Stochastic volatility with leverage: fast likelihood inference

Yasuhiro Omori

Faculty of Economics, University of Tokyo, Tokyo 113-0033, Japan
omori@e.u-tokyo.ac.jp

Siddhartha Chib

Olin School of Business, Washington University, St Louis, USA
chib@olin.wustl.edu

Neil Shephard

Nuffield College, Oxford OX1 1NF, UK
neil.shephard@nuf.ox.ac.uk

Jouchi Nakajima

Faculty of Economics, University of Tokyo, Tokyo 113-0033, Japan
n_sophia@f3.dion.ne.jp

August 22, 2004

Abstract

Kim, Shephard, and Chib (1998) provided a Bayesian analysis of stochastic volatility models based on a fast and reliable Markov chain Monte Carlo (MCMC) algorithm. Their method ruled out the leverage effect, which is known to be important in applications. Despite this, their basic method has been extensively used in the financial economics literature and more recently in macroeconometrics. In this paper we show how the basic approach can be extended in a novel way to stochastic volatility models with leverage without altering the essence of the original approach. Several illustrative examples are provided.

Key words: Leverage effect, Markov chain Monte Carlo, Mixture sampler, Stochastic volatility, Stock returns.

1 Introduction

The stochastic volatility (SV) model, a specific non-linear state space model, has been the subject of considerable attention in the econometric and statistical literatures because of the many interesting challenges it raises for estimation and inference (see, for example, the reviews in Ghysels, Harvey, and Renault (1996) and Shephard (2004)). It is also an important model because of its significance in financial applications where it has been used to understand time-varying volatility in high frequency asset returns. Kim, Shephard, and Chib (1998) developed an approach for fitting and comparing the SV model that has been extensively employed (for

example, Mahieu and Schotman (1998), Primiceri (2005) and Stroud, Muller, and Polson (2003)). This approach relies on a Bayesian Markov chain Monte Carlo (MCMC) sampling method to summarize the posterior distribution of the model parameters and the latent time varying volatilities. The approach is highly efficient in terms of the common metrics (for example inefficiency factors) that are used to study the mixing properties of sequences that are produced by the sampling algorithm. The Kim, Shephard and Chib approach, which is based on a certain approximation to a log-chisqured distribution was developed for SV models without leverage (correlation between the errors in the measurement and evolution equations). Leverage, however, is known to be important in applications. The goal of this article, therefore, is to develop the corresponding inferential methodology for SV models with leverage without altering the essence of the Kim, Shephard and Chib approach.

The simplest model we study is the well known log-normal *stochastic volatility* (SV) model given by

$$
\begin{aligned}
y_t &= \epsilon_t \exp(h_t/2), \\
h_{t+1} &= \mu + \phi(h_t - \mu) + \eta_t, \quad t = 0, 1, \ldots, n,
\end{aligned}
\tag{1}
$$

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

$$
\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(0, \Sigma\right), \quad \text{and} \quad \Sigma = \begin{pmatrix} 1 & \rho\sigma \\ \rho\sigma & \sigma^2 \end{pmatrix}.
$$

The parameter ρ measures the leverage effect. The leverage effect refers to the increase in volatility following a drop in equity returns and, in this model, corresponds to a negative correlation between ϵ_t and η_t (e.g. Black (1976), Nelson (1991) and Yu (2004)). The latter reference also gives a discussion of various alternative MCMC schemes put forward in the literature.

In the Kim, Shephard, and Chib (1998) approach the distribution of $\log \epsilon_t^2$ is approximated by a mixture of seven Gaussian distributions such that the first four moments of both densities are equal. They then wrote the mixture distribution hierarchically in terms of a latent component indicator (one for each time period) and conducted the MCMC sampling on the posterior distribution of the latent component indicators, the vector of latent volatilities $h = \{h_t\}_{t=1}^n$ and the parameters. One key feature of their method is that it permits for the joint sampling of h conditioned on the latent component indicators thus leading to posterior draws that mix better than alternative approaches that rely on one-at-a time sampling of the volatilities. The sampling is finished by a reweighting step to overcome any error arising from the mixture approximation. Although this approach has proved valuable and has formed the basis of many subsequent studies it was designed only for the case where $\rho = 0$. In this paper we show that the approach can be extended in a novel way to SV models with leverage by starting with the joint distribution of $\log \epsilon_t^2, \eta_t | sign(y_t)$ and approximating this distribution by a suitably constructed ten-component mixture of normal distributions. We discuss how this is done and show that it 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).

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 working of this method, while in Section 4 we illustrate the methods on some data from the Japanese stock market. In Section 5 we show that the analysis extends to much wider classes of SV models, while Section 6 concludes.

2 Efficient auxiliary mixture sampler

2.1 Reformulation in the no leverage case

Following Nelson (1988) and Harvey and Shephard (1996), without loss of information we replace (1) by (d_t, y_t^*), the bivariate observations, where

$$d_t = sign(y_t) = I(\epsilon_t > 0) - I(\epsilon_t \leq 0), \tag{2}$$
$$y_t^* = \log y_t^2 = h_t + \xi_t, \tag{3}$$

and

$$\xi_t = \log \epsilon_t^2.$$

Thus

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

In the case where $\rho = 0$ the signs of $y = (y_1, ..., y_n)'$ are independent of $y^* = (y_1^*, ..., y_n^*)'$ and we can neglect $d = (d_1, ..., d_n)'$. This greatly simplifies the development of an inferential methodology because y^* is a linear process (e.g. Harvey, Ruiz, and Shephard (1994)) with an i.i.d. error ξ_t in (3) that follows a $\log \chi_1^2$ density

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

Kim, Shephard, and Chib (1998) introduced the idea of accurately approximating this distribution by a mixture of normal distributions, selected to ensure that moments up to a certain order are equal. In the Bayesian MCMC context, the resulting approximation error can be corrected by reweighting the sequences sampled from the posterior distribution, as we discuss below in subsection 2.3. The mixture approximation has the form of

$$g(\xi_t) = \sum_{i=1}^{K} p_i f_N(\xi_t | m_i, v_i^2), \quad \xi_t \in R \tag{4}$$

where $f_N(\xi_t | m_i, v_i^2)$ denotes the density function of a normal distribution with mean m_i and variance v_i^2. The constants m_i and v_i^2 were determined by Kim, Shephard, and Chib (1998) on the basis of $K = 7$ components. These values are reproduced in the first block of columns in Table 1. In this paper we have favoured a tighter approximation, based on $K = 10$, which is given in the second block of Table 1.

