η_{t-1} are correlated. Moreover, Yu (2005) shows that the model with the correlation between ϵ_t and η_t fits the data better than that with the correlation between ϵ_t and η_{t-1}.

3 Block sampler and posterior mode estimation

In our block sampler, we divide $(\alpha_1, \ldots, \alpha_n)$ into $K+1$ blocks, $(\alpha_{k_{i-1}+1}, \ldots, \alpha_{k_i})'$ for $i = 1, \ldots, K+1$, with $k_0 = 0$ and $k_{K+1} = n$, where $k_i - k_{i-1} \geq 2$. Following Shephard and Pitt (1997), we select K knots, $(k_1, \ldots, k_K)$, randomly (see Section 5.1.1 for the detail). We sample the error term $(\eta_{k_{i-1}}, \ldots, \eta_{k_i-1})$ instead of $(\alpha_{k_{i-1}+1}, \ldots, \alpha_{k_i})$ simultaneously from their full conditional distribution.

Suppose that $k_{i-1} = s$ and $k_i = s + m$ for the i-th block. Then $(\eta_s, \ldots, \eta_{s+m-1})$ are sampled simultaneously from the following full conditional distribution.

$$\begin{aligned} & f(\eta_s, \ldots, \eta_{s+m-1} | \alpha_s, \alpha_{s+m+1}, y_s, \ldots, y_{s+m}) \\ & \quad \propto \prod_{t=s}^{s+m} f(y_t|\alpha_t, \alpha_{t+1}) \prod_{t=s}^{s+m-1} f(\eta_t), \quad s+m < n, \end{aligned} \tag{5}$$

$$\begin{aligned} & f(\eta_s, \ldots, \eta_{s+m-1} | \alpha_s, y_s, \ldots, y_{s+m}) \\ & \quad \propto \prod_{t=s}^{s+m-1} f(y_t|\alpha_t, \alpha_{t+1}) f(y_n|\alpha_n) \prod_{t=s}^{s+m-1} f(\eta_t), \quad s+m = n. \end{aligned} \tag{6}$$

The conditional distribution of y_t given α_t and α_{t+1} for $t < n$ and that given α_t for $t = n$ are normal with mean μ_t and variance σ_t^2 where

$$\mu_t = \begin{cases} \rho\sigma_\epsilon\sigma_\eta^{-1}(\alpha_{t+1} - \phi\alpha_t)\exp(\alpha_t/2), & t < n, \\ 0, & t = n, \end{cases} \tag{7}$$

$$\sigma_t^2 = \begin{cases} (1-\rho^2)\sigma_\epsilon^2\exp(\alpha_t), & t < n, \\ \sigma_\epsilon^2\exp(\alpha_n), & t = n. \end{cases} \tag{8}$$

The logarithm of $f(y_t|\alpha_t, \alpha_{t+1})$ or $f(y_n|\alpha_n)$ in equations (5) and (6) (excluding constant term) is given by

$$l_t = -\frac{\alpha_t}{2} - \frac{(y_t - \mu_t)^2}{2\sigma_t^2}. \tag{9}$$

Then the logarithm of (5) or (6) is $-\sum_{t=s}^{s+m-1} \eta_t^2/2 + L$ (excluding a constant term) where

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

Further define

$$\begin{aligned} d &= (d'_{s+1}, \dots, d'_{s+m})', \quad d_t = \frac{\partial L}{\partial \alpha_t}, \quad t = s+1, \dots, s+m, & (10) \\ Q &= -E\left[\frac{\partial^2 L}{\partial\alpha\partial\alpha'}\right] = \begin{pmatrix} A_{s+1} & B'_{s+2} & O & \dots & O \\ B_{s+2} & A_{s+2} & B'_{s+3} & \dots & O \\ O & B_{s+3} & A_{s+3} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & B'_{s+m} \\ O & \dots & O & B_{s+m} & A_{s+m} \end{pmatrix}, & \\ A_t &= -E\left[\frac{\partial^2 L}{\partial\alpha_t\partial\alpha'_t}\right], \quad t = s+1, \dots, s+m, & (11) \\ B_t &= -E\left[\frac{\partial^2 L}{\partial\alpha_t\partial\alpha'_{t-1}}\right], \quad t = s+2, \dots, s+m, \quad B_{s+1} = O, & (12) \end{aligned}$$

