ELSEVIER

Stochastic volatility with leverage: Fast and efficient likelihood inference

Yasuhiro Omori[a],[*], Siddhartha Chib[b], Neil Shephard[c],
Jouchi Nakajima[d]

[a]*Faculty of Economics, University of Tokyo, Tokyo 113-0033, Japan*
[b]*Olin School of Business, Washington University in St. Louis, St. Louis, USA*
[c]*Nuffield College, University of Oxford, Oxford OX1 1NF, UK*
[d]*Graduate School of Economics, University of Tokyo, Tokyo 113-0033, Japan*

Available online 23 August 2006

Abstract

   This paper is concerned with the Bayesian analysis of stochastic volatility (SV) models with leverage. Specifically, the paper shows how the often used Kim et al. [1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361–393] method that was developed for SV models without leverage can be extended to models with leverage. The approach relies on the novel idea of approximating the joint distribution of the outcome and volatility innovations by a suitably constructed ten-component mixture of bivariate normal distributions. The resulting posterior distribution is summarized by MCMC methods and the small approximation error in working with the mixture approximation is corrected by a reweighting procedure. The overall procedure is fast and highly efficient. We illustrate the ideas on daily returns of the Tokyo Stock Price Index. Finally, extensions of the method are described for superposition models (where the log-volatility is made up of a linear combination of heterogenous and independent autoregressions) and heavy-tailed error distributions (student and log-normal).
© 2006 Elsevier B.V. All rights reserved.

Keywords: Leverage effect; Markov chain Monte Carlo; Mixture sampler; Stochastic volatility; Stock returns

*Corresponding author. Fax: +1 314 935 6359.
   E-mail addresses: omori@e.u-tokyo.ac.jp (Y. Omori), chib@wustl.edu (S. Chib),
neil.shephard@economics.ox.ac.uk (N. Shephard), n_sophia@f3.dion.ne.jp (J. Nakajima).

1. Introduction

There is by now a large literature on the fitting of stochastic volatility (SV) models (see, for example, the reviews in Ghysels et al. (1996) and Shephard (2005)). Particularly well studied are SV models without leverage for which the Bayesian approach of Kim et al. (1998) (KSC) has become well established (for example, Mahieu and Schotman, 1998; Primiceri, 2005; Stroud et al., 2003). A salient feature of the KSC approach is that it produces samples from the posterior distribution of interest by sampling a more tractable approximate posterior distribution; the small approximation error is corrected by reweighting the sampled draws. This procedure is highly efficient in the sense that the sampled draws display weak serial dependence, a desirable feature of any well-constructed MCMC algorithm. It has long been believed, however, that the KSC approach cannot be extended to the broader and more realistic class of SV models with leverage. The purpose of this paper is to show that in fact such an extension is eminently possible and that the modified approach retains all the appealing characteristics of the original approach— simplicity, ease of implementation, and weak serial dependence in the sampled draws.

The simplest model we study is the discrete time log-normal SV model given by

$$y_t = \varepsilon_t \exp(h_t/2),$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t, \quad t = 0, 1, \ldots, n, \tag{1}$$

where y_t is the observed response, $\{h_t\}$ are unobserved log-volatilities, $|\phi| < 1$,

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix} \bigg| (\rho, \sigma) \sim \text{i.i.d. } \mathcal{N}_2(\mathbf{0}, \mathbf{\Sigma}), \quad \mathbf{\Sigma} = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}, \tag{2}$$

and $\mathcal{N}_p(\mathbf{m}, \mathbf{V})$ is the p-variate normal distribution with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{V}$. In this model, the parameter ρ measures the correlation between ε_t and η_t and, when negative, captures the increase in volatility that follows a drop in equity returns (e.g. Black, 1976; Nelson, 1991).

In the context of models without leverage, KSC approximate the distribution of $\log \varepsilon_t^2$ by a mixture of seven Gaussian distributions to match its first four moments. Conditioned on the latent mixture component indicators s_t ($t = 1, 2, \ldots, n$), this produces a model that is linear and Gaussian, with all its attendant benefits. Essentially, it then becomes possible to efficiently sample the posterior distribution of $\mathbf{s}_n = \{s_t\}_{t=1}^n$, the latent volatilities $\mathbf{h}_n = \{h_t\}_{t=1}^n$ and the parameters by MCMC methods. One key feature of their method is that it permits the joint sampling of $\mathbf{h}_n$ conditioned on $\mathbf{s}_n$ thus leading to posterior draws that mix better than approaches that rely on one-at-a-time sampling of the volatilities. The sampling is finished by a reweighting step to compensate for any error arising from the mixture approximation. In this paper, we show how this basic idea can be extended to SV models with leverage by starting with the joint distribution of $\log \varepsilon_t^2, \eta_t | \text{sign}(y_t), \rho, \sigma$ and approximating this distribution by a suitably constructed ten-component mixture of normal distributions. Our approach effectively solves the problems of fitting SV models with leverage. We also show how our new approach can be further extended to cover more general SV models than those given in (1).

For a deeper understanding of the model we analyze in this paper it is worth noting that the model in Jacquier et al. (2004) where ε_t and η_{t-1} are correlated is distinct and different. The model in (2) is appealing because it is an Euler approximation to the log-normal OU

SV model with leverage. Thus, the methods we develop in this paper, combined with those of Elerian et al. (2001), Eraker (2001) and Roberts and Stramer (2001), can be used to fit the corresponding continuous time model with discretely sampled data. Letting $\mathbf{y}_{t-1} = (y_1, \ldots, y_{t-1})$, another distinction is that a model with correlated ε_t and η_{t-1} implies that $y_t|\mathbf{y}_{t-1}$ can be skewed, while models which correlate ε_t and η_t have symmetric $y_t|\mathbf{y}_{t-1}$ unless ε_t is skewed itself. On the other hand, in the alternative specification ρ has two roles, leverage and skewness. In our view, the use of a single parameter to model two effects is not appealing because it makes the parameter difficult to interpret. Another downside of correlating ε_t and η_{t-1} is that y_t is no longer a martingale difference sequence. A more desirable way of introducing skewness in the distribution of $y_t|\mathbf{y}_{t-1}$ is by modeling ε_t as asymmetric within the setup of (2). This allows the SV model to maintain the martingale difference property in parallel fashion to what is done in the GARCH literature and in the literature on time-changed Lévy processes and Lévy-based SV models. Yu (2005) provides further discussion of some of these issues alongside empirical evidence that the model in (2) is better supported in a real data example.

The rest of the paper is organized as follows. In Section 2, we develop in detail our approach to dealing with SV models with leverage. Section 3 illustrates the performance of this method both on its own terms and in relation to a single move method that appears in Jacquier et al. (2004) in which the $\{h_t\}$ and the parameters are sampled one at a time, conditioned on the remaining values. In Section 4, we illustrate our techniques on data from the Japanese stock market. Section 5 deals with extensions of the method to superposition models (where the log-volatility is made up of a linear combination of heterogenous and independent autoregressions) and heavy-tailed error distributions (student and log-normal). Concluding remarks are contained in Section 6.

2. Efficient auxiliary mixture sampler

2.1. Reformulation in the no leverage case

To motivate our technique, recall from Nelson (1988), Harvey et al. (1994) and Harvey and Shephard (1996) that the process for y_t in (1) can equivalently be expressed in terms of the bivariate observations (d_t, y_t^*) where

$$d_t = I(\varepsilon_t \geqslant 0) - I(\varepsilon_t < 0), \tag{3}$$

$$y_t^* = \log y_t^2 = h_t + \varepsilon_{tt}^*, \tag{4}$$

and

$$\varepsilon_t^* = \log \varepsilon_t^2.$$

In other words, we have the map

$$y_t = d_t \exp(y_t^*/2).$$

When $\rho = 0$ it is easy to see that the signs of $\mathbf{y}_n = (y_1, \ldots, y_n)'$ are independent of $\mathbf{y}_n^* = (y_1^*, \ldots, y_n^*)'$ which means that we can neglect $\mathbf{d}_n = (d_1, \ldots, d_n)'$ and focus on the model in

terms of $\{y_t^*\}$ which is linear in $\{h_t\}$ with an i.i.d. error ε_t^* that follows a $\log \chi_1^2$ density

$$f(\varepsilon_t^*) = \frac{1}{\sqrt{2\pi}} \exp\left\{\frac{\varepsilon_t^* - \exp(\varepsilon_t^*)}{2}\right\}, \quad \varepsilon_t^* \in R.$$

Although the latter distributional form still precludes direct and simple inference, KSC introduced the idea of accurately approximating the $\log \chi_1^2$ distribution by a matched mixture of normal distributions

$$g(\varepsilon_t^*) = \sum_{j=1}^{K} p_j \mathcal{N}(\varepsilon_t^* | m_j, v_j^2), \quad \varepsilon_t^* \in R, \tag{5}$$

where $\mathcal{N}(\varepsilon_t^* | m_j, v_j^2)$ denotes the density function of a normal distribution with mean m_j and variance v_j^2. The values of p_j, m_j and v_j^2 found by KSC on the basis of $K = 7$ components are reproduced in the first block of columns in Table 1. They proceeded to develop an efficient Bayesian MCMC method for sampling the resulting posterior distribution and then reweighted the sampled draws in a way to ensure that the variates corresponded to the posterior under the $\log \chi_1^2$ sampling density. The entire approach was shown to be efficient and readily implementable.

In our current work, we have favored a tighter approximation to the density of the $\log \chi_1^2$ distribution that utilizes $K = 10$ components. The component parameters are given in the second block of Table 1. For the moment, the columns in the table labeled a_j and b_j can be ignored.

That the move to $K = 10$ components leads to a superior approximation is illustrated in Fig. 1 where we plot the difference between the density of the $\log \chi_1^2$ distribution and the approximating mixture distribution, evaluated over the range from the first to the 99th percentiles. It can be seen from the first row of this figure that the new mixture with $K = 10$ components provides a much closer fit. The second row of the figure shows that the new approximation is also close for the density of $\sqrt{\chi_1^2}$.