Figure 1 shows the differences between the approximate and the true densities of the $\log \chi_1^2$

	KSC			$K = 10$				
i	p_i	m_i	v_i^2	p_i	m_i	v_i^2	a_i	b_i
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

Table 1: *Selection of* $(p_i, m_i, v_i^2, a_i, b_i)$. *Left hand side was determined by Kim, Shephard and Chib, the ones on the right hand side are new and represent a better approximation.*

and $\sqrt{\chi_1^2}$ (for the range from the 1st percentile to the 99th percentile) for the two mixtures. We can see that the move to $K = 10$ components improves the approximation.
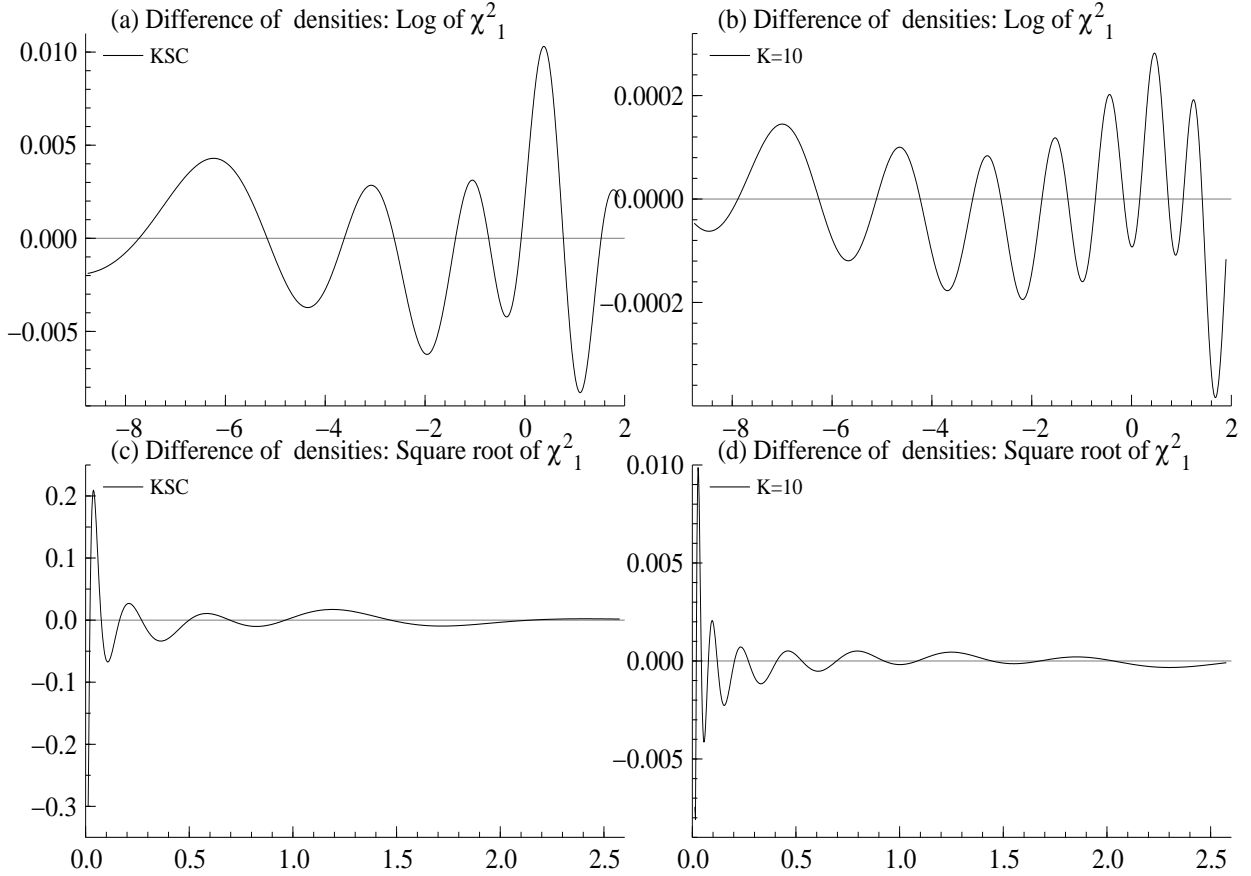

Figure 1: *The difference between the approximate and the true densities (for the range from the 1st percentile to the 99th percentile). The* $\log \chi_1^2$ *density (top) and the* $\sqrt{\chi_1^2}$ *density (bottom).*

2.2 Reformulation in general case

Now consider the general case of $\rho \neq 0$. The main complication is that d_t is not ignorable because, for example,

$$\eta_t | d_t, \xi_t \sim \mathrm{N}\left(d_t \rho \sigma \exp\left(\xi_t/2\right), \sigma^2(1 - \rho^2)\right). \tag{5}$$

Another complication is that ξ_t now enters both (3) and (5). To extend the approach of Kim, Shephard, and Chib (1998) we consider the novel strategy of approximating the bivariate conditional density of

$$\xi_t, \eta_t | d_t$$

This bivariate density is key as

$$(\xi_t, \eta_t | d_t) \perp\!\!\!\perp (\xi_s, \eta_s | d_s)$$

for all $t \neq s$, where $\perp\!\!\!\perp$ denotes probabilistic independence. Clearly

$$
\begin{aligned}
f(\xi_t, \eta_t | d_t) &= f(\xi_t | d_t) f(\eta_t | \xi_t, d_t) \\
&= f(\xi_t) f(\eta_t | \xi_t, d_t). \tag{6}
\end{aligned}
$$

Our idea now is to maintain the mixture approximation $g(\xi_t)$ given in (4) and to consider the approximation

$$g(\xi_t, \eta_t | d_t) = \sum_{i=1}^{K} p_i f_N(\xi_t | m_i, v_i^2) f_N\left[\eta_t | d_t \rho \sigma \exp(m_i/2)\left\{a_i + b_i\left(\xi_t - m_i\right)\right\}, \sigma^2(1 - \rho^2)\right], \tag{7}$$

where (a_i, b_i) are known constants. In other words, we utilize a mixture of bivariate Gaussian densities to approximate the distribution of $\xi_t, \eta_t | d_t$. The remaining question is the determination of the density

$$f_N\left[\eta_t | d_t \rho \sigma \exp(m_i/2)\left\{a_i + b_i\left(\xi_t - m_i\right)\right\}, \sigma^2(1 - \rho^2)\right]$$

to well approximate the density of $\eta_t | d_t, \xi_t$ in the i-th component of the mixture distribution. Due to the form of (5) this amounts to approximating

$$\exp\left(\xi_t/2\right)\exp(-m_i/2)$$

by

$$(a_i + b_i\left(\xi_t - m_i\right)),$$

given

$$\xi_t \sim N(m_i, v_i^2).$$

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. We find the values of (a_i, b_i) by considering the mean square norm and setting

$$(a_i, b_i) = \arg\min_{a,b} E\{\exp\left(\xi_t/2\right)\exp(-m_i/2) - a - b\left(\xi_t - m_i\right)\}^2, \quad \xi_t \sim N(m_i, v_i^2), \quad i = 1, 2, ..., K.$$

By calculation we find that the solutions to this minimization problem are given by

$$
\begin{aligned}
a_i &= \exp(v_i^2/8), \\
b_i &= E\{z_t \exp(v_i z_t/2)\} = \frac{1}{2}\exp\left(\frac{v_i^2}{8}\right), \quad i = 1, 2, ..., K.
\end{aligned}
$$

The implied values of (a_i, b_i) are given in Table 1.