where $s \geq 0$ ($t \geq 1$ for (11) and $t \geq 2$ for (12)). As for the asymmetric stochastic volatility model, the first derivative of L with respect to α_t is given by

$$d_t = \frac{\partial L}{\partial \alpha_t} = \begin{cases} -\frac{1}{2} + \frac{(y_t - \mu_t)^2}{2\sigma_t^2} + \frac{(y_t - \mu_t)}{\sigma_t^2}\frac{\partial \mu_t}{\partial \alpha_t} + \frac{(y_{t-1} - \mu_{t-1})}{\sigma_{t-1}^2}\frac{\partial \mu_{t-1}}{\partial \alpha_t}, \\ \qquad t = s+1, \dots, s+m-1, \quad \text{or} \quad t = s+m = n, \\ -\frac{1}{2} + \frac{(y_t - \mu_t)^2}{2\sigma_t^2} + \frac{(y_t - \mu_t)}{\sigma_t^2}\frac{\partial \mu_t}{\partial \alpha_t} + \frac{(y_{t-1} - \mu_{t-1})}{\sigma_{t-1}^2}\frac{\partial \mu_{t-1}}{\partial \alpha_t} + \frac{\phi(\alpha_{t+1} - \phi\alpha_t)}{\sigma_\eta^2}, \\ \qquad t = s+m < n, \end{cases} \tag{13}$$

where

$$\frac{\partial \mu_t}{\partial \alpha_t} = \begin{cases} \frac{\rho\sigma_\epsilon}{\sigma_\eta}\left\{-\phi + \frac{(\alpha_{t+1} - \phi\alpha_t)}{2}\right\}\exp\left(\frac{\alpha_t}{2}\right), & t = 1, \dots, n-1, \\ 0, & t = n, \end{cases} \tag{14}$$

$$
\frac{\partial \mu_{t-1}}{\partial \alpha_t} = \begin{cases} 0, & t=1, \\ \dfrac{\rho\sigma_\epsilon}{\sigma_\eta}\exp\left(\dfrac{\alpha_{t-1}}{2}\right), & t=2,\ldots,n. \end{cases} \tag{15}
$$

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.

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

Applying the second order Taylor expansion to (5) will produce the approximating normal density $f^*(\eta_s,\ldots,\eta_{s+m-1}|\alpha_s,\alpha_{s+m+1},y_s,\ldots,y_{s+m})$ as follows (see Appendix A1).

$$
\begin{aligned}
&\log f(\eta_s,\ldots,\eta_{s+m-1}|\alpha_s,\alpha_{s+m+1},y_s,\ldots,y_{s+m}) \\
\approx\ & \text{const} - \frac{1}{2}\sum_{t=s}^{s+m-1}\eta_t'\eta_t + \hat{L} + \left.\frac{\partial L}{\partial \eta'}\right|_{\eta=\hat{\eta}}(\eta-\hat{\eta}) + \frac{1}{2}(\eta-\hat{\eta})'\, E\left.\left(\frac{\partial L^2}{\partial\eta\partial\eta'}\right)\right|_{\eta=\hat{\eta}}(\eta-\hat{\eta}) \\
=\ & \text{const} - \frac{1}{2}\sum_{t=s}^{s+m-1}\eta_t'\eta_t + \hat{L} + \hat{d}'(\alpha-\hat{\alpha}) - \frac{1}{2}(\alpha-\hat{\alpha})'\hat{Q}(\alpha-\hat{\alpha}) && (16) \\
=\ & \text{const} + \log f^*(\eta_s,\ldots,\eta_{s+m-1}|\alpha_s,\alpha_{s+m+1},y_s,\ldots,y_{s+m}) && (17)
\end{aligned}
$$

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. Similarly, we can obtain the normal density which approximates (6).