Table 1
Selection of $(p_j, m_j, v_j^2, a_j, b_j)$

j	KSC			$K = 10$				
	p_j	m_j	v_j^2	p_j	m_j	v_j^2	a_j	b_j
1	0.04395	1.50746	0.16735	0.00609	1.92677	0.11265	1.01418	0.50710
2	0.24566	0.52478	0.34023	0.04775	1.34744	0.17788	1.02248	0.51124
3	0.34001	−0.65098	0.64009	0.13057	0.73504	0.26768	1.03403	0.51701
4	0.25750	−2.35859	1.26261	0.20674	0.02266	0.40611	1.05207	0.52604
5	0.10556	−5.24321	2.61369	0.22715	−0.85173	0.62699	1.08153	0.54076
6	0.00002	−9.83726	5.17950	0.18842	−1.97278	0.98583	1.13114	0.56557
7	0.00730	−11.40039	5.79596	0.12047	−3.46788	1.57469	1.21754	0.60877
8				0.05591	−5.55246	2.54498	1.37454	0.68728
9				0.01575	−8.68384	4.16591	1.68327	0.84163
10				0.00115	−14.65000	7.33342	2.50097	1.25049

Left-hand side was determined by Kim, Shephard and Chib, the ones on the right-hand side are new and represent a better approximation.
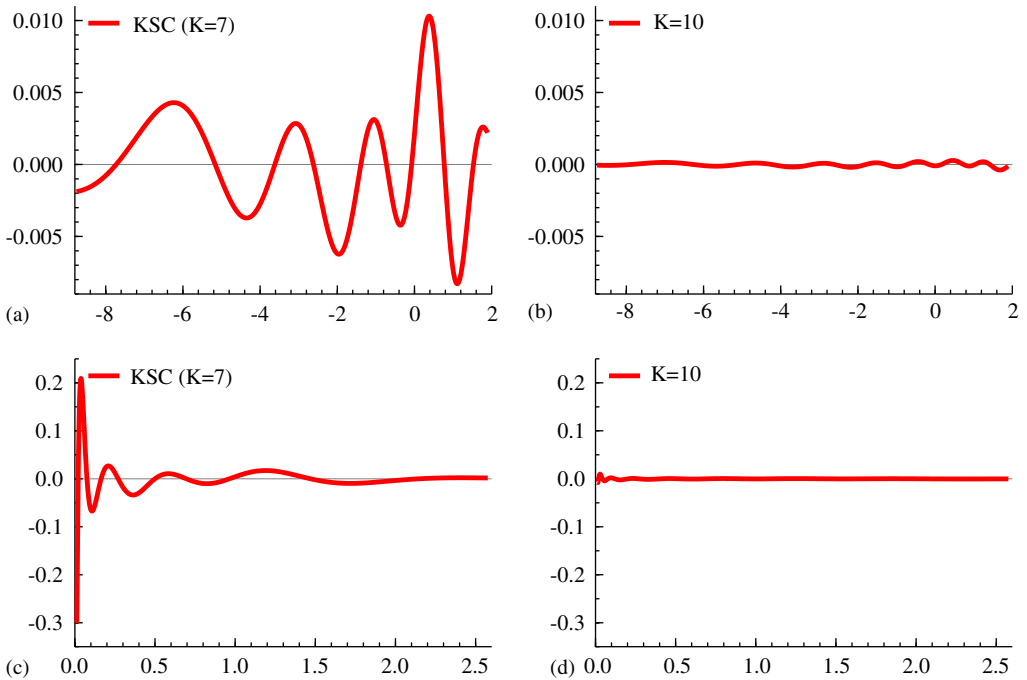
Fig. 1. The difference between the approximate and the true densities (for the range from the first percentile to the 99th percentile). The $\log \chi_1^2$ density (top) and the $\sqrt{\chi_1^2}$ density (bottom). (a) Difference of densities: Log of χ_1^2 (b) Difference of densities: Log of χ_1^2 (c) Difference of densities: Square root of χ_1^2 (d) Difference of densities: Square root of χ_1^2.

2.2. Reformulation in general case

Now consider the general case of $\rho \neq 0$. The conditional distribution of η is

$$\eta_t | d_t, \varepsilon_t^*, \rho, \sigma \sim \mathcal{N}(d_t \rho \sigma \exp(\varepsilon_t^*/2), \sigma^2(1 - \rho^2)), \tag{6}$$

which shows that we have two complications: d_t is not ignorable and ε_t^* enters both (4) and (6). Our idea now is to approximate the bivariate conditional density of

$$\varepsilon_t^*, \eta_t | d_t, \rho, \sigma$$

as a mixture of bivariate normal densities. To develop our approximation we start with the usual marginal-conditional decomposition

$$
\begin{aligned}
f(\varepsilon_t^*, \eta_t | d_t, \rho, \sigma) &= f(\varepsilon_t^* | d_t) f(\eta_t | \varepsilon_t^*, d_t, \rho, \sigma) \\
&= f(\varepsilon_t^*) f(\eta_t | \varepsilon_t^*, d_t, \rho, \sigma).
\end{aligned} \tag{7}
$$

We now approximate $f(\varepsilon_t^*)$ by the mixture distribution given in (5) and let the approximation of $f(\varepsilon_t^*, \eta_t | d_t, \rho, \sigma)$ take the form

$$g(\varepsilon_t^*, \eta_t | d_t, \rho, \sigma) = \sum_{j=1}^{K} p_j \mathcal{N}(\varepsilon_t^* | m_j, v_j^2) \mathcal{N}[\eta_t | d_t \rho \sigma \exp(m_j/2)\{a_j + b_j(\varepsilon_t^* - m_j)\}, \sigma^2(1 - \rho^2)],$$

$$\tag{8}$$

where the second term in the jth component

$$\mathcal{N}[\eta_t | d_t \rho \sigma \exp(m_j/2)\{a_j + b_j(\varepsilon_t^* - m_j)\}, \sigma^2(1 - \rho^2)]$$

is intended to match the density of $\eta_t | d_t, \varepsilon_t^*, \rho, \sigma$ given in (6). Specifically, on inspection of the form of the density in (6), it can be seen that we are approximating

$$\exp(\varepsilon_t^*/2)$$

by

$$\exp(m_j/2)(a_j + b_j(\varepsilon_t^* - m_j)),$$

given $\varepsilon_t^* \sim \mathcal{N}(m_j, v_j^2)$. We focus on this approximation because it does not depend upon ρ. Interestingly, ρ does not affect the quality of the approximation as we show below. To find the values (a_j, b_j), $j = 1, \ldots, 10$, we consider the mean square norm and let

$$(a_j, b_j) = \arg\min_{a,b} E\{\exp(\varepsilon_t^*/2)\exp(-m_j/2) - a - b(\varepsilon_t^* - m_j)\}^2, \quad \varepsilon_t^* \sim \mathcal{N}(m_j, v_j^2),$$

$$j = 1, 2, \ldots, 10.$$

By calculation, we get

$$a_j = \exp(v_j^2/8),$$

$$b_j = E\{z_t \exp(v_j z_t/2)\} = \frac{1}{2}\exp\left(\frac{v_j^2}{8}\right), \quad j = 1, 2, \ldots, 10,$$

which are evaluated and reproduced in Table 1.

Remark 1. The key question is how well (8) approximates (7). We give results for $\rho = -0.3, -0.6$ and -0.9. Fig. 2 shows f and g for $\eta_t | \varepsilon_t^*, d_t = 1$ evaluated with ε_t^* set at its 25th, 50th and 75th percentiles. Likewise Fig. 3 shows f and g for $\varepsilon_t^* | \eta_t, d_t = 1$ evaluated with $\eta_t = -0.67\sigma, 0, 0.67\sigma$. The results suggest the approximation is quite good for it is very hard to see any difference between the true densities f and the approximations g. Further, Fig. 4 shows the marginal density of η_t given $d_t = 1$. It is clear that the true conditional joint density given d_t is well approximated by the stated bivariate normal mixture.

2.3. MCMC algorithm

2.3.1. Broad principles

The SV model can be expressed as

$$\begin{pmatrix} y_t^* \\ h_{t+1} \end{pmatrix} = \begin{pmatrix} h_t \\ \mu + \phi(h_t - \mu) \end{pmatrix} + \begin{pmatrix} \varepsilon_t^* \\ \eta_t \end{pmatrix}.$$

Now on using the mixture approximation (8) to the density $\varepsilon_t^*, \eta_t | d_t, \rho, \sigma$ and introducing the mixture component indicator $s_t \in \{1, 2, \ldots, K\}$ we have that

$$\left\{ \begin{pmatrix} \varepsilon_t^* \\ \eta_t \end{pmatrix} \middle| d_t, s_t = j, \rho, \sigma \right\} \overset{L}{=} \begin{pmatrix} m_j + v_j z_t \\ d_t \rho \sigma (a_j + b_j v_j z_t) \exp(m_j/2) + \sigma \sqrt{1 - \rho^2} z_t^* \end{pmatrix},$$

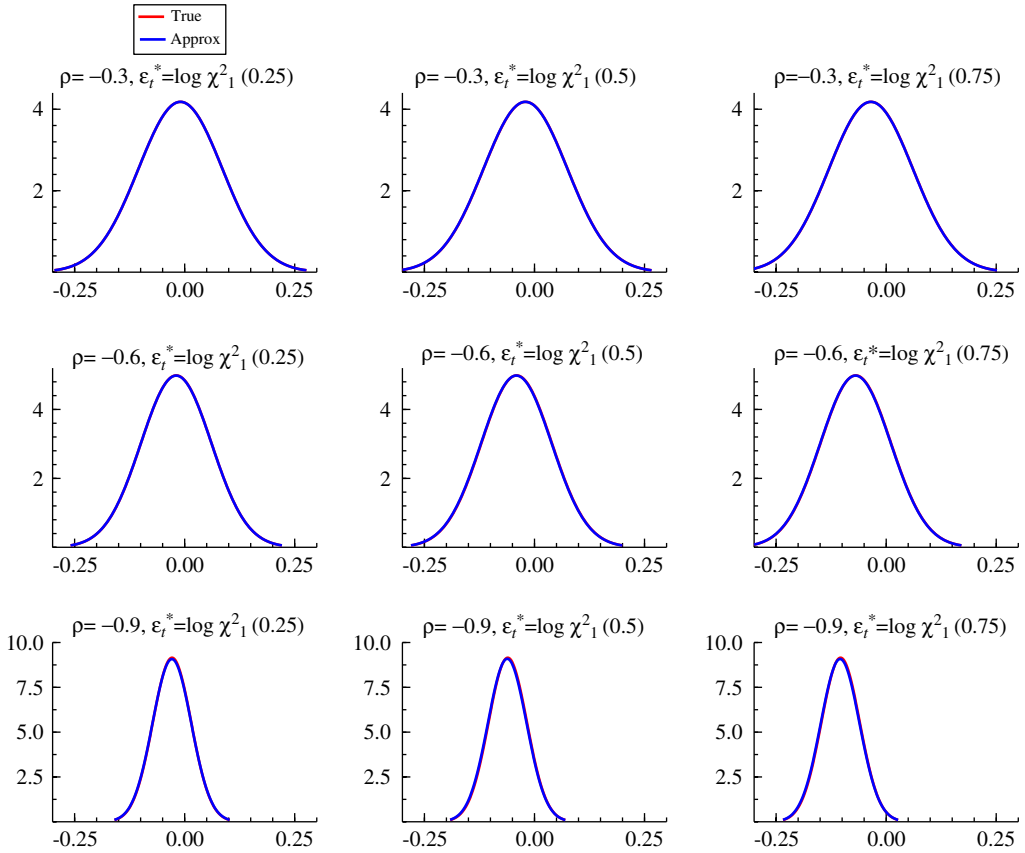$$\begin{pmatrix} z_t \\ z_t^* \end{pmatrix} \overset{\text{i.i.d.}}{\sim} \mathcal{N}_2(\mathbf{0}, \mathbf{I}),$$