Remark 1 *The key question is how well (7) approximate (6). We give results for $\rho = -0.3, -0.6$ and -0.9. Figure 2 shows f and g for $\eta_t | \xi_t, d_t = 1$ evaluated with ξ_t set at its 25th, 50th and 75th percentiles. Likewise Figure 1 shows f and g for $\xi_t | \eta_t, d_t = 1$ evaluated with $\eta_t = -0.67\sigma$,*
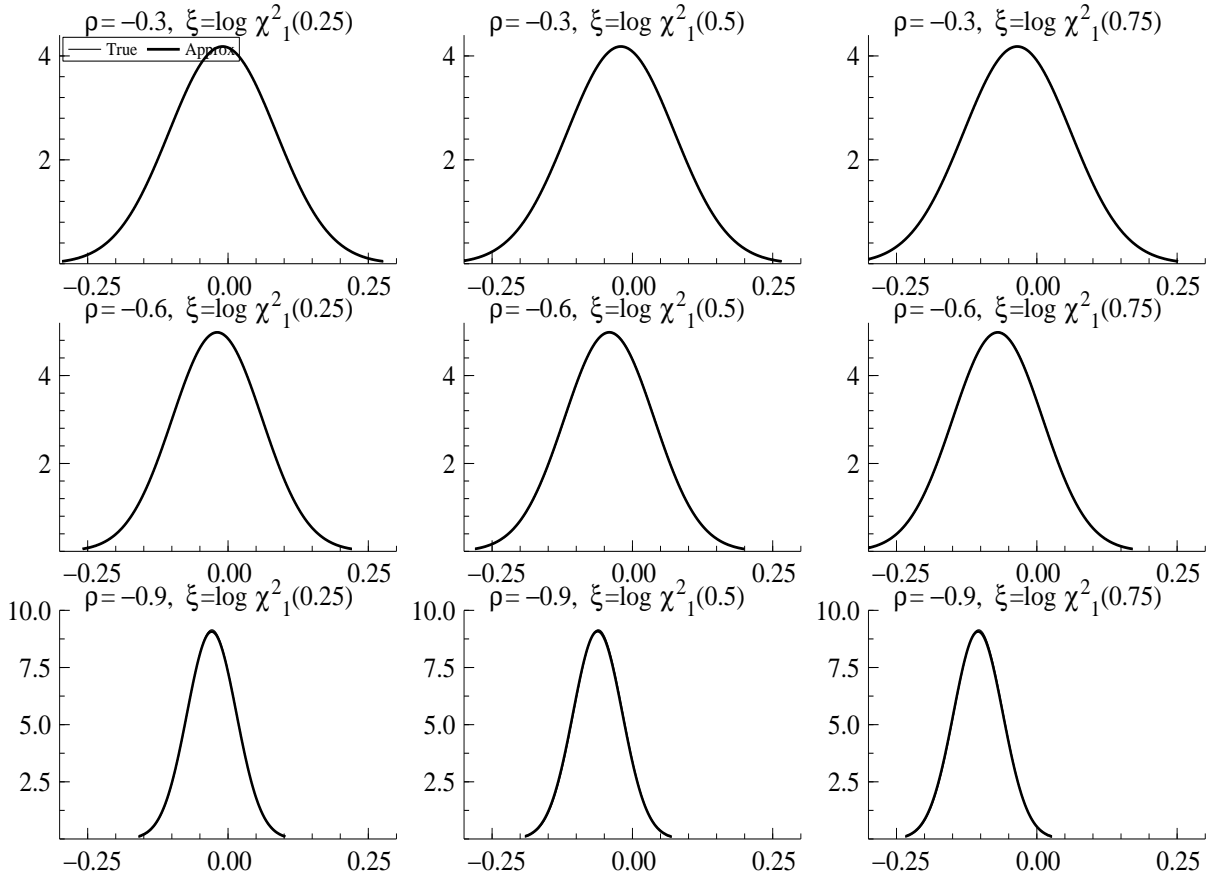

Figure 2: *The conditional density of η_t given $d_t = 1$ and $v_t = \log \chi_1^2(0.25)$, $\log \chi_1^2(0.5)$, $\log \chi_1^2(0.75)$ (left, middle, right) for $\rho = -0.3, -0.6, -0.9$ (top, middle, bottom).*

0, 0.67σ. 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, Figure 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.

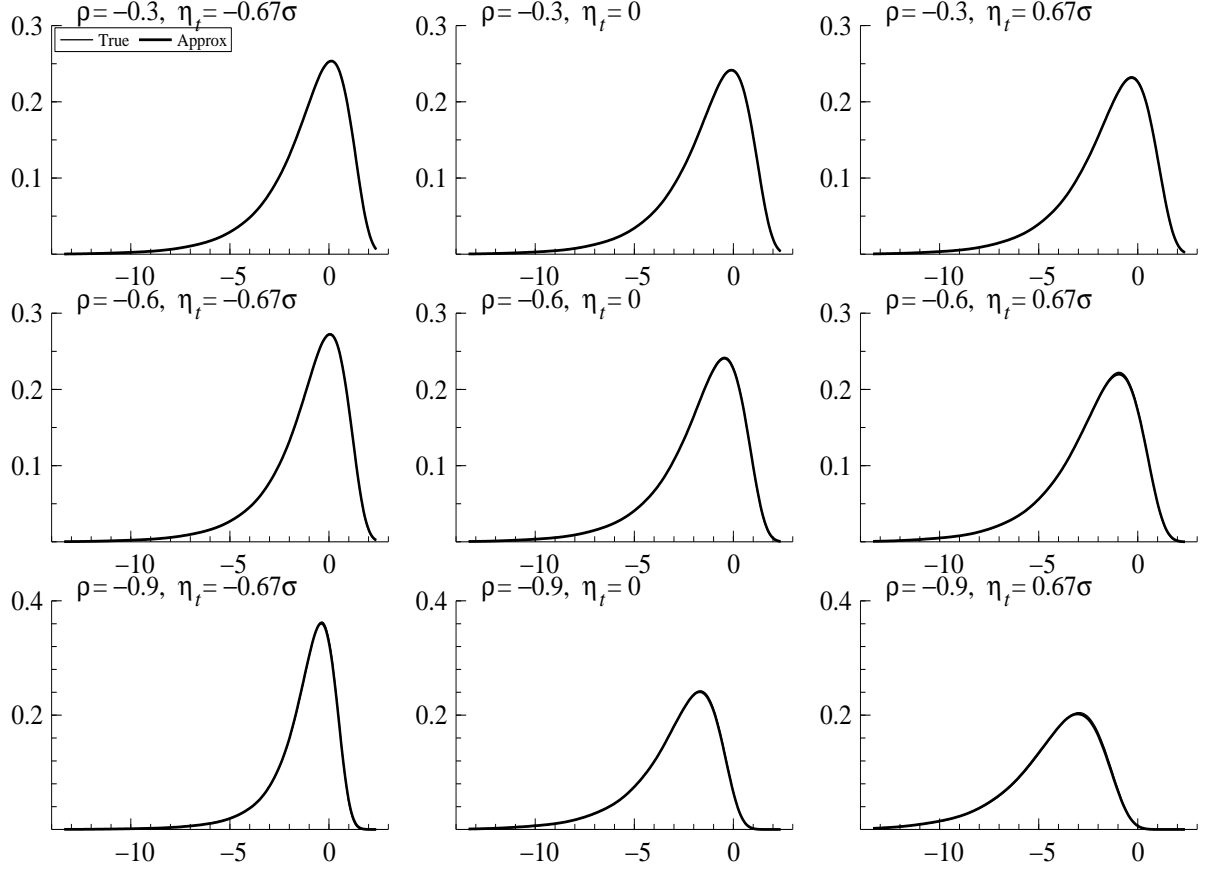Figure 3: *The conditional density of v_t given $d_t = 1$ and $\eta_t = -0.67\sigma$, 0, 0.67σ (left, middle, right) for $\rho = -0.3$, -0.6, -0.9 (top, middle, bottom). The value of σ is set to 1 in this example.*