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 the posterior mode.

Algorithm 1.1 (Posterior mode disturbance smoother):

1. Initialize $\hat{\eta}$ and compute $\hat{\alpha}$ at $\eta=\hat{\eta}$ using (2) recursively.

2. Evaluate $\hat{d}_t$'s, $\hat{A}_t$'s, and $\hat{B}_t$'s using (10)–(12) 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

$$\hat{y}_t = Z_t \alpha_t + G_t \xi_t, \quad t = s+1, \ldots, s+m, \tag{18}$$

$$\alpha_{t+1} = \phi \alpha_t + H_t \xi_t, \quad t = s, s+1, \ldots, s+m, \tag{19}$$

$$\xi_t = (\epsilon_t', \eta_t')' \sim N(0, I),$$

   where

$$Z_t = I + K_t^{\prime -1} J_{t+1}' \phi, \quad G_t = K_t^{\prime -1} [I, J_{t+1}' \sigma_\eta], \quad H_t = [O, \sigma_\eta].$$

   Apply Kalman filter and a disturbance smoother (e.g. Koopman (1993)) to the linear Gaussian system (18) and (19) 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 (18) and (19) is the same as f^* in (17). Thus, applying Kalman filter and a disturbance smoother to the linear Gaussian system (18) and (19), we first obtain a smoothed estimate of η_t and then substitute it recursively to the linear state equation (2) 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.

Fahrmeir and Wagenpfeil (1997) and Fahrmeir and Tutz (2001) proposed a closely related algorithm for the non-Gaussian dynamic regression models assuming the exponential family distribution for the measurement equations. However, their algorithm assumed the independence between the measurement error ϵ_t and η_t, and hence cannot be applied to the asymmetric stochastic volatility models. Our algorithm can be applied to the models with more general distribution family and correlated errors.

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 (18)–(19).

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 (16).

   (i) Generate $\eta_y \sim f^*$ using the multimove simulation smoother (e.g. de Jong and Shephard (1995), Durbin and Koopman (2002)) for the approximating linear Gaussian state-space model (18)–(19).

   (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 (18) and (19) reduce to

$$\begin{aligned}
\hat{y}_t &= \alpha_t + K_t^{'-1}\epsilon_t, \quad \epsilon_t \sim N(0, I), \quad t = s+1, \ldots, s+m, \\
\alpha_{t+1} &= \phi\alpha_t + \sigma_\eta\eta_t, \quad \eta_t \sim N(0, I). \quad t = s, s+1, \ldots, s+m,
\end{aligned}$$

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

4 Extension

It is straightforward to extend our method for more general models. Thus, we also consider an asymmetric stochastic volatility models with state-dependent variances.

Stroud *et al.* (2003) considered state-dependent variance models (but without asymmetry) 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 the asymmetric stochastic volatility model. Specifically, we replace state equations (2) and (3) by

$$\alpha_{t+1} = \phi\alpha_t + \eta_t\sigma_\eta \left\{1 + \frac{1}{1+\exp(-\alpha_t)}\right\}, \quad |\phi| < 1, \quad t = 1, \ldots, n-1, \tag{20}$$

$$\alpha_1 \sim N(0, \sigma_0^2), \quad (\sigma_0^2\text{: known}). \tag{21}$$

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.

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

$$\beta_{t+1} = \hat{T}_t\beta_t + \hat{R}_t\eta_t, \quad \eta_t \sim N(0, I), \tag{22}$$

$$\hat{T}_t = \left.\frac{\partial\alpha_{t+1}}{\partial\alpha_t'}\right|_{\eta=\hat{\eta}}, \quad \hat{R}_t = \left.\frac{\partial\alpha_{t+1}}{\partial\eta_t'}\right|_{\eta=\hat{\eta}}, \tag{23}$$

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.