Fig. 2. The conditional density of η_t given $d_t = 1$ and $\varepsilon_t^* = \log \chi_1^2(0.25), \log \chi_1^2(0.5), \log \chi_1^2(0.75)$ (left, middle, right) for $\rho = -0.3, -0.6, -0.9$ (top, middle, bottom).

where $A \overset{L}{=} B$ implies that A and B have the same distribution. If we let $\boldsymbol{\theta} = (\phi, \rho, \sigma)$ and assume that $h_1 | \mu, \boldsymbol{\theta} \sim \mathcal{N}(\mu, \sigma^2/(1 - \phi^2))$, then under the auxiliary notation

$$\widetilde{\mu}_1 = \widetilde{\mu}_2 = \cdots = \widetilde{\mu}_n = \mu,$$

we have that the SV model with leverage can be expressed in linear Gaussian state space form (e.g. Harvey, 1989; West and Harrison, 1997; Durbin and Koopman, 2001)

$$\begin{pmatrix} y_t^* \\ h_{t+1} \\ \widetilde{\mu}_{t+1} \end{pmatrix} = \begin{pmatrix} h_t \\ \widetilde{\mu}_t + \phi(h_t - \widetilde{\mu}_t) \\ \widetilde{\mu}_t \end{pmatrix} + \begin{pmatrix} \varepsilon_t^* \\ \eta_t \\ 0 \end{pmatrix}, \tag{9}$$

for $t = 1, 2, \ldots, n$, and

$$\begin{pmatrix} h_1 \\ \widetilde{\mu}_1 \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu_0 \\ \mu_0 \end{pmatrix}, \begin{pmatrix} \sigma^2/(1 - \phi^2) + \sigma_0^2 & \sigma_0^2 \\ \sigma_0^2 & \sigma_0^2 \end{pmatrix} \right). \tag{10}$$
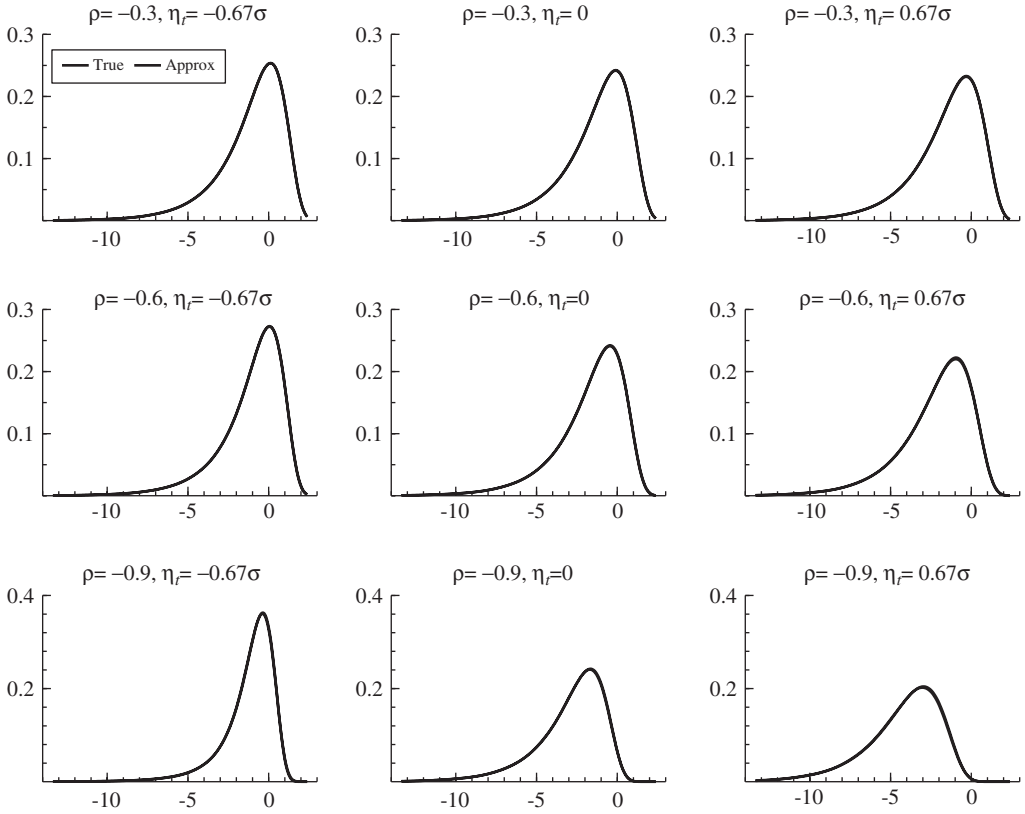
Fig. 3. The conditional density of ξ_t given $d_t = 1$ and $\eta_t = -0.67\sigma$, 0, 0.67σ (left, middle, right) for $\rho = -0.3$, -0.6, -0.9 (top, middle, bottom). The value of σ is set to 1 in this example.
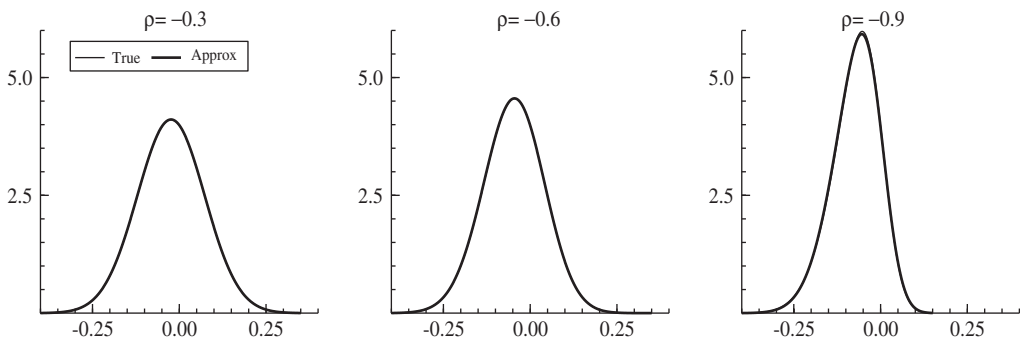

Fig. 4. The marginal density of η_t given $d_t = 1$ for $\rho = -0.3$, -0.6, -0.9 (left, middle, right). The value of σ is set to 1 in this example.

Under a given prior $\pi(\boldsymbol{\theta})$ on $\boldsymbol{\theta}$ and a normal prior on μ ($\mu \sim \mathcal{N}(\mu_0, \sigma_0^2)$), it is now possible to efficiently sample the posterior density

$$g(\mathbf{s}_n, \mathbf{h}_n, \mu, \boldsymbol{\theta} | \mathbf{y}_n^*, \mathbf{d}_n) \tag{11}$$

by MCMC techniques (see for example Chib, 2001 for a review of these methods). Of course, due to the approximation of $\varepsilon_t^*, \eta_t | d_t, \rho, \sigma$, this posterior is not exactly the correct one, but we will see in Section 2.4 that it is easy to correct the small error by reweighting the sampled draws.

There are a number of different ways of sampling the posterior density above but the scheme given next is relatively simple, fast and effective as we will show. First we initialize $\mathbf{s}_n, \mathbf{h}_n, \mu$ and $\boldsymbol{\theta}$, and then iterate the following steps to obtain a posterior sample.

1. Draw $\mathbf{s}_n | \mathbf{y}_n^*, \mathbf{d}_n, \mathbf{h}_n, \mu, \boldsymbol{\theta}$.
2. Draw $(\mathbf{h}_n, \mu, \boldsymbol{\theta}) | \mathbf{y}_n^*, \mathbf{d}_n, \mathbf{s}_n$ by
    (a) drawing $\boldsymbol{\theta} | \mathbf{y}_n^*, \mathbf{d}_n, \mathbf{s}_n$;
    (b) drawing $\mathbf{h}_n, \mu | \mathbf{y}_n^*, \mathbf{d}_n, \mathbf{s}_n, \boldsymbol{\theta}$.

2.3.2. Step 1

We first define

$$\varepsilon_t^* = y_t^* - h_t, \quad \eta_t = (h_{t+1} - \mu) - \phi(h_t - \mu),$$

then evaluate for each $j = 1, 2, \ldots, K$

$$\pi(s_t = j | \mathbf{y}_n^*, \mathbf{d}_n, \mathbf{h}_n, \mu, \boldsymbol{\theta})$$
$$\propto \pi(s_t = j | \varepsilon_t^*, \eta_t, d_t, \mu, \boldsymbol{\theta})$$
$$\propto \Pr(s_t = j) v_j^{-1} \exp\left\{ -\frac{(\varepsilon_t^* - m_j)^2}{2v_j^2} - \frac{[\eta_t - d_t \rho \sigma \exp(m_j/2)\{a_j + b_j(\varepsilon_t^* - m_j)\}]^2}{\sigma^2(1 - \rho^2)} \right\}.$$

This discrete distribution is sampled by the inverse distribution method.

2.3.3. Step 2

In Step 2a we sample the density

$$g(\boldsymbol{\theta} | \mathbf{y}_n^*, \mathbf{d}_n, \mathbf{s}_n) \propto g(\mathbf{y}_n^* | \mathbf{d}_n, \mathbf{s}_n, \boldsymbol{\theta})\pi(\boldsymbol{\theta}),$$

marginalized over μ. The density $g(\mathbf{y}_n^* | \mathbf{d}_n, \mathbf{s}_n, \boldsymbol{\theta})$ is found from the output of the Kalman filter recursions applied to the models in (9) and (10). As one of the elements of the state vector is μ, which is time-invariant, this density can also be computed by the so-called augmented Kalman filter (e.g. Durbin and Koopman, 2001) but this procedure is computationally more involved. For the sampling we rely on the Metropolis–Hastings algorithm with a proposal density based on truncated Gaussian approximation of $\pi(\boldsymbol{\theta} | \mathbf{y}_n^*, \mathbf{d}_n, \mathbf{s}_n)$ (Chib and Greenberg, 1994, 1995). We define $\widehat{\boldsymbol{\theta}} = (\hat{\phi}, \widehat{\sigma^2}, \hat{\rho})'$ which maximizes (or approximately maximizes) $g(\mathbf{y}_n^* | \mathbf{d}_n, \mathbf{s}_n, \boldsymbol{\theta})\pi(\boldsymbol{\theta})$. Then we generate a candidate $\boldsymbol{\gamma}^*$ from the normal distribution truncated on the region R, $\mathrm{TN}_R(\widehat{\boldsymbol{\theta}}, \boldsymbol{\Sigma}_*)$, where

$$\Sigma_*^{-1} = -\left.\frac{\partial^2 \log g(\mathbf{y}_n^* | \mathbf{d}_n, \mathbf{s}_n, \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right|_{\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}},$$

and $R = \{\boldsymbol{\gamma} : |\phi| < 1, \sigma^2 > 0, |\rho| < 1\}$. Alternatively, we may generate a candidate from an untruncated Gaussian proposal using a transformation $\theta_1 = \log(1 + \phi) - \log(1 - \phi), \theta_2 = \log \sigma_1^2, \theta_3 = \log(1 + \rho) - \log(1 - \rho)$.