2.3 MCMC algorithm

2.3.1 Broad principles

The SV model can be expressed as

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

Now on using the mixture approximation (7) to the density $\xi_t, \eta_t | d_t$ and introducing the mixture component indicator $s_t \in \{1, 2, ..., K\}$ we have that

$$
\left\{ \begin{pmatrix} \xi_t \\ \eta_t \end{pmatrix} | d_t, s_t = i \right\} \overset{L}{=} \begin{pmatrix} m_i + v_i z_t \\ d_t \rho \sigma (a_i + b_i v_i z_t) \exp(m_i/2) + \sigma\sqrt{1-\rho^2} z_t^* \end{pmatrix}, \quad \begin{pmatrix} z_t \\ z_t^* \end{pmatrix} \overset{i.i.d.}{\sim} N(0, I).
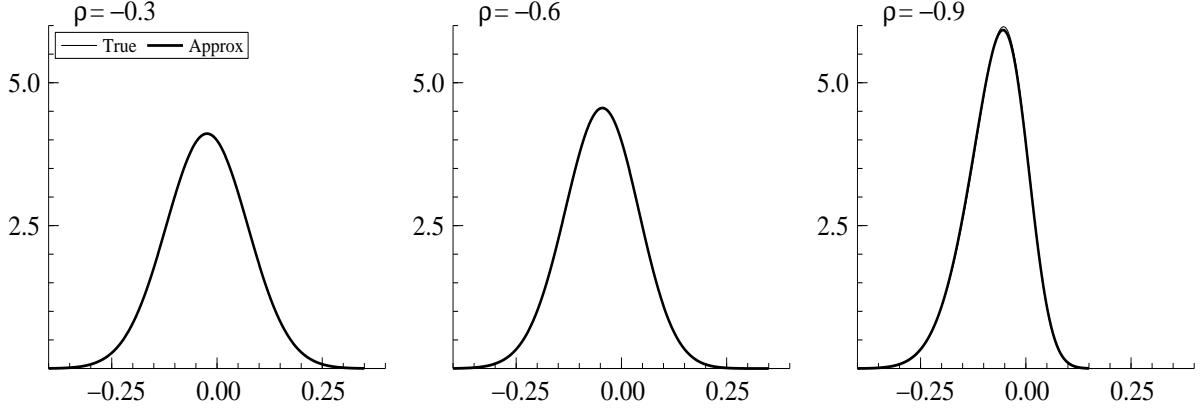$$

Figure 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.*

If we let $s = (s_1, ..., s_n)$, $\theta = (\phi, \rho, \sigma)$ and assume that $\mu \sim N(\mu_0, \sigma_0^2)$ and $h_1|\mu \sim N(\mu_0, \sigma^2/(1 - \phi^2))$, then under the auxiliary notation

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

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) and Durbin and Koopman (2001))

$$\begin{pmatrix} y_t^* \\ h_{t+1} \\ \widetilde{\mu}_{t+1} \end{pmatrix} = \begin{pmatrix} h_t \\ \widetilde{\mu}_t + \phi(h_t - \widetilde{\mu}_t) \\ \widetilde{\mu}_t \end{pmatrix} + \begin{pmatrix} \xi_t \\ \eta_t \\ 0 \end{pmatrix}, \tag{8}$$

where

$$\begin{pmatrix} h_1 \\ \widetilde{\mu}_1 \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_0 \\ \mu_0 \end{pmatrix}, \begin{pmatrix} \sigma^2/(1 - \phi^2) + \sigma_0^2 & \sigma_0^2 \\ \sigma_0^2 & \sigma_0^2 \end{pmatrix} \right). \tag{9}$$

Under a given prior $\pi(\theta)$ on θ, it is now possible to efficiently sample the posterior density

$$g(s, h, \theta, \mu|y^*, d), \quad \text{where } h = (h_1, ..., h_n), \tag{10}$$

by MCMC techniques (see for example Chib (2001) for a review of these methods). Of course, this posterior is not exactly the correct one, but we will see in subsection 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 efficient as we will show.

1. Initialize s, h, μ and θ.

2. Sample $s|h, \mu, \theta, y^*, d$.

3. Sample $(h, \mu, \theta)|s, y^*, d$ by

    (a) Sampling $\theta|s, y^*, d$.

    (b) Sampling $\mu, h|\theta, s, y^*, d$.

4. Go to 2.

2.3.2 Step 2

We first define

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

then evaluate for each $i = 1, 2, ..., K$

$$
\begin{aligned}
\pi(s_t &= i|h, \mu, \theta, y^*, d) \\
&\propto \pi(s_t = i|\xi_t, \eta_t, d_t, \mu, \theta) \\
&\propto \Pr(s_t = i)v_i^{-1} \exp\left\{ -\frac{(\xi_t - m_i)^2}{2v_i^2} - \frac{[\eta_t - d_t\rho\sigma\exp(m_i/2)\{a_i + b_i(\xi_t - m_i)\}]^2}{\sigma^2(1-\rho^2)} \right\}.
\end{aligned}
$$

This discrete distribution is sampled by the inverse distribution method.

2.3.3 Step 3

In Step 3a we sample the density

$$\pi(\theta|s, y^*, d) \propto g(y^*|d, s, \theta)\pi(\theta),$$

marginalized over μ. The density $g(y^*|d, s, \theta)$ is found from the output of the Kalman filter recursions applied to the model in (8) and (9). 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 a Laplace approximation of $\pi(\theta|s, y^*, d)$ (e.g. Chib and Greenberg (1995)). We define $\widehat{\theta} = (\hat{\phi}, \widehat{\sigma}^2, \hat{\rho})'$ which maximizes (or approximately maximizes) $g(y^*|d, s, \theta)\pi(\theta)$. Then we generate a candidate γ^* from the truncated normal distribution, $\text{TN}_R(\widehat{\theta}, \Sigma_*)$, where

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

and $R = \{\gamma : |\phi| < 1, \sigma^2 > 0, |\rho| < 1\}$. Alternatively, we may generate a candidate 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 be negative definite (e.g. when $|\hat{\rho}| \approx 1$), we take a flat proposal $\mu_* = \widehat{\theta}$ and $\Sigma_* = c_0 I$ using some constant c_0.

    Step 3b, the sampling of $g(h, \mu|d, s, \theta)$, 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) and Durbin and Koopman (2002)). Software for carrying out Gaussian simulation smoothing is widely available (Koopman, Shephard, and Doornik (1999)).

2.4 Correcting for misspecification

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

$$h^j, \mu^j, \theta^j, \quad j = 1, 2, ..., M,$$

are from the approximate posterior density $g(h, \mu, \theta | y^*, d)$. To produce draws from the correct posterior density $f(h, \mu, \theta | y^*, d)$ we simply re-weight the sampled draws. Define

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

Then we compute the weights

$$w_j^* = \prod_{t=1}^{n} \frac{f(\xi_t^j, \eta_t^j | d_t, \theta^j)}{g(\xi_t^j, \eta_t^j | d_t, \theta^j)}, \quad j = 1, 2, ..., M,$$

and let

$$w_j = \frac{w_j^*}{\sum_{i=1}^{M} w_i^*}.$$

We can now produce a sample from $f(h, \mu, \theta | y^*, d)$ by resampling the sampled variates with weights proportional to w_j. Furthermore, posterior moments can be computed by weighted averaging of the MCMC draws. 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, and so the effect of reweighting is modest.

2.5 Associated particle filter

In order to complete our inferential approach for this model we discuss a simulation-based approach to filtering. In particular, we show how we can recursively sample the distributions $h_t | y_1, ..., y_t$, $h_{t+1} | y_1, ..., y_t$ and $y_{t+1} | y_1, ..., y_t$, all conditional on μ, σ, ρ, ϕ. These sampled variates allow us to calculate marginal likelihoods, Bayes factors and goodness of fit statistics. The filtering and the associated computations are carried out by particle filter methods (e.g., in this context, Kim, Shephard, and Chib (1998) and Pitt and Shephard (1999) or more generally Doucet, de Freitas, and Gordon (2001)).