As for the state equation (20), $\hat{T}_t$ and $\hat{R}_t$ in the auxiliary state equation (22) 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

$$\mu_t = \rho\sigma_\epsilon\sigma_\eta^{-1}(\alpha_{t+1} - \phi\alpha_t)\left\{1 + \frac{1}{1+\exp(-\alpha_t)}\right\}^{-1}\exp(\alpha_t/2), \tag{24}$$

and σ_t^2 given by (8). The logarithm of conditional likelihood of y_t (excluding constant term) is the same as in (9).

To sample a block $(\alpha_{s+1}, \dots, \alpha_{s+m})$ given α_s, α_{s+m+1} and other parameters, we consider the log conditional posterior for η_t $(t = s, s+1, \dots, s+m-1)$ 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 (13) but replacing (14) and (15) by

$$\frac{\partial\mu_t}{\partial\alpha_t} = \begin{cases} \frac{\partial\mu_t}{\partial\alpha_{t+1}}\left[-\phi + (\alpha_{t+1} - \phi\alpha_t)\left\{\frac{1}{2} - \frac{1}{3 + 2\exp(\alpha_t) + \exp(-\alpha_t)}\right\}\right], \\ \qquad t = 1, \dots, n-1, \\ 0, \qquad t = n, \end{cases} \tag{25}$$

$$\frac{\partial\mu_{t-1}}{\partial\alpha_t} = \begin{cases} 0, & t = 1, \\ \frac{\rho\sigma_\epsilon}{\sigma_\eta}\left\{1 + \frac{1}{1+\exp(-\alpha_{t-1})}\right\}^{-1}\exp\left(\frac{\alpha_t}{2}\right), & t = 2, \dots, n, \end{cases} \tag{26}$$

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

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

Let $L = \sum_{t=s}^{s+m} l_t$ and $\eta = (\eta_s', \dots, \eta_{s+m-1}')'$. Then

$$\begin{aligned} &\log f(\eta|\alpha_s, \alpha_{s+m+1}, y_s, \dots, y_{s+m}) \\ &= \text{const} - \frac{1}{2}\sum_{t=s}^{s+m-1}\eta_t'\eta_t + L + \log p(\alpha_{s+m+1}|\alpha_{s+m}) \\ &\approx \text{const} - \frac{1}{2}\sum_{t=s}^{s+m-1}\eta_t'\eta_t + \hat{L} + \hat{d}'(\beta - \hat{\beta}) - \frac{1}{2}(\beta - \hat{\beta})'\hat{Q}(\beta - \hat{\beta}) + \log p(\alpha_{s+m+1}|\hat{\alpha}_{s+m}) \\ &= \text{const} + \log f^*(\eta|\alpha_s, \alpha_{s+m+1}, y_s, \dots, y_{s+m}) + \log p(\alpha_{s+m+1}|\hat{\alpha}_{s+m}), \end{aligned} \tag{27}$$

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.

Posterior 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 (23) at $\eta = \hat{\eta}$ and compute $\hat{\alpha}_t$'s and $\hat{\beta}_t$'s recursively.

$$
\begin{aligned}
\hat{\alpha}_{t+1} &= \phi\hat{\alpha}_t + \hat{\eta}_t\sigma_\eta\left\{1 + \frac{1}{1+\exp(-\hat{\alpha}_t)}\right\}, \\
\hat{\beta}_{t+1} &= \hat{T}_t\hat{\beta}_t + \hat{R}_t\hat{\eta}_t,
\end{aligned}
$$

   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 (10)–(12) 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

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

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

$$
\begin{aligned}
\hat{y}_t &= Z_t\beta_t + G_t\xi_t, \quad t = s+1, \ldots, s+m, && (28) \\
\beta_{t+1} &= \hat{T}_t\beta_t + H_t\xi_t, \quad t = s, s+1, \ldots, s+m-1, && (29) \\
& \xi_t = (\epsilon_t', \eta_t')' \sim N(0, I),
\end{aligned}
$$

   where

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

Apply Kalman filter and a disturbance smoother to the linear Gaussian system (28) and (29) 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

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

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 (28)–(29).