The proposal values are accepted or rejected according to the Metropolis–Hastings probability of move. When the Hessian matrix is not negative definite (e.g. when $|\hat{\rho}| \approx 1$), we take a normal proposal $TN_R(\hat{\theta}, c_0 I)$ using some large constant c_0.

Step 2b, the sampling of $\mathbf{h}_n, \mu | \mathbf{y}_n^*, \mathbf{d}_n, \mathbf{s}_n, \theta$ is simple and is implemented with the help of the Gaussian simulation smoother (Frühwirth-Schnatter, 1994; Carter and Kohn, 1994; de Jong and Shephard, 1995; Durbin and Koopman, 2002). Software for carrying out Gaussian simulation smoothing is widely available (Koopman et al., 1999).

2.4. Correcting for misspecification

In our approach, we approximate the true bivariate density $f(\varepsilon_t^*, \eta_t | d_t, \mu, \theta)$ with our convenient mixture density $g(\varepsilon_t^*, \eta_t | d_t, \mu, \theta)$. Thus the draws from our MCMC procedure

$$\mathbf{h}_n^k, \mu^k, \theta^k, \quad k = 1, 2, \ldots, M,$$

are from the approximate posterior density $g(\mathbf{h}_n, \mu, \theta | \mathbf{y}_n^*, \mathbf{d}_n)$. To produce draws from the correct posterior density $\pi(\mathbf{h}_n, \mu, \theta | \mathbf{y}_n^*, \mathbf{d}_n)$ we simply reweight the sampled draws. Define

$$\varepsilon_t^{*k} = y_t^* - h_t^k, \quad \eta_t^k = (h_{t+1}^k - \mu^k) - \phi^k(h_t^k - \mu^k).$$

Then we compute the weights

$$w_k^* = \prod_{t=1}^{n} \frac{f(\xi_t^k, \eta_t^k | d_t, \mu^k, \theta^k)}{g(\xi_t^k, \eta_t^k | d_t, \mu^k, \theta^k)}, \quad k = 1, 2, \ldots, M,$$

and let

$$w_k = \frac{w_k^*}{\sum_{l=1}^{M} w_l^*}.$$

We can now acquire a sample from $\pi(h, \mu, \theta | y^*, d)$ by resampling the sampled variates with weights proportional to w_j. Furthermore, posterior moments can be computed by weighted averaging of the MCMC draws as in KSC. For example, we can estimate the posterior mean of $g(\theta)$ as

$$\hat{E}[g(\mu, \theta)] = \sum_{k=1}^{M} w_k g(\mu^k, \theta^k).$$

We will see in the Monte Carlo experiments and in the empirical work that the variance of these weights is small, a consequence of the accuracy of our approximation. This implies, therefore, that the effect of the reweighting is modest.

2.5. Associated particle filter

We complete our methodological developments for this model by presenting a simulation-based approach to filtering. In particular, we show how we can recursively sample the distributions $(h_t | \mathbf{y}_t, \mu, \theta)$, $(h_{t+1} | \mathbf{y}_t, \mu, \theta)$ and $(y_{t+1} | \mathbf{y}_t, \mu, \theta)$. These sampled variates are needed in order to calculate marginal likelihoods, Bayes factors and goodness of fit statistics. We implement the filtering and associated computations by particle filter methods (e.g. in this context, KSC; Pitt and Shephard, 1999 or more generally Doucet et al., 2001).

The SV model with leverage can be expressed in the form of a non-linear, non-Gaussian state space model consisting of the measurement density

$$f(y_t|h_t) = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}h_t - \frac{1}{2}y_t^2 \exp(-h_t) \right\} \tag{12}$$

and the evolution density

$$f(h_{t+1}|y_t, h_t, \mu, \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma} \exp\left\{ -\frac{(h_{t+1} - \mu_{t+1})^2}{2(1-\rho^2)\sigma^2} \right\}, \tag{13}$$

where $\mu_{t+1} = \mu + \phi(h_t - \mu) + \rho\sigma \exp(-h_t/2)y_t$. To develop our particle filtering method we start with the fact (from Bayes theorem) that

$$f(h_{t+1}, h_t|\mathbf{y}_{t+1}, \mu, \boldsymbol{\theta}) \propto f(y_{t+1}|h_{t+1})f(h_{t+1}|\mathbf{y}_t, h_t, \mu, \boldsymbol{\theta})f(h_t|\mathbf{y}_t, \mu, \boldsymbol{\theta})$$

where we assume that we have samples (particles) from $f(h_t|\mathbf{y}_t, \mu, \boldsymbol{\theta})$, and a discrete uniform approximation $\hat{f}(h_t|\mathbf{y}_t, \mu, \boldsymbol{\theta})$ to $f(h_t|\mathbf{y}_t, \mu, \boldsymbol{\theta})$. In principle, given this discrete distribution we could sample h_t from $\hat{f}(h_t|\mathbf{y}_t, \mu, \boldsymbol{\theta})$, then h_{t+1} from $f(h_{t+1}|\mathbf{y}_t, h_t, \mu, \boldsymbol{\theta})$, a process we could repeat many times to generate a sample of values on h_{t+1}. We finish the process by resampling the latter values with weights proportional to $f(y_{t+1}|h_{t+1}, \boldsymbol{\theta})$. By appealing to the theory of importance sampling it can be shown that these resampled particles are from $f(h_{t+1}|\mathbf{y}_{t+1}, \mu, \boldsymbol{\theta})$.

Effectively, in the process just described, the target posterior density, namely $f(h_{t+1}, h_t|\mathbf{y}_{t+1}, \mu, \boldsymbol{\theta})$, is sampled with the help of $f(h_{t+1}|\mathbf{y}_t, h_t, \mu, \boldsymbol{\theta}) \times \hat{f}(h_t|\mathbf{y}_t, \mu, \boldsymbol{\theta})$ as the importance function. It turns out, however, that it is advantageous to also involve y_{t+1} in the importance function. To this end, given the ith particle h_t^i from $f(h_t|\mathbf{y}_t, \mu, \boldsymbol{\theta})$, we let the importance function be

$$\begin{aligned} g(h_{t+1}, h_t^i|\mathbf{y}_{t+1}, \mu, \boldsymbol{\theta}) &\propto f(y_{t+1}|\mu_{t+1}^i)f(h_{t+1}|y_t, h_t^i, \mu, \boldsymbol{\theta})\hat{f}(h_t^i|\mathbf{y}_t, \mu, \boldsymbol{\theta}) \\ &\propto f(h_{t+1}|y_t, h_t^i, \mu, \boldsymbol{\theta})g(h_t^i|\mathbf{y}_{t+1}, \mu, \theta), \end{aligned} \tag{14}$$

where

$$g(h_t^i|\mathbf{y}_{t+1}, \mu, \theta) = \frac{f(y_{t+1}|\mu_{t+1}^i)\hat{f}(h_t^i|\mathbf{y}_t, \mu, \boldsymbol{\theta})}{\sum_{j=1}^I f(y_{t+1}|\mu_{t+1}^j)\hat{f}(h_t^j|\mathbf{y}_t, \mu, \boldsymbol{\theta})},$$

$$f(y_{t+1}|\mu_{t+1}^i) = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\mu_{t+1}^i - \frac{1}{2}y_{t+1}^2 \exp(-\mu_{t+1}^i) \right\},$$

$$\mu_{t+1}^i = \mu + \phi(h_t^i - \mu) + \rho\sigma \exp(-h_t^i/2)y_t.$$

This leads to the following particle filtering method.

1. Initialize $t = 1$, h_1^i from its unconditional distribution for $i = 1, 2, \ldots, I$.
   (a) Compute $w_i = f(y_1|h_1^i)$ and $W_i = F(y_1|h_1^i)$, (where F denotes the distribution function of y_t given h_t) and record

   $$\overline{w}_1 = \frac{1}{I}\sum_{i=1}^I w_i, \quad \overline{W}_1 = \frac{1}{I}\sum_{i=1}^I W_i.$$

   (b) Let $\hat{f}(h_1^i|y_1, \mu, \theta) = \pi_1^i = w_i/\sum_{j=1}^I w_j$, $i = 1, 2, \ldots, I$.

2. For each i, simulate h_t^i and h_{t+1}^i, $i = 1, \ldots, I$, using the importance function $g(h_{t+1}, h_t | Y_{t+1}, \mu, \theta)$ given by (14). Compute

$$w_i = \frac{f(y_{t+1}|h_{t+1}^i)f(h_{t+1}^i|y_t, h_t^i, \mu, \theta)\hat{f}(h_t^i|\mathbf{y}_t, \mu, \theta)}{g(h_{t+1}, h_t^i|\mathbf{y}_{t+1}, \mu, \theta)} = \frac{f(y_{t+1}|h_{t+1}^i)\hat{f}(h_t^i|\mathbf{y}_t, \mu, \theta)}{g(h_t^i|\mathbf{y}_{t+1}, \mu, \theta)},$$

$$W_i = \frac{F(y_{t+1}|h_{t+1}^i)\hat{f}(h_t^i|\mathbf{y}_t, \mu, \theta)}{g(h_t^i|\mathbf{y}_{t+1}, \mu, \theta)}, \quad i = 1, \ldots, I,$$

and record

$$\overline{w}_t = \frac{1}{I}\sum_{i=1}^{I} w_i, \quad \overline{W}_t = \frac{1}{I}\sum_{i=1}^{I} W_i. \tag{15}$$

Let $\hat{f}(h_{t+1}^i|\mathbf{y}_{t+1}, \mu, \theta) = \pi_{t+1}^i = w_i / \sum_{j=1}^{I} w_j$, $i = 1, 2, \ldots, I$.
3. Increment t and go to 2.

It can be shown that as $I \to \infty$, $\overline{w}_{t+1} \overset{p}{\to} f(y_{t+1}|\mathbf{y}_t, \mu, \theta)$, and

$$\overline{W}_{t+1} \overset{p}{\to} F(y_{t+1}|\mathbf{y}_t, \mu, \theta),$$

the predictive distribution function. In addition, the draws on h_{t+1} are particles from $h_{t+1}|\mathbf{y}_t, \mu, \theta$, while the resampled items at stage 3 are samples from $h_{t+1}|\mathbf{y}_{t+1}, \mu, \theta$. It therefore follows that

$$\sum_{t=1}^{n} \log \overline{w}_t \overset{p}{\to} \sum_{t=1}^{n} \log f(y_t|\mathbf{y}_{t-1}, \mu, \theta),$$

is a consistent estimate of the conditional log-likelihood and can be used in the method of Chib (1995) to calculate the marginal likelihood. Likewise the sequence of $\overline{W}_t$, and its reflected version $2|\overline{W}_t - \frac{1}{2}|$, can be used to check for model fit as these are approximately i.i.d. standard uniform if the model is correctly specified. This diagnostic was introduced into econometrics by KSC, following the earlier work of Shephard (1994), Smith (1985) and Rosenblatt (1952). Diagnostic checking of this type has been further popularized by Diebold et al. (1998).