To implement particle filtering it is helpful to define the state as $\alpha_t = (h_{t+1}, h_t)$ in which case the SV model with leverage can be expressed in the form of a non-linear, non-Gaussian state space model with measurement density

$$f(y_t | \alpha_t) = f_N(y_t | \rho \eta_t \sigma^{-1} \exp(h_t/2), (1 - \rho^2) \exp(h_t)), \tag{11}$$

where

$$\eta_t = (h_{t+1} - \mu) - \phi(h_t - \mu),$$

and evolution equations

$$\begin{pmatrix} h_{t+2} \\ h_{t+1} \end{pmatrix} = \begin{pmatrix} \mu(1-\phi) \\ 0 \end{pmatrix} + \begin{pmatrix} \phi & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} h_{t+1} \\ h_t \end{pmatrix} + \begin{pmatrix} \eta_{t+1} \\ 0 \end{pmatrix}. \tag{12}$$

Our particle filtering method, which we now describe, is based on draws from (12) followed by evaluations of the density in (11).

1. Initialize $t = 0$, α_0^i from its unconditional distribution for $i = 1, 2, ..., I$.

2. For each i simulate

$$\alpha_{t+1}^{i,j} | \alpha_t^i, \quad j = 1, 2, ..., J, \tag{13}$$

and compute $w_{i,j} = f(y_{t+1} | \alpha_{t+1}^{i,j})$ and $W_{i,j} = F(y_{t+1} | \alpha_{t+1}^{i,j})$. Record

$$\overline{w}_{t+1} = \frac{1}{IJ} \sum_{j=1}^{J} \sum_{i=1}^{I} w_{i,j}, \quad \overline{W}_{t+1} = \frac{1}{IJ} \sum_{j=1}^{J} \sum_{i=1}^{I} W_{i,j}.$$

3. Resample $\alpha_{t+1}^{i,j}$ with probabilities proportional to $w_{i,j}$ to produce a sample of size I, which we label as $\alpha_{t+1}^1, \alpha_{t+1}^2, ..., \alpha_{t+1}^I$.

4. Increment t and go to 2.

It can be shown that as $I, J \to \infty$, $\overline{w}_{t+1} \xrightarrow{p} f(y_{t+1} | y_1, ..., y_t)$, and $\overline{W}_{t+1} \xrightarrow{p} F(y_{t+1} | y_1, ..., y_t)$, the predictive distribution function. In addition, the draws on α_{t+1} are particles from $\alpha_{t+1} | y_1, ..., y_t$, while the resampled items at stage 3 are samples from $\alpha_{t+1} | y_1, ..., y_{t+1}$. It therefore follows that

$$\sum_{t=1}^{n} \log \overline{w}_t \xrightarrow{p} \sum_{t=1}^{n} \log f(y_t | y_1, ..., y_{t-1}, \mu, \sigma, \rho, \phi),$$

is a consistent estimate of the conditional log-likelihood and can be used as an input in the calculation of the marginal likelihood by the method of Chib (1995). Likewise the sequence of $\overline{W}_t$, and its reflected version $2 |\overline{W}_t - 1/2|$, 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 Kim, Shephard, and Chib (1998), while earlier work along these lines in statistics include Shephard (1994), Smith (1985) and Rosenblatt (1952). Diagnostic checking of this type has been further popularized by Diebold, Gunther, and Tay (1998).

3   Illustrative example

This section gives illustrative examples to show the performance of the approximation discussed above. In the examples, we use $y_t^* = \log(y_t^2 + c)$ where the offset c is introduced to deal with very small values of y_t^2 as in Kim, Shephard, and Chib (1998). Because our ten component mixture

approximation provides an improved fit to the left tail of the $\log \chi_1^2$ 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 stochastic volatility model (1) by letting $\phi = 0.97$, $\beta \equiv \exp(\mu/2) = 0.65$, $\sigma = 0.15$ and $\rho = -0.3$. These values are based on the estimates reported by KSC and Yu (2004) 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 $n = 1,000$ observations.

The results are based on the prior distribution

$$
\begin{aligned}
\mu &\sim N(0,1), \quad \frac{\phi+1}{2} \sim \text{Beta}(20, 1.5), \\
\sigma^{-2} &\sim Gamma\left(\frac{5}{2}, \frac{0.05}{2}\right), \quad \rho \sim U(-1,1),
\end{aligned}
$$

where $U(-1,1)$ denotes a uniform distribution on $(-1,1)$. In the MCMC sampling of the posterior distribution, the initial 500 variates are discarded and the subsequent $M = 5,000$ values are retained for purposes of analysis. Figure 3 shows the sample paths, the sample autocorrelations function and the posterior densities of parameters for the case $\rho = -0.3$. The sample paths look stable and the sample 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 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. In KSC, where the $\rho = 0$ case was considered, the inefficiency factors were in the range $30 \sim 150$ (Table 5, KSC) for the original mixture sampler and $10 \sim 16$ 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.

In order to judge the quality of our approximation we next report the distribution of the weights as discussed above. Figures 3 and 7 shows the distribution of $\log(w^{(i)} \times M)$, which would all have been zero if the approximation were exact. Figure 3 looks at the case of $\rho = 0$ and compares the $K = 7$ component analysis advocated by Kim, Shephard, and Chib (1998) 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 $K = 10$. Kim, Shephard, and Chib (1998) demonstrated that reweighting had little impact on posterior inference about θ, μ, so we would expect that the improvement here is gratifying but small from a practical viewpoint.

In Figure 7, the distributions of $\log(w^{(i)} \times M)$ 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 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 $\rho = -0.3 \sim -0.5$.

12
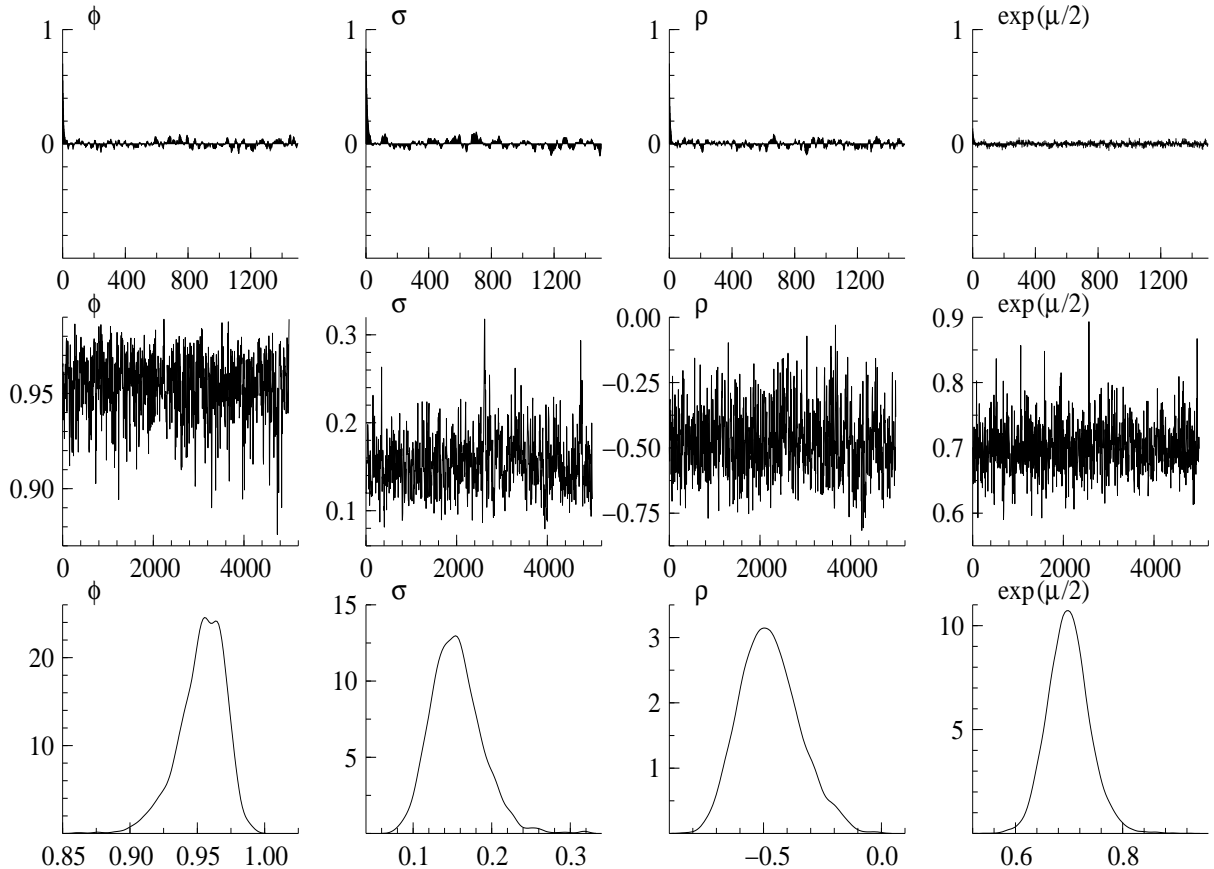
Figure 5: Asymmetric stochastic volatility model ($\rho = -0.3$). Sample autocorrelation functions, sample paths and estimated posterior densities.