3. Generate a candidate η_y from $f^*(\eta_y)$ using a simulation smoother for the approximating linear Gaussian state-space model (28)–(29). 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.

5 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 $\alpha_{-t} = (\alpha_1, \ldots, \alpha_{t-1}, \alpha_{t+1}, \ldots, \alpha_n)$).

5.1 Asymmetric stochastic volatility model

5.1.1 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 $\{\alpha_t\}_{t=k_{i-1}+1}^{k_i}|\{\alpha_t|t \le k_{i-1}, t > k_i\}, \phi, \Sigma, y$ 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 $k_0 = 0$ and $k_{K+1} = n$ where $k_i - k_{i-1} \ge 2$ for $i = 1, \ldots, K+1$. The K knots, $(k_1, \ldots, k_K)$, are generated randomly using

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

where U_i's are independent uniform random variables on $(0, 1)$ (see e.g. Shephard and Pitt (1997), Watanabe and Omori (2004)) . As discussed in Shephard and Pitt (1997), these stochastic knots have advantages to allow the points of conditioning to change over the MCMC iterations and are expected to accelerate the convergence of the distribution of MCMC samples to the posterior distibution. We control the single tuning parameter K to obtain the efficient sampler. 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

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

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

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

5.1.2 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.

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.

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).

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.

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 asymmetry 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 proposed sampler reduces sample autocorrelations considerably and that it produces more accurate estimation results than the single move sampler.

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).

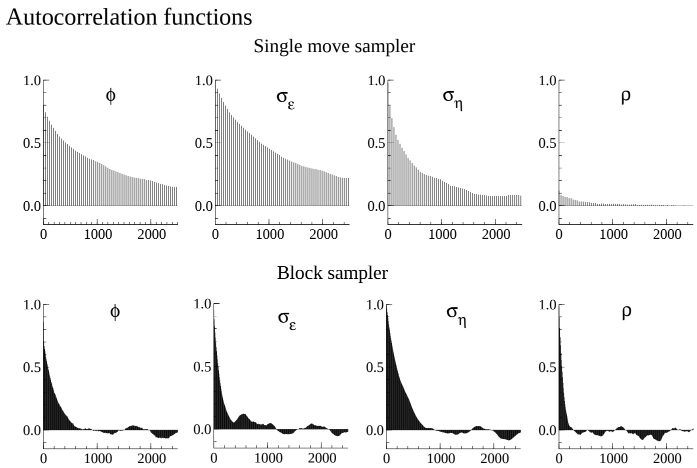

Figure 1: Sample autocorrelation functions of MCMC samples.

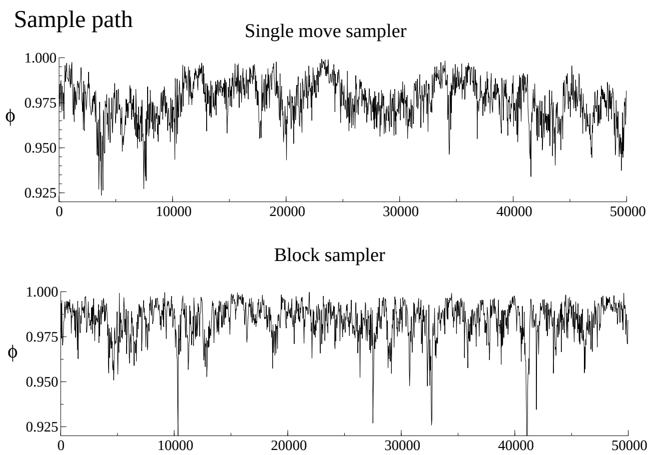

Figure 2: Sample path of ϕ's using first 50,000 MCMC samples.

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 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.

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.

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.