3. Illustrative example

3.1. Auxiliary mixture sampler

This section gives illustrative examples to show the performance of the approximation discussed above. Throughout we use $y_t^* = \log(y_t^2 + c)$ where the offset c is used to deal with very small values of y_t^2 as in KSC. Because our ten-component mixture approximation provides an improved fit to the left tail of the log χ_1^2 density we set c equal to 0.0001 which is smaller than the value of $c = 0.001$ used by KSC.

We simulated the data from the SV model (1) where we set $\phi = 0.97$, $\beta \equiv \exp(\mu/2) = 0.65$, $\sigma = 0.15$ and $\rho = -0.3$. These values are based on the estimates reported by KSC and Yu (2005) in their analysis of daily returns on foreign exchange rates and the S&P500 index. In addition, we also consider models with $\rho = 0, -0.6, -0.9$ to

investigate the effect of ρ on the quality of our inferences. In each case, we consider samples with $n = 1000$ observations.

The results are based on the prior distribution

$$\mu \sim \mathcal{N}(0, 1), \quad \frac{\phi + 1}{2} \sim \text{Beta}(20, 1.5),$$

$$\sigma^{-2} \sim \text{Gamma}\left(\frac{5}{2}, \frac{0.05}{2}\right), \quad \rho \sim U(-1, 1),$$

where $U(-1, 1)$ denotes a uniform distribution on $(-1, 1)$. The priors for ϕ and σ^2 are the same as those used by KSC. In particular, we use a beta prior on $(\phi + 1)/2$ to ensure the stationarity of the latent volatility process. To reflect the high persistence of the process in the previous empirical literature, we set $E(\phi) = 0.86$ and $\sqrt{Var(\phi)} = 0.11$. In the MCMC sampling of the posterior distribution, the initial 500 variates are discarded and the subsequent $M = 5000$ values are retained for purposes of analysis. Fig. 5 shows the sample autocorrelations function, the sample paths and the posterior densities of parameters for the case $\rho = -0.3$. The sample paths look stable and the autocorrelations decay quickly. In Table 2, the summary statistics are given for the cases $\rho = -0.3$, -0.6 and -0.9. The posterior means are close to the true values, and all true values are contained in the 95% credible intervals.

To measure how well the chain mixes, we calculate the inefficiency factors (the inverse of inefficiency factor is also known as numerical efficiency in Geweke (1992)). The inefficiency factor (equivalently the autocorrelation time) is defined as $1 + 2\sum_{s=1}^{\infty} \rho_s$ where ρ_s is the sample autocorrelation at lag s calculated from the sampled values.[1] It is also the ratio of the numerical variance of the posterior sample mean to the variance of the posterior sample mean from the hypothetical uncorrelated draws. Thus it suggests the relative number of correlated draws necessary to attain the same variance of the posterior sample mean from the uncorrelated draws. In KSC, where the $\rho = 0$ case was considered, the inefficiency factors were between 30 and 150 (Table 5, KSC) for the original mixture sampler and between 10 and 16 for the improved integration sampler (Table 6, KSC). In our MCMC implementation, these values are still small for $\rho = -0.3, -0.6$ and -0.9, showing that our sampler is highly effective. Among the parameters $(\phi, \sigma, \rho, \beta)$, the inefficiency factor of β is the smallest while that of σ tends to be the largest. The leverage effect parameter ρ may have relatively larger values for higher negative correlations.

In order to judge the quality of our approximation we next report the distribution of the weights as discussed above. Figs. 6 and 7 shows the distribution of $\log(w_k \times M)$, which would all have been zero if the approximation were exact. Fig. 6 looks at the case of $\rho = 0$ and compares the $K = 7$ component analysis used by KSC to our more refined $K = 10$ component analysis. While the standard deviation of the log-weights based on $K = 7$ is 0.92, it is 0.05 when $K = 10$. KSC demonstrated that reweighting had little impact on posterior inference about θ, μ, so we would expect that the improvement here is small from a practical viewpoint.

In Fig. 7, the distributions of $\log(w_k \times M)$ are shown for our new approximation in an asymmetric volatility model ($\rho = -0.3, -0.6, -0.9$). For $\rho = -0.3$, its standard deviation is 0.41, which is much smaller than that of KSC in the symmetric volatility model. For $\rho = -0.6$, the distribution is skewed to the left, and we have a slightly larger but still small

---

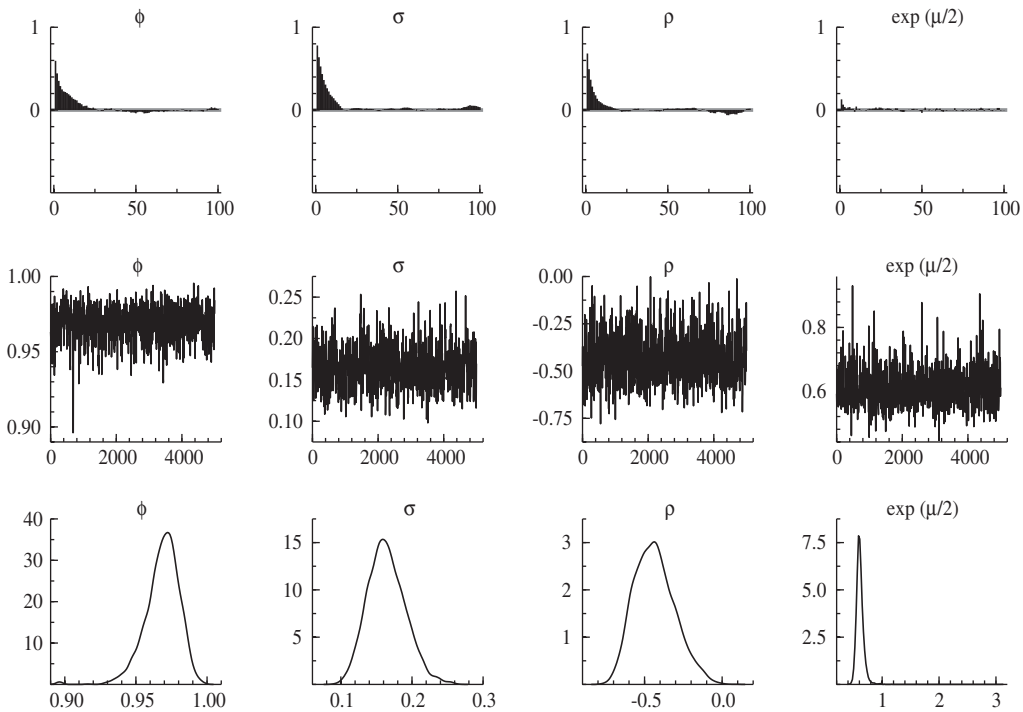[1] We have used 2500 lags in the estimation of the inefficiency factors.

Fig. 5. Asymmetric stochastic volatility model ($\rho = -0.3$). Sample autocorrelation functions, sample paths and estimated posterior densities.

Table 2
Auxiliary mixture sampler

Parameter	True value	Unweighted		Weighted		95% interval	Inefficiency
		Mean	St. dev.	Mean	St. dev.		
ϕ	0.97	0.9685	0.0122	0.9686	0.0121	[0.9420, 0.9880]	8.4
σ	0.15	0.1644	0.0265	0.1644	0.0264	[0.1174, 0.2184]	10.1
ρ	**−0.3**	−0.4351	0.1295	−0.4338	0.1301	[−0.6579, −0.1532]	6.8
β	0.65	0.6105	0.0723	0.6108	0.0751	[0.5118, 0.7475]	2.1
ϕ	0.97	0.9680	0.0113	0.9685	0.0112	[0.9429, 0.9865]	7.4
σ	0.15	0.1684	0.0261	0.1681	0.0260	[0.1214, 0.2273]	7.8
ρ	**−0.6**	−0.6528	0.1004	−0.6501	0.0996	[−0.8242, −0.4340]	7.2
β	0.65	0.6130	0.0585	0.6132	0.0580	[0.5284, 0.7479]	3.1
ϕ	0.97	0.9595	0.0102	0.9592	0.0101	[0.9349, 0.9766]	8.7
σ	0.15	0.1694	0.0236	0.1688	0.0240	[0.1271, 0.2195]	11.2
ρ	**−0.9**	−0.8613	0.0662	−0.8515	0.0627	[−0.9693, −0.7159]	14.7
β	0.65	0.6807	0.0345	0.6841	0.0225	[0.6106, 0.7466]	5.3