4 Real data example

In this section, we apply our approximate bivariate mixture model to the stock returns data. The data are daily returns of TOPIX (Tokyo Stock Price Index), and 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 $1,232$ days on which the market was open. Table 3 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 $y_t \leq 0$.

In our analysis of these data, we use the same prior distribution given above. Again as above, in the MCMC design, the initial 500 iterations are discarded and the following $5,000$ values are recorded. Figure 8 shows the sample paths, the posterior densities and the sample autocorrelation functions of $(\phi, \sigma, \rho, \beta = \exp(\mu/2))$. The sample autocorrelations decay quickly and the output mixes well.

Table 4 shows the estimated posterior means, standard deviations, the 95% credible intervals and inefficiency measures. Inefficiency factors are small, suggesting that $1,000$ variates would
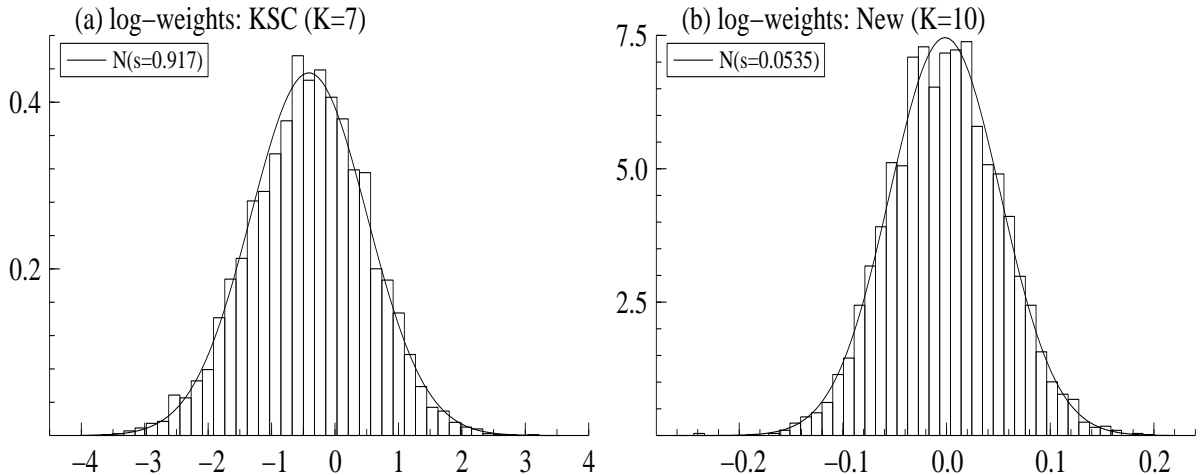
Figure 6: Histogram of the $\log(w^{(i)} \times M)$ where $M = 5,000$ is a number of samples for a symmetric stochastic volatility model ($\rho = 0$). Left: KSC ($K = 7$). Right: New ($K = 10$). Other line: the normal density function setting its mean and variance equal to the sample mean and sample variance.

be enough to calculate posterior moments of the parameters. 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.

The posterior mean of ρ is -0.36 and negative as expected. Since its 95% credible interval is $[-0.59, -0.11]$, the correlation coefficient ρ is significantly below zero. The negative value of ρ indicates that the leverage effect is present in these data.

Figure 9 shows the distributions of $\log(w^{(i)} \times M)$ 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, Kim, Shephard, and Chib (1998) 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 4 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 5 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). The log-likelihood ordinate, which is an input into this computation, was calculated from a run of the particle filter run with $I = 2,500$ and $J = 10$. 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.
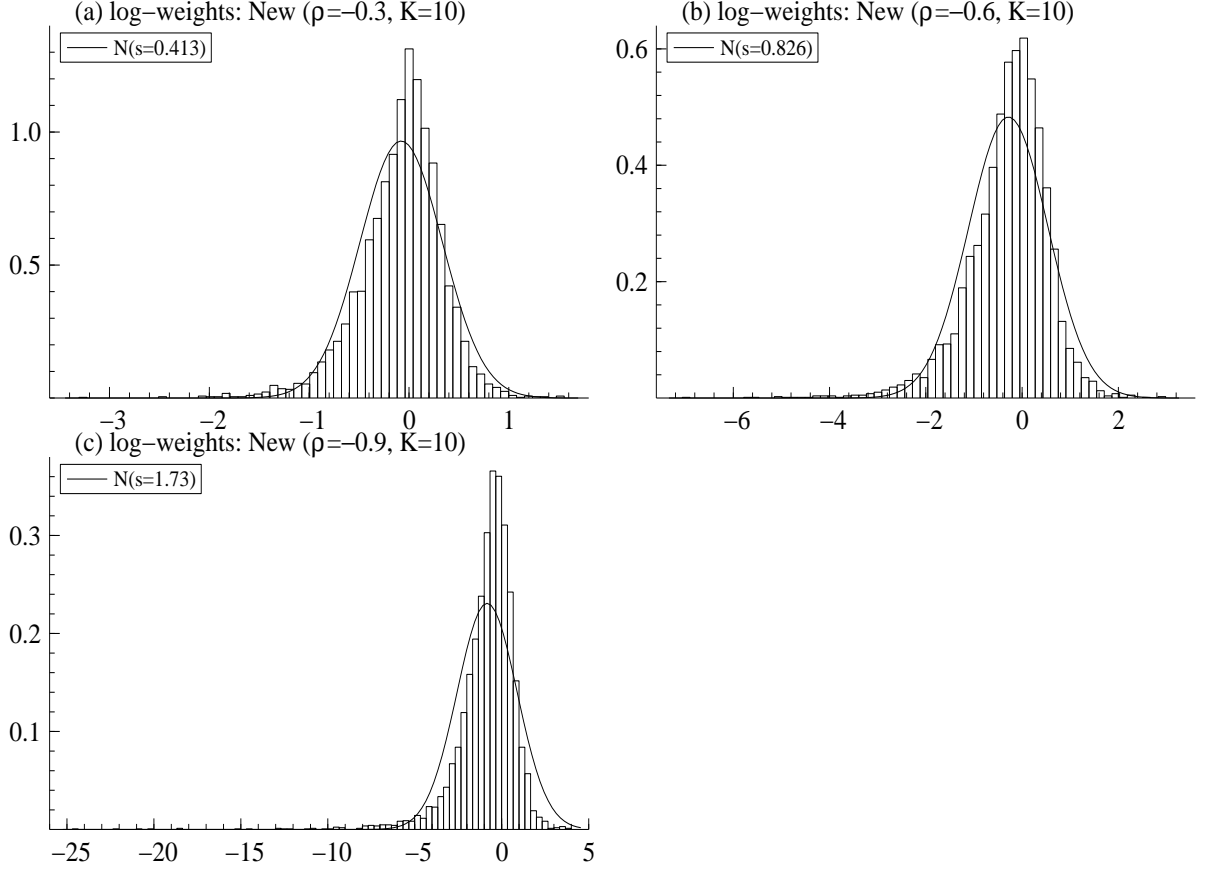
14

Figure 7: Histogram of the $\log(w^{(i)} \times M)$ where $M = 5,000$ 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.