5.1.3 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.

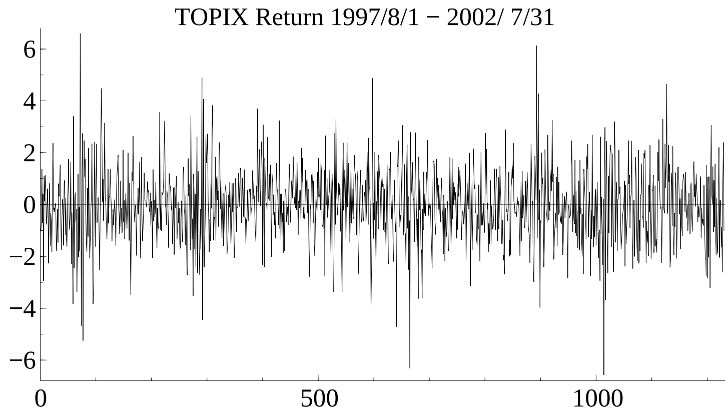

Figure 3: TOPIX return data. 1997/8/1 – 2002/7/31.

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 asymmetry 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.

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.

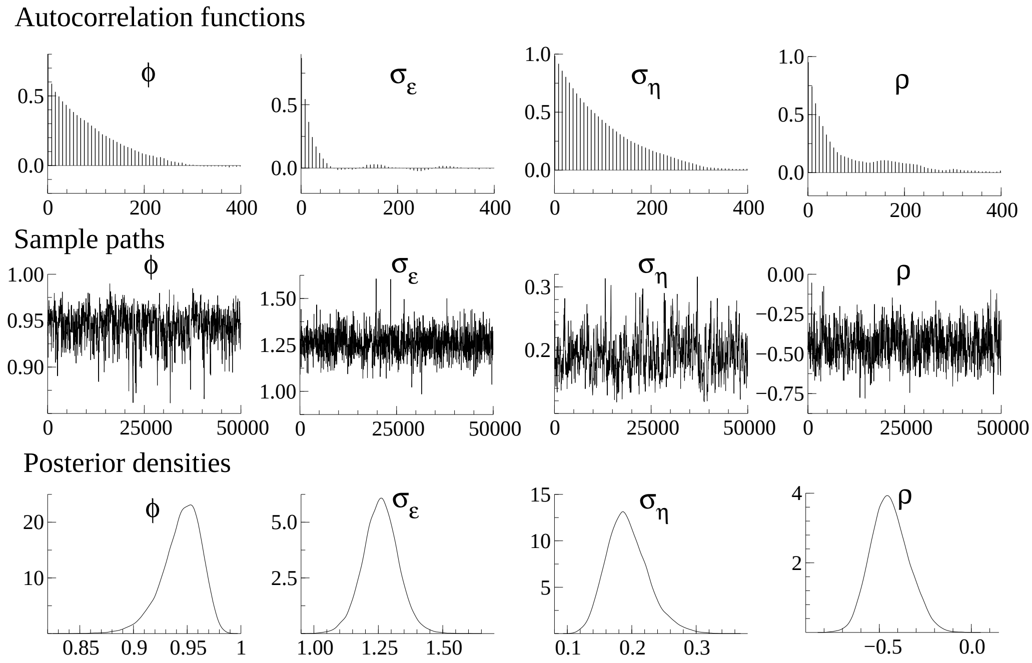

Figure 4: Sample autocorrelation functions of MCMC samples.

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.

5.2 Asymmetric stochastic volatility model with state-dependent variances

This subsection illustrates our method using simulated data generated by the stochastic volatility model in (20) and (21). 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. 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.

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.

Table 8 shows 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.

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.

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 (27). The appropriate number of blocks for this particular example would be from 20 to 40.

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.

6 Conclusion

In this article, we described a disturbance smoother and a simulation smoother for asymmetric stochastic volatility models. The high performance of our proposed method is shown in estimation efficiencies using illustrative numerical examples in comparison with a single move sampler.