Summary statistics for three simulation experiments using a variety of values of ρ. Sample size is 1000 throughout.
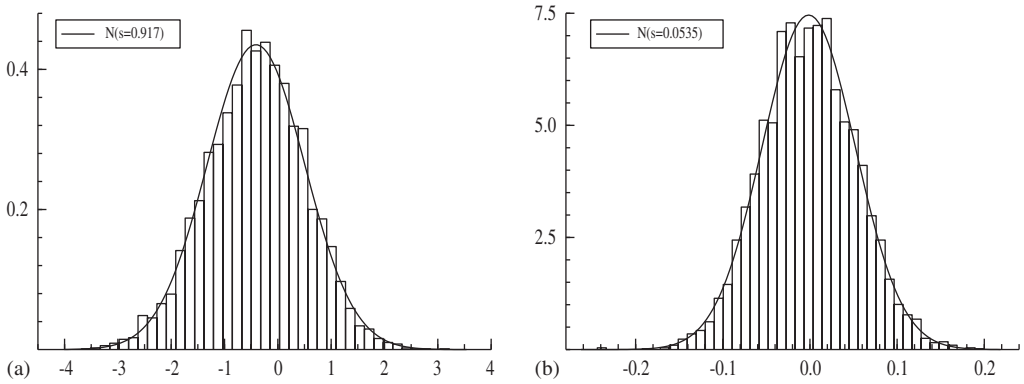
Fig. 6. Histogram of the $\log(w_k \times M)$ where $M = 5000$ is a number of samples for a symmetric stochastic volatility model ($\rho = 0$). Left: KSC ($K = 7$). Right: new ($K = 10$). Other line: the normal density function setting its mean and variance equal to the sample mean and sample variance. (a) log-weights: KSC ($K = 7$) (b) log-weights: New ($K = 10$).
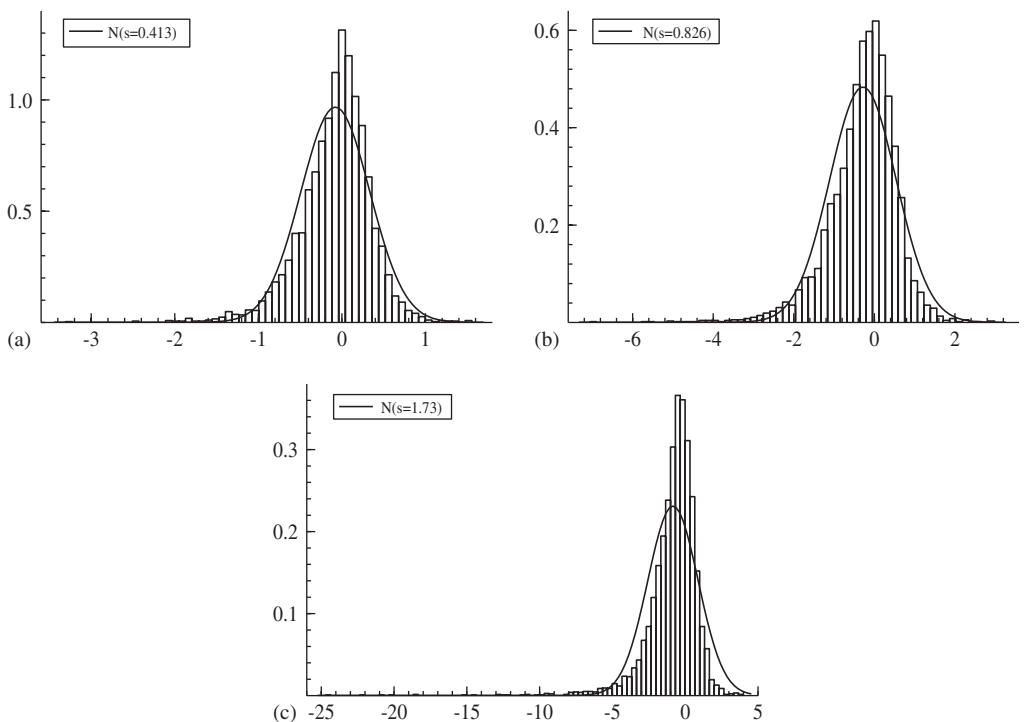
Fig. 7. Histogram of the $\log(w_k \times M)$ where $M = 5000$ is a number of samples for an asymmetric stochastic volatility model. Top right: new ($\rho = -0.3, K = 10$). Top left: new ($\rho = -0.6, K = 10$). Bottom left: new ($\rho = -0.9, K = 10$). Other line: the normal density function setting its mean and variance equal to the sample mean and sample variance. (a) log-weights: New ($\rho = -0.3, K = 10$) (b) log-weights: New ($\rho = -0.6, K = 10$) (c) log-weights: New ($\rho = -0.9, K = 10$).

Table 3
Single move sampler

Parameter	True value	Mean	St. dev.	95% interval	Inefficiency
ϕ	0.97	0.9629	0.0133	[0.9330, 0.9848]	598.7
σ	0.15	0.1712	0.0295	[0.1201, 0.2342]	1350.4
ρ	**−0.3**	−0.4043	0.1235	[−0.6223, −0.1418]	620.5
β	0.65	0.5910	0.0522	[0.5046, 0.7042]	73.6
ϕ	0.97	0.9615	0.0129	[0.9309, 0.9819]	1427.5
σ	0.15	0.1763	0.0258	[0.1351, 0.2416]	1722.8
ρ	**−0.6**	−0.6026	0.0979	[−0.7769, −0.3909]	2591.0
β	0.65	0.5743	0.0437	[0.5008, 0.6686]	384.1
ϕ	0.97	0.9619	0.0043	[0.9532, 0.9702]	1166.8
σ	0.15	0.1467	0.0041	[0.1390, 0.1549]	5928.3
ρ	**−0.9**	−0.8777	0.0185	[−0.9091, −0.8374]	8405.3
β	0.65	0.7085	0.0221	[0.6645, 0.7516]	510.0

Summary statistics for three simulation experiments using a variety of values of ρ. Sample size is 1000 throughout.

standard deviation, 0.83. For $\rho = -0.9$, the distribution is skewed to the left and the standard deviation is 1.73. This latter case is, however, somewhat special because in our analysis of real financial data we usually find that ρ is between -0.3 and -0.5.

3.2. Comparison with a single move sampler

In this subsection, we compare our mixture sampler with a single move sampler in which the $\{h_t\}$ and the parameters are sampled one-at-a-time, conditioned on the remaining values. KSC provide a similar comparison in the no leverage case.

The single move algorithm we employ in this experiment is essentially the algorithm developed by Jacquier et al. (2004). A straightforward modification of their algorithm is needed, however, to account for the different manner in which leverage is modeled; as discussed in the Introduction, in the latter paper, ε_t and η_{t-1} are correlated whereas in our case the dependence is in terms of ε_t and η_t.

For comparability, the data set and prior is the same as in the previous section although the burn-in is now 25,000 and the MCMC sample size is $M = 250,000$, both considerably larger than before because of the high serial correlation in the output from the single move method. The average acceptance rates of the latent variable h_t's in the Metropolis–Hastings algorithms are 71%, 55% and 29% for $\rho = -0.3, -0.6$ and -0.9. It appears that it gets increasingly difficult to find competitive proposal values in this method as the leverage correlation becomes more negative.

Our results from the single move method are given in Table 3 and Fig. 8. Clearly, the sample autocorrelation functions decay markedly slowly and the inefficiency factors in several cases are in the thousands.[2] Notice also a considerable worsening of the sample autocorrelations and inefficiency factors of (σ, ρ) for the cases $\rho = -0.6$ and $\rho = -0.9$.

---

[2]Due to the slowly declining serial correlations the inefficiency factors are based on 25,000 lags. In addition, even though this algorithm is about three times faster per cycle than our algorithm, the high-inefficiency factors
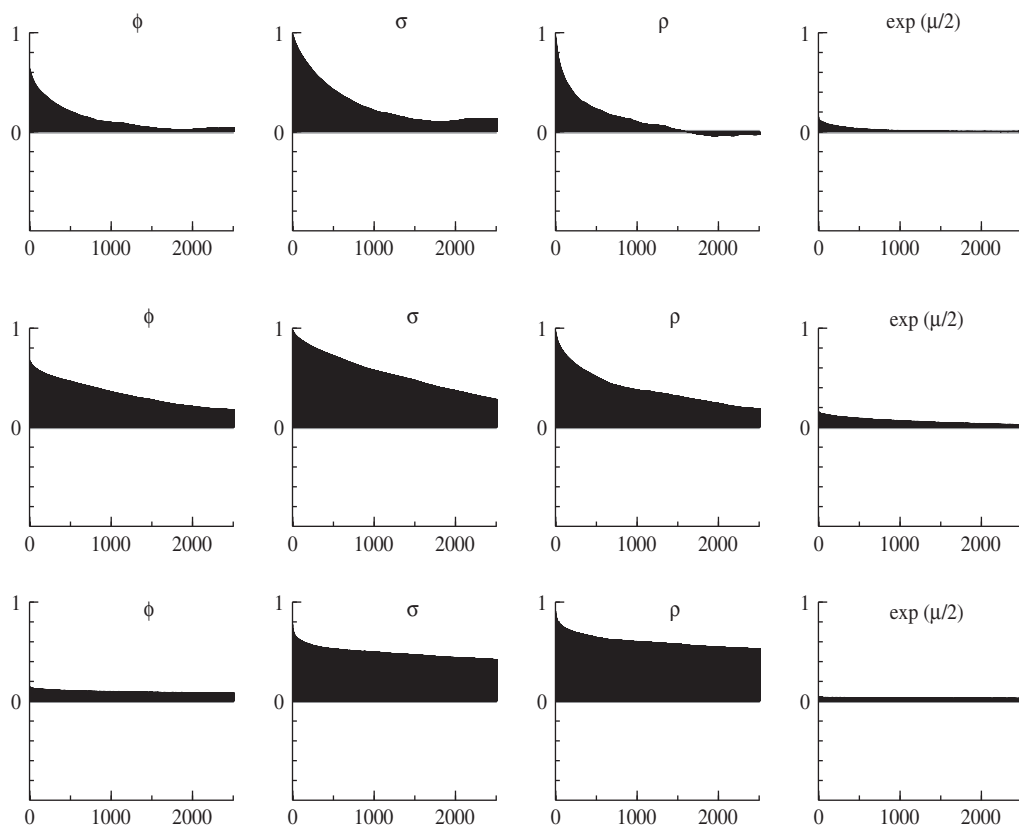
Fig. 8. Single move sampler. Sample autocorrelation functions of parameters for asymmetric stochastic volatility models with $\rho = -0.3, -0.6, -0.9$.

On comparing the corresponding results given in Table 2 it is apparent that the mixing properties of our method are strikingly better than those of the single move method.

4. Empirical example

In this section, we apply our approach to daily returns on the TOPIX (Tokyo Stock Price Index) which are calculated as the differences in the logarithm of the daily closing value of TOPIX. The sample period is from January 5, 1998 through December 30, 2002 leading to a sample of 1232 days on which the market was open. Table 4 gives the summary statistics of the data. The mean and standard deviation of the returns are -0.026 and 1.284, respectively. In addition, there were 602 days when $y_t > 0$ and 630 days when $y_t \leqslant 0$.

(footnote continued)

imply that the single move method must be run for a substantially longer period to generate the same effective sample size.

Table 4
Summary statistics for TOPIX return data (log-difference)

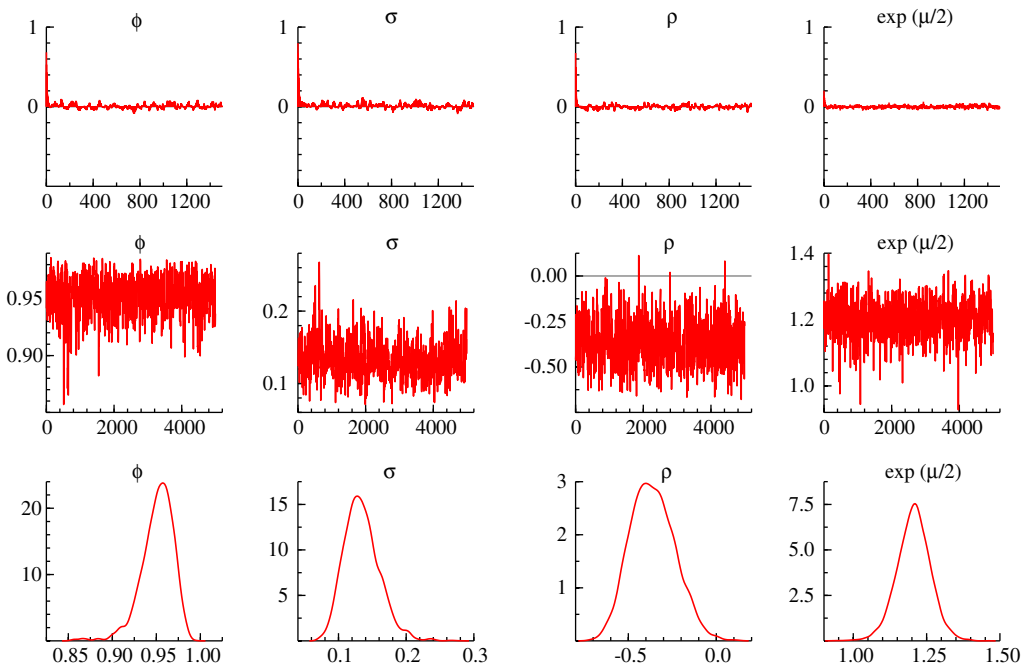| Obs.	Mean	St. dev.	Max	Min	Pos(+)	Neg(−)
1232	−0.0255	1.2839	5.3749	−5.6819	602	630

TOPIX (1998/1/5–2002/12/30).