5 More general dynamics

5.1 Framework

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

$$y_t = \epsilon_t \exp(h_t/2), \quad h_t = z_t'\alpha_t, \tag{14}$$

$$\alpha_{t+1} = b_t + T_t\alpha_t + \eta_t, \tag{15}$$

where z_t, b_t and T_t are non-stochastic processes, potentially dependent on some parameter θ. We assume that

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N(0, \Sigma), \quad \Sigma = \begin{pmatrix} 1 & \sigma' \\ \sigma & \Omega \end{pmatrix}. \tag{16}$$

In order to simplify the exposition assume that Ω is non-singular. In principle this framework can allow general forms of leverage wherein the dependence between ϵ_t and the elements of η_t is allowed to vary over the individual elements.

Parameter	True value	Unweighted Mean	Stdev.	Weighted Mean	Stdev.	95% interval	Inefficiency
ϕ	0.97	0.9533	0.0173	0.9532	0.0172	[0.9123, 0.9795]	8.5
σ	0.15	0.1554	0.0325	0.1556	0.0326	[0.1009, 0.2257]	14.7
ρ	**-0.3**	-0.4728	0.1253	-0.4673	0.1253	[-0.6915, -0.1984]	7.9
β	0.65	0.6993	0.0393	0.6988	0.0392	[0.6267, 0.7812]	2.2
ϕ	0.97	0.9563	0.0145	0.9562	0.0147	[0.9237, 0.9781]	12.6
σ	0.15	0.1667	0.0305	0.1669	0.0310	[0.1152, 0.2337]	16.1
ρ	**-0.6**	-0.7069	0.0951	-0.7027	0.0962	[-0.8623, -0.4922]	11.0
β	0.65	0.6862	0.0355	0.6865	0.0355	[0.6195, 0.7593]	2.5
ϕ	0.97	0.9700	0.0074	0.9703	0.0073	[0.9537, 0.9827]	7.5
σ	0.15	0.1844	0.0268	0.1828	0.0268	[0.1360, 0.2399]	9.5
ρ	**-0.9**	-0.8296	0.0607	-0.8290	0.0610	[-0.9301, -0.6859]	11.0
β	0.65	0.6370	0.0446	0.6367	0.0442	[0.5584, 0.7345]	3.8

Table 2: Summary statistics for three simulation experiments using a variety of values of ρ. Sample size is 1,000 throughout.

TOPIX (1998/1/5 - 2002/12/30)						
Obs.	Mean	Stdev	Max	Min	pos(+)	neg(-)
1,232	-0.0255	1.2839	5.3749	-5.6819	602	630

Table 3: Summary statistics for TOPIX return data (log-difference).

This structure implies that

$$\eta_t | d_t, \xi_t \sim \mathrm{N}\left(d_t \sigma \exp\left(\xi_t/2\right), \Omega - \sigma\sigma'\right).$$

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

$$\left\{ \begin{pmatrix} \xi_t \\ \eta_t \end{pmatrix} | d_t, s_t = i \right\} \overset{L}{=} \begin{pmatrix} m_i + v_i z_t \\ d_t \sigma(a_i + b_i v_i z_t)\exp(m_i/2) + (\Omega - \sigma\sigma')^{1/2} z_t^* \end{pmatrix},$$

where $(z_t, z_t^*)' \overset{i.i.d.}{\sim} N(0, I)$. Therefore, except for an increase in the dimension of the problem, this extension raises no new issues for our MCMC implementation.

Parameter	Unweighted Mean	Stdev.	Weighted Mean	Stdev.	95% interval	Ineff	Posterior correlation			
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

Table 4: Estimation result for TOPIX. Sample size was 5,000, based on 5,500 MCMC draws, discarding the first 500. Posterior correlation denotes the posterior correlation matrix.
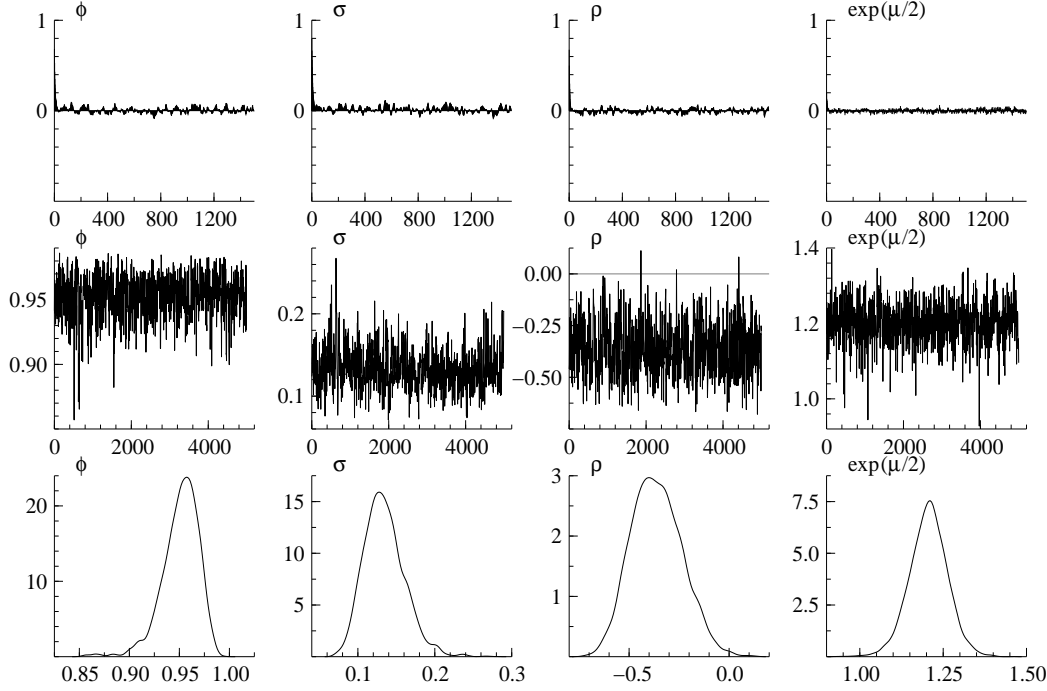
Figure 8: Estimation result for TOPIX. Sample autocorrelation functions, sample paths of MCMC output and estimated posterior densities.