As mentioned in Section 1, Melino and Turnbull (1990) use the GMM and Harvey and Shephard (1996) use the QML via the Kalman filter for the estimation of asymmetric stochastic volatility models. However, they are less efficient than the MCMC-based Bayesian method. Bartolucci and De Luca (2003) propose the maximum likelihood estimation via the quadrature method for the estimation of asymmetric stochastic volatility models, and Celeux *et al.* (2006) propose the population Monte Carlo scheme. It would be interesting to compare our method with these methods regarding the estimation efficiencies.

Although we concentrated on asymmetric stochastic volatility models in this article, our method can be applied to general non-Gaussian and nonlinear state-space models with correlated errors (Omori and Watanabe (2007)). It is also worthwile applying our method to other models. Watanabe, Yamada and Tanaka (2007) have applied our method to the Cox-Ingersoll-Ross model of the term structure of interest rates.

Acknowledgement

The authors thank Herman K. van Dijk, Hisashi Tanizaki, Erricos John Kontoghiorghes and three anonymous referees for their helpful discussions and comments. 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

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

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

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

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

$$\beta_t = \sum_{j=s}^{t-1} \left. \frac{\partial \alpha_t}{\partial \eta_j'} \right|_{\eta=\hat{\eta}} \eta_j + \hat{T}_{t-1} \cdots \hat{T}_s \beta_s,$$

we obtain

$$\begin{aligned} \left. \frac{\partial L}{\partial \eta} \right|_{\eta=\hat{\eta}}' (\eta - \hat{\eta}) &= \sum_{j=s}^{s+m-1} \sum_{t=j+1}^{s+m} \left. \frac{\partial L}{\partial \alpha_t'} \right|_{\alpha=\hat{\alpha}} \left. \frac{\partial \alpha_t}{\partial \eta_j'} \right|_{\eta=\hat{\eta}} (\eta_j - \hat{\eta}_j) \\ &= \sum_{t=s+1}^{s+m} \hat{d}_t' (\beta_t - \hat{\beta}_t) = \hat{d}' (\beta - \hat{\beta}). \end{aligned} \tag{30}$$

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

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

and its expected value is

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

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

$$\begin{aligned}
&(\eta-\hat{\eta})'\, E\left(\frac{\partial L^2}{\partial \eta \partial \eta'}\right)\Bigg|_{\eta=\hat{\eta}} (\eta-\hat{\eta}) \\
&= \sum_{t_1=s+1}^{s+m}\sum_{t_2=s+1}^{s+m}\sum_{i=s}^{t_1-1}\sum_{j=s}^{t_2-1} (\eta_i-\hat{\eta}_i)' \left.\frac{\partial \alpha_{t_1}}{\partial \eta_i}\right|_{\eta=\hat{\eta}} E\left(\frac{\partial L^2}{\partial \alpha_{t_1}\partial \alpha_{t_2}'}\right)\Bigg|_{\eta=\hat{\eta}} \left.\frac{\partial \alpha_{t_2}}{\partial \eta_j'}\right|_{\eta=\hat{\eta}} (\eta_j-\hat{\eta}_j) \\
&= \sum_{t_1=s+1}^{s+m}\sum_{t_2=s+1}^{s+m} (\beta_{t_1}-\hat{\beta}_{t_1})'\, E\left(\frac{\partial L^2}{\partial \alpha_{t_1}\partial \alpha_{t_2}'}\right)\Bigg|_{\eta=\hat{\eta}} (\beta_{t_2}-\hat{\beta}_{t_2}) \\
&= -(\beta-\hat{\beta})'\hat{Q}(\beta-\hat{\beta}). \qquad (31)
\end{aligned}$$

The results are obtained from equations (30) and (31).