Fig. 9. Estimation result for TOPIX. Sample autocorrelation functions, sample paths of MCMC output and estimated posterior densities.

We use the same prior distribution given in Section 3.1, while again the initial 500 MCMC iterations are discarded and the following 5000 values are recorded. Fig. 9 shows the sample autocorrelation functions, the sample paths and the posterior densities of $(\phi, \sigma, \rho, \beta = \exp(\mu/2))$. The sample autocorrelations decay quickly and the output mixes well.

Table 5 shows the estimated posterior means, standard deviations, the 95% credible intervals and inefficiency factors.[3] These factors are small, suggesting that the posterior moments of the parameters could be estimated with relatively economical sample sizes. The posterior means of ϕ, σ, β are $0.95, 0.13$ and 1.21, respectively, which are typical of the values found in prior analysis of these data.

---

[3]The inefficiency factors are based on 100 lags of the autocorrelation functions and do not change materially when more lags (say a 1000) are involved.

Table 5
Estimation result for TOPIX

Parameter	Unweighted		Weighted		95% interval	Ineff	Posterior correlation			
	Mean	St. dev.	Mean	St. dev.						
ϕ	0.9511	0.0185	0.9512	0.0185	[0.908, 0.980]	9.3	1	−.66	−.30	−.06
σ	0.1343	0.0262	0.1341	0.0264	[0.091, 0.193]	13.0		1	.19	−.08
ρ	−0.3617	0.1265	−0.3578	0.1257	[−0.593,−0.107]	6.8			1	.13
β	1.2056	0.0573	1.2052	0.0571	[1.089, 1.318]	2.7				1

Sample size was 5000, based on 5500 MCMC draws, discarding the first 500. Posterior correlation denotes the posterior correlation matrix.
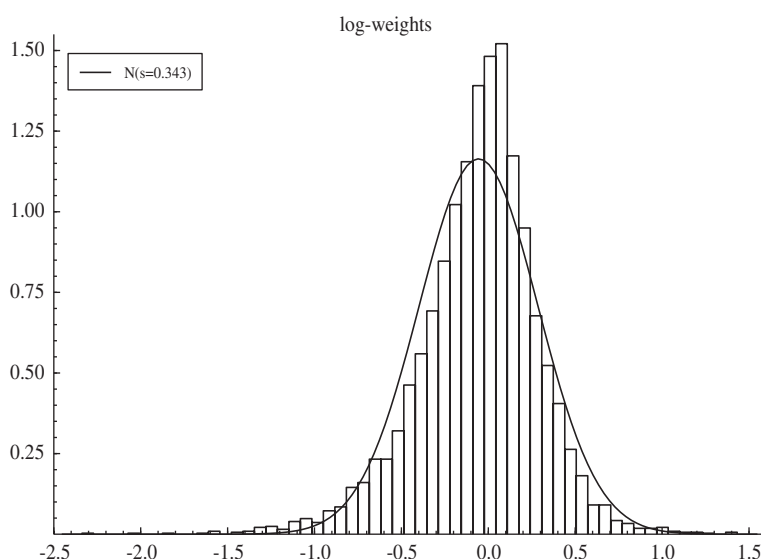

Fig. 10. Sampling result of log-weights $\log(w_k \times M)$ for the TOPIX series. Shows histogram and fitted normal density.

The posterior mean of ρ is -0.36 and negative as expected, suggesting the presence of leverage. Since its 95% credible interval is $[-0.59, -0.11]$, the posterior probability that ρ is negative is greater than 0.95.

Fig. 10 shows the distributions of $\log(w_k \times M)$ for the proposed sampler. As in the illustrative examples when $\rho = -0.3$, the log weights are concentrated around zero, and the standard deviation is 0.34. In contrast, KSC in the context of the basic SV model report a standard deviation of around 1.0 in their analysis of similar data. This shows that our overall approach is well behaved.

Table 5 shows the effect of reweighting on inference. We see that reweighting has a small effect on the estimates of the posterior mean. In the first column of Table 6, we present the log of the marginal likelihood for the SV model in the $\rho = 0$ case. The marginal likelihood here and elsewhere was calculated by the method of Chib (1995), alongside the

Table 6
Marginal likelihood estimation by the Chib (1995) method for the TOPIX data

	SV	SV-t	ASV	ASV-t	ASV-g	SP
Likelihood ordinate	−2033.83	−2033.21	−2029.55	−2029.12	−2029.20	−2029.68
(S.E.)	(0.16)	(0.19)	(0.10)	(0.12)	(0.17)	(0.13)
Prior ordinate	3.75	0.16	3.09	0.99	3.38	7.84
Posterior ordinate	8.87	5.17	10.25	8.46	12.50	14.81
(S.E.)	(0.02)	(0.20)	(0.02)	(0.08)	(0.04)	(0.04)
Marg likelihood	−2038.95	−2038.21	−2036.71	−2036.59	−2038.31	−2036.65
(S.E.)	(0.16)	(0.28)	(0.10)	(0.14)	(0.17)	(0.14)

All values are in the natural-log scale. SV, SV-t and SV-g denote the SV models with Gaussian, student-t and normal log-normal errors. ASV allows $\rho \neq 0$. SP denotes superposition model.

modification for Metropolis–Hastings chains given in Chib and Jeliazkov (2001). The log-likelihood ordinate, which is an input into this computation, was calculated by the particle filter method with $I = 10,000$ using (15). The marginal likelihood of this model can be compared to the SV model with leverage given in the third column of the table. The results show that the model with leverage improves the likelihood, evaluated at the posterior mean, by around 4 at the cost of a single parameter. On the basis of the log marginal likelihood, which contains an automatic penalty for model complexity, we find that the log of the Bayes factor in favor of the leverage model is around 2.

5. More general dynamics

5.1. Framework

Precisely the same methods can be used to handle flexible models of the type

$$y_t = \varepsilon_t \exp(h_t/2), \quad h_t = \mathbf{z}_t' \boldsymbol{\alpha}_t, \tag{16}$$

$$\boldsymbol{\alpha}_{t+1} = \mathbf{b}_t + \mathbf{T}_t \boldsymbol{\alpha}_t + \boldsymbol{\eta}_t, \tag{17}$$

where z_t, b_t and T_t are non-stochastic processes, potentially dependent on some parameter θ and $\boldsymbol{\alpha}_t$ is r-dimensional. We assume that

$$\begin{pmatrix} \varepsilon_t \\ \boldsymbol{\eta}_t \end{pmatrix} \sim \mathcal{N}_{r+1}(0, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & \boldsymbol{\sigma}' \\ \boldsymbol{\sigma} & \boldsymbol{\Omega} \end{pmatrix}. \tag{18}$$

In order to simplify the exposition assume that $\boldsymbol{\Omega}$ is non-singular. In principle this framework can allow general forms of leverage wherein the dependence between ε_t and the elements of $\boldsymbol{\eta}_t$ is allowed to vary over the individual elements.

This structure implies that

$$\boldsymbol{\eta}_t | d_t, \varepsilon_t^* \sim \mathcal{N}(d_t \boldsymbol{\sigma} \exp(\varepsilon_t^*/2), \boldsymbol{\Omega} - \boldsymbol{\sigma}\boldsymbol{\sigma}').$$

Then, if we use our bivariate mixture approximation, we get that

$$\left\{ \begin{pmatrix} \varepsilon_t^* \\ \boldsymbol{\eta}_t \end{pmatrix} \middle| d_t, s_t = j \right\} \stackrel{L}{=} \begin{pmatrix} m_j + v_j z_t \\ d_t \boldsymbol{\sigma}(a_j + b_j v_j z_t) \exp(m_j/2) + (\boldsymbol{\Omega} - \boldsymbol{\sigma}\boldsymbol{\sigma}')^{1/2} \mathbf{z}_t^* \end{pmatrix},$$

where $(z_t, \mathbf{z}_t^*)' \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}_{r+1}(\mathbf{0}, \mathbf{I})$. Therefore, except for an increase in the dimension of the problem, this extension raises no new issues for our MCMC implementation.

5.2. Example: superposition model

Suppose that

$$h_t = \sum_{i=1}^{r} \alpha_{i,t},$$

where

$$\boldsymbol{\alpha}_{t+1} = \begin{pmatrix} \mu \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} \phi_1 & 0 & 0 & \cdots & 0 \\ 0 & \phi_2 & 0 & \cdots & 0 \\ 0 & 0 & \phi_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \phi_r \end{pmatrix} \left(\boldsymbol{\alpha}_t - \begin{pmatrix} \mu \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right) + \boldsymbol{\eta}_t,$$

and

$$\begin{pmatrix} \varepsilon_t \\ \boldsymbol{\eta}_t \end{pmatrix} \sim \mathcal{N}_{r+1}(\mathbf{0}, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \begin{pmatrix} 1 & \rho_1\sigma_1 & \rho_2\sigma_2 & \cdots & \rho_r\sigma_r \\ \rho_1\sigma_1 & \sigma_1^2 & 0 & \cdots & 0 \\ \rho_2\sigma_2 & 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_r\sigma_r & 0 & & & \sigma_r^2 \end{pmatrix}.$$

Then the log-volatility is made up of the sum of independent autoregressions, each with a different persistence level and degree of leverage. Superposition models of this type have become popular in financial econometrics as they are more general than empirically limiting Markov volatility models while close to corresponding continuous time models (Shephard, 1996; Engle and Lee, 1999; Barndorff–Nielsen and Shephard, 2001; Chernov et al., 2003). It is easy to check that for Σ to be positive definite we need $\sum_{i=1}^{r} \rho_i^2 < 1$.