5.2 Example: superposition model

Suppose that

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

where

$$\alpha_{t+1} = \begin{pmatrix} \mu \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} \phi_1 & 0 & 0 & \cdots & 0 \\ 0 & \phi_2 & 0 & \cdots & 0 \\ 0 & 0 & \phi_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \phi_I \end{pmatrix} \left(\alpha_t - \begin{pmatrix} \mu \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \right) + \eta_t,$$

and

$$\begin{pmatrix} \epsilon_t \\ \eta_t \end{pmatrix} \sim N\left(0, \Sigma\right), \quad \Sigma = \begin{pmatrix} 1 & \rho_1\sigma_1 & \rho_2\sigma_2 & \cdots & \rho_I\sigma_I \\ \rho_1\sigma_1 & \sigma_1^2 & 0 & \cdots & 0 \\ \rho_2\sigma_2 & 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_I\sigma_I & 0 & & & \sigma_I^2 \end{pmatrix}.$$

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
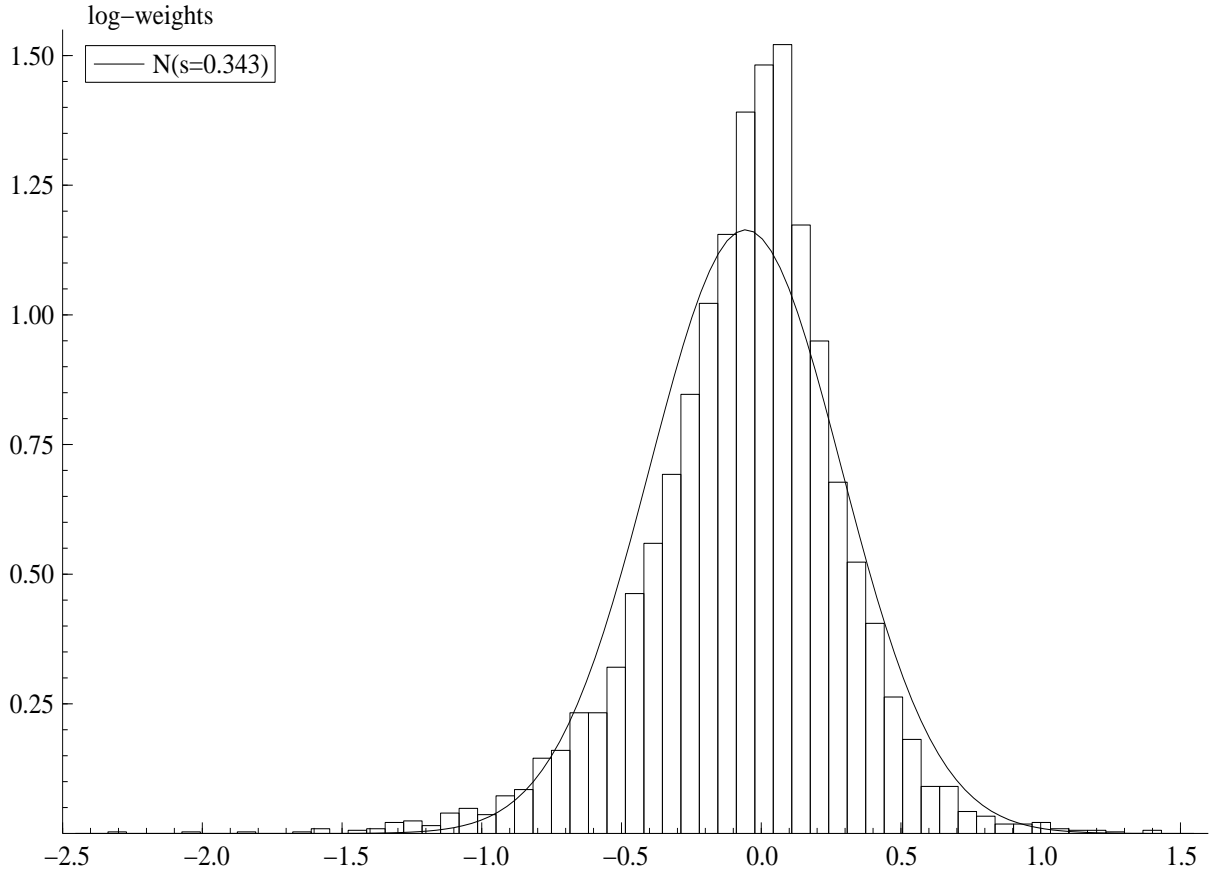
17

Figure 9: Sampling result of log-weights $\log(w^{(i)} \times M)$ for the TOPIX series. Shows histogram and fitted normal density.

volatility models while close to corresponding continuous time models (Shephard (1996), Engle and Lee (1999), Barndorff-Nielsen and Shephard (2001) and Chernov, Gallant, Ghysels, and Tauchen (2003)). It is easy to check that for Σ to be positive definite we need $\sum_{i=1}^{I} \rho_i^2 < 1$.

Column 6 of Table 5 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. These results are based on a prior where $(\phi_2 + 1)/2 \sim Beta(10, 10)$ with the side constraint that $\phi_2 < \phi_1$. Further, we assume $(\rho_2 + 1)/2 \sim Beta(10, 10)$ with the constraint that $0 < \rho_1^2 + \rho_2^2 < 1$. Finally, $\sigma_2^{-2} \sim Gamma(5/2, 0.05/2)$. To generate a candidate with such constraints, we may consider 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 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.

	SV	SV-t	ASV	ASV-t	ASV-g	SP
Likelihood ordinate	-2033.98	-2033.24	-2029.85	-2029.30	-2028.87	-2029.75
(S.E.)	(0.38)	(0.33)	(0.54)	(0.48)	(0.57)	(0.25)
Prior ordinate	3.75	0.16	3.09	3.38	3.38	7.84
Posterior ordinate	8.87	5.17	10.25	12.50	12.50	14.81
(S.E.)	(0.02)	(0.20)	(0.02)	(0.04)	(0.04)	(0.04)
Marg Likelihood	-2039.10	-2038.24	-2037.00	-2038.41	-2037.98	-2036.72
(S.E.)	(0.38)	(0.39)	(0.54)	(0.49)	(0.57)	(0.25)

Table 5: Marginal likelihood estimation by the Chib (1995) method for the TOPIX data. 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.

5.3 Example: heavy-tailed error distribution

Many writers have followed Harvey, Ruiz, and Shephard (1994) in extending the SV model to allow for heavier tailed returns. For example, Chib, Nardari, and Shephard (2002) extended the basic Kim, Shephard, and Chib (1998) approach by letting

$$
\begin{aligned}
y_t &= \sqrt{\lambda_t}\epsilon_t \exp(h_t/2), &(17)\\
h_{t+1} &= \mu + \phi(h_t - \mu) + \eta_t, \quad t = 0, 1, \ldots, n, &(18)
\end{aligned}
$$

where λ_t is an i.i.d. scale mixture variable and $\lambda_t \perp\!\!\!\perp (\epsilon_t, \eta_t)$. This is relevant empirically and also maps into the literature on time-change Lévy processes and Lévy based SV models (Carr, Geman, Madan, and Yor (2003), Carr and Wu (2004) and Cont and Tankov (2004)). Papers on various inferential aspects of these models include Barndorff-Nielsen and Shephard (2003) and Li, Wells, and Yu (2004). In this subsection we will assume that

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

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

The above model fits into the framework put forward in (14)-(16) by writing

$$
\begin{aligned}
y_t &= \epsilon_t \exp(h_t/2), &(19)\\
h_t &= h_{t+1}^* + \lambda_t, &(20)\\
h_{t+1}^* &= \mu + \phi(h_t^* - \mu) + \eta_t, \quad t = 0, 1, \ldots, n, &(21)
\end{aligned}
$$

where

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

Therefore, this extension again raises no new inferential issues.

19

Table 5 gives results for the three different heavier tailed specifications. In the second column we report the results when $\sqrt{\lambda_t}\epsilon_t$ follows a student-t distribution with ν degrees of freedom, where $\nu \sim Gamma(16, 0.8)$. The fit of this model is better than the basic model, but not over the leverage model. The fifth column reports the results for the Gaussian SV model which is marginally better