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

$$\begin{aligned}
\hat{A}_t &= J_tJ_t' + K_tK_t', \quad t = s+1,\dots,s+m, \\
\hat{B}_t &= J_tK_{t-1}', \quad t = s+2,\dots,s+m,
\end{aligned}$$

and $B_{s+1} = J_{s+1} = O$. Denote $C_t = J_tJ_t', D_t = K_tK_t'$ and we obtain

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

for $t = s+2,\dots,s+m$, 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 $K = \mathrm{diag}(K_{s+1},\dots,K_{s+m})$, $D = \mathrm{diag}(D_{s+1},\dots,D_{s+m})$,

$b = KU^{-1}\hat{d}$, $\gamma = K'^{-1}U'\beta$, and $\hat{\gamma} = K'^{-1}U'\hat{\beta}$. Then

$$\begin{aligned} \hat{d}'(\beta-\hat{\beta}) - \frac{1}{2}(\beta-\hat{\beta})'\hat{Q}(\beta-\hat{\beta}) &= b'(\gamma-\hat{\gamma}) - \frac{1}{2}(\gamma-\hat{\gamma})'D(\gamma-\hat{\gamma}) \\ &= -\frac{1}{2}(\hat{y}-\gamma)'D(\hat{y}-\gamma) \end{aligned} \tag{32}$$

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 (32) is a likelihood function for

$$\begin{aligned} \hat{y}_t &= \beta_t + K_t'^{-1}J_{t+1}'\beta_{t+1} + K_t'^{-1}\epsilon_t = Z_t\beta_t + G_t\xi_t, \\ & \quad \xi_t = (\epsilon_t', \eta_t')' \sim N(0, I). \end{aligned} \tag{33}$$

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

References

Avramov, D., Chordia, T. and Goyal, A. (2006), "The Impact of Trades on Daily Volatility," *Review of Financial Studies,* 19, 1241-1277.

Bartolucci, F. and De Luca, G. (2003), "Likelihood-based inference for asymmetric stochastic volatility models," *Computational Statistics & Data Analysis,* 42, 445-449.

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

Celeux, G., Marin, J.-M. and Robert, C. P. (2006), "Iterated importance sampling in missing data problems," *Computational Statistics & Data Analysis*, 50, 3386-3404.

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.

Christie, A. A. (1982), "The Stochastic Behavior of Common Stock Variances–Value, Leverage and Interest Rate Effects," *Journal of Financial Economics,* 10, 407–432.

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

Ding, Z., Granger, C. W. J. and Engle, R. F. (1993), "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance,* 1, 83-106.

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

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. 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.

Fahrmeir, L. and Tutz, G. (2001), *Multivariate Statistical Modelling Based on Generalized Linear Models*, 2nd ed. New York: Springer.

Frühwirth-Schnatter, S. and Frühwirth, R. (2007), "Auxiliary mixture sampling with applications to logistic models," *Computational Statistics & Data Analysis,* 51, 3509-3528.

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.

Glosten, L. R., Jagannathan, R. and Runkle, D. E. (1993), "On the Relation between the Expected Value and the Volatility of Nominal Excess Returns on Stocks," *Journal of Finance,* 48, 1779-1801.

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.

Jacquier, E., Polson, N. G. and Rossi, P. E. (1994), "Bayesian analysis of stochastic volatility models (with discussion)," *Journal of Business and Economic Statistics*, 12, 371-417.

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.

Koopman, S. J. (1993), "Disturbance smoother for state space models," *Biometrika,* 80, 117-126.

Melino, A. and Turnbull, S. M. (1990), "Pricing Foreign Currency Options with Stochastic Volatility," *Journal of Econometrics*, 45, 239-265.

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. (2007), "Stochastic volatility model with leverage: fast likelihood inference," *Journal of Econometrics*, 140, 425-449.

Omori, Y. and Watanabe, T. (2007), "Block sampler and posterior mode estimation for a nonlinear and non-Gaussian state-space model with correlated errors," CIRJE Discussion Paper F-508, Center for International Research on the Japanese Economy, University of Tokyo.

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

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.

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., Yamada, T. and Tanaka, K. (2007), "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.