Column 6 of Table 6 shows that for the TOPIX data set adding a second volatility component to the model has a modest effect on the fit of the model as measured by the log marginal likelihood. The log of Bayes factors are also given in Table 7. These results are based on a prior where $(\phi_2 + 1)/2 \sim \text{Beta}(10, 10)$ with the side constraint that $\phi_2 < \phi_1$. Further, we assume $(\rho_2 + 1)/2 \sim \text{Beta}(10, 10)$ with the constraint that $0 < \rho_1^2 + \rho_2^2 < 1$. Finally, $\sigma_2^{-2} \sim \text{Gamma}(5/2, 0.05/2)$. To generate a candidate with such constraints, we draw a candidate from the untruncated Gaussian proposal using a transformation $\theta_1 = \log(1 + \phi_1) - \log(1 - \phi_1)$, $\theta_2 = \log(1 + \phi_2) - \log(\phi_1 - \phi_2)$, $\theta_3 = \log \sigma_1^2$, $\theta_4 = \log \sigma_2^2$, $\theta_5 = \log(1 + \rho_1) - \log(1 - \rho_1)$ and $\theta_6 = \log(\sqrt{1 - \rho_1^2} + \rho_2) - \log(\sqrt{1 - \rho_1^2} - \rho_2)$. Even

Table 7
Bayes factors for the TOPIX data

	SV-t	ASV	ASV-t	ASV-g	SP
SV	−0.32	−0.97	−1.02	−0.28	−1.00
SV-t		−0.65	−0.70	−0.04	−0.68
ASV			−0.05	0.69	−0.03
ASV-t				0.75	−0.03
ASV-g					−0.72

The figures are log (base 10) of Bayes factors for the row model against the column model.

though the log-likelihood, evaluated at the posterior mean of the parameters, is higher than the one component model, the new model has three extra parameters σ_2, ρ_2 and ϕ_2, which is obviously penalized in the marginal likelihood computation.

5.3. Example: heavy-tailed error distribution

Many writers have followed Harvey et al. (1994) in extending the SV model to allow for heavier tailed returns. For example, Chib et al. (2002) extended the basic KSC approach by letting

$$y_t = \sqrt{\lambda_t}\varepsilon_t \exp(h_t/2), \tag{19}$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \eta_t, \quad t = 0, 1, \ldots, n, \tag{20}$$

where λ_t is an i.i.d. scale mixture variable and $\lambda_t \perp\!\!\!\perp (\varepsilon_t, \eta_t)$ ($\perp\!\!\!\perp$ denotes probabilistic independence). This is relevant empirically and also corresponds to the literature on time-change Lévy processes and Lévy-based SV models (Carr et al., 2003; Carr and Wu, 2004; Cont and Tankov, 2004). Papers on various inferential aspects of these models include Barndorff-Nielsen and Shephard (2006) and Li et al. (2004). In this subsection we will assume that

$$\log \lambda_t \sim \mathcal{N}(-0.5\tau^2, \tau^2),$$

in which case $\lambda_t^{1/2}\varepsilon_t$ has a normal log-normal distribution. This specification is closed in the empirical work by assuming that $\tau^2 \sim \text{Gamma}(1, 1)$.

The above model fits into the framework put forward in (16)–(18) by writing

$$y_t = \varepsilon_t \exp(h_t/2), \tag{21}$$

$$h_t = h_{t+1}^* + \lambda_t, \tag{22}$$

$$h_{t+1}^* = \mu + \phi(h_t^* - \mu) + \eta_t, \quad t = 0, 1, \ldots, n, \tag{23}$$

where

$$\begin{pmatrix} \varepsilon_t \\ \eta_t \\ \lambda_t \end{pmatrix} \sim \mathcal{N}_3 \left(\begin{pmatrix} 0 \\ 0 \\ -0.5\tau^2 \end{pmatrix}, \begin{pmatrix} 1 & \rho\sigma & 0 \\ \rho\sigma & \sigma^2 & 0 \\ 0 & 0 & \tau^2 \end{pmatrix} \right).$$

Therefore, this extension again raises no new inferential issues.

Table 6 gives results for the three different heavier tailed specifications. In the second (fourth) column we report the results when $\rho = 0$ ($\rho \neq 0$) and $\sqrt{\lambda_t}\varepsilon_t$ follows a student-t distribution with v degrees of freedom, where $v \sim \text{Gamma}(16, 0.8)$. The fifth column reports the results for the Gaussian scale mixture SV model with leverage. The fit of the second model is better than the basic model, but not over the leverage model. Overall, however, the simple Gaussian SV model with leverage is preferred for these data.

6. Conclusion

In this paper, we have extended the KSC approach to SV models with leverage. This starts with the joint distribution of $\log \varepsilon_t^2, \eta_t | sign(y_t)$ which is then approximated by a suitably constructed ten-component mixture of bivariate normal distributions. We show that this method, which is easy to implement and produces output that mixes well, effectively solves the problems of fitting SV models with leverage. We also show how our new method can be further extended to cover even more general SV models such as those with heavy-tailed distributions and superposition effects. In each case, our algorithm performs as well as the original KSC algorithm but is applicable under wider conditions. We also discuss the computation of the marginal likelihood and Bayes factors and provide an empirical analysis of real Japanese stock return data where the SV model with leverage is preferred over competing models.

Acknowledgments

References

Barndorff-Nielsen, O.E., Shephard, N., 2001. Non-Gaussian Ornstein–Uhlenbeck-based models and some of their uses in financial economics (with discussion). Journal of the Royal Statistical Society, Series B 63, 167–241.

Barndorff-Nielsen, O.E., Shephard, N., 2006. Impact of jumps on returns and realised variances: econometric analysis of time-deformed Lévy processes. Journal of Econometrics 131, 217–252.

Black, F., 1976. Studies of stock price volatility changes. Proceedings of the Business and Economic Statistics Section, American Statistical Association, pp. 177–181.

Carr, P., Wu, L., 2004. Time-changed Lévy processes and option pricing. Journal of Financial Economics 71, 113–141.

Carr, P., Geman, H., Madan, D.B., Yor, M., 2003. Stochastic volatility for Lévy processes. Mathematical Finance 13, 345–382.

Carter, C.K., Kohn, R., 1994. On Gibbs sampling for state space models. Biometrika 81, 541–553.

Chernov, M., Gallant, A.R., Ghysels, E., Tauchen, G., 2003. Alternative models of stock price dynamics. Journal of Econometrics 116, 225–257.

Chib, S., 1995. Marginal likelihood from the Gibbs output. Journal of the American Statistical Association 90, 1313–1321.

Chib, S., 2001. Markov chain Monte Carlo methods: computation and inference. In: Heckman, J.J., Leamer, E. (Eds.), Handbook of Econometrics, vol. 5. North-Holland, Amsterdam, pp. 3569–3649.

Chib, S., Greenberg, E., 1994. Bayes inference for regression models with ARMA(p, q) errors. Journal of Econometrics 64, 183–206.

Chib, S., Greenberg, E., 1995. Understanding the Metropolis–Hastings algorithm. The American Statistician 49, 327–335.

Chib, S., Jeliazkov, I., 2001. Marginal likelihood from the Metropolis–Hastings output. Journal of the American Statistical Association 96, 270–281.

Chib, S., Nardari, F., Shephard, N., 2002. Markov chain Monte Carlo methods for generalized stochastic volatility models. Journal of Econometrics 108, 281–316.

Cont, R., Tankov, P., 2004. Financial Modelling with Jump Processes. Chapman & Hall, London.

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

Diebold, F.X., Gunther, T.A., Tay, T.S., 1998. Evaluating density forecasts with applications to financial risk management. International Economic Review 39, 863–883.

Doornik, J.A., 2001. Ox: Object Oriented Matrix Programming, 3.0. Timberlake Consultants Press, London.

Doucet, A., de Freitas, N., Gordon, N.J. (Eds.), 2001. Sequential Monte Carlo Methods in Practice. Springer, New York.

Durbin, J., Koopman, S.J., 2001. Time Series Analysis by State Space Methods. Oxford University Press, Oxford.

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

Elerian, O., Chib, S., Shephard, N., 2001. Likelihood inference for discretely observed non-linear diffusions. Econometrica 69, 959–993.

Engle, R.F., Lee, G.G.J., 1999. A permanent and transitory component model of stock return volatility. In: Engle, R.F., White, H. (Eds.), Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger. Oxford University Press, Oxford, pp. 475–497.

Eraker, B., 2001. Markov chain Monte Carlo analysis of diffusion models with application to finance. Journal of Business and Economic Statistics 19, 177–191.

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

Geweke, J., 1992. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints. Computing Science and Statistics: Proceedings of the 23rd Symposium, pp. 571–578.

Ghysels, E., Harvey, A.C., Renault, E., 1996. Stochastic volatility. In: Rao, C.R., Maddala, G.S. (Eds.), Statistical Methods in Finance. North-Holland, Amsterdam, pp. 119–191.

Harvey, A.C., 1989. Forecasting, Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.

Harvey, A.C., Shephard, N., 1996. The estimation of an asymmetric stochastic volatility model for asset returns. Journal of Business and Economic Statistics 14, 429–434.

Harvey, A.C., Ruiz, E., Shephard, N., 1994. Multivariate stochastic variance models. Review of Economic Studies 61, 247–264.

Jacquier, E., Polson, N.G., Rossi, P.E., 2004. Stochastic volatility models: univariate and multivariate extensions. Journal of Econometrics 122, 185–212.

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

Koopman, S.J., Shephard, N., Doornik, J.A., 1999. Statistical algorithms for models in state space using SsfPack 2.2. The Econometrics Journal 2, 107–166.

Li, H., Wells, M., Yu, L., 2004. A MCMC analysis of time-changed Lévy processes of stock returns, unpublished paper. Johnson Graduate School of Management, Cornell University, Ithaca, U.S.A.

Mahieu, R., Schotman, P., 1998. An empirical application of stochastic volatility models. Journal of Applied Econometrics 16, 333–359.

Nelson, D.B., 1988. The time series behaviour of stock market volatility and returns, Unpublished Ph.D., MIT, Cambridge, MA.

Nelson, D.B., 1991. Conditional heteroskedasticity in asset pricing: a new approach. Econometrica 59, 347–370.

Pitt, M.K., Shephard, N., 1999. Filtering via simulation: auxiliary particle filter. Journal of the American Statistical Association 94, 590–599.

Primiceri, G., 2005. Time varying structural vector autoregressions and monetary policy. Review of Economic Studies 72, 821–852.

Roberts, G.O., Stramer, O., 2001. On inference for nonlinear diffusion models using the Hastings–Metropolis algorithms. Biometrika 88, 603–621.

Rosenblatt, M., 1952. Remarks on a multivariate transformation. Annals of Mathematical Statistics 23, 470–472.

Shephard, N., 1994. Partial non-Gaussian state space. Biometrika 81, 115–131.

Shephard, N., 1996. Statistical aspects of ARCH and stochastic volatility. In: Cox, D.R., Hinkley, D.V., Barndorff-Nielsen, O.E. (Eds.), Time Series Models in Econometrics, Finance and Other Fields. Chapman & Hall, London, pp. 1–67.

Shephard, N., 2005. Stochastic Volatility: Selected Readings. Oxford University Press, Oxford.

Smith, J.Q., 1985. Diagnostic checks of non-standard time series models. Journal of Forecasting 4, 283–291.

Stroud, J.R., Muller, P., Polson, N.G., 2003. Nonlinear state-space models with state-dependent variances. Journal of the American Statistical Association 98, 377–386.

West, M., Harrison, J., 1997. Bayesian Forecasting and Dynamic Models, second ed. Springer, New York.

Yu, J., 2005. On leverage in a stochastic volatility model. Journal of Econometrics 127, 165–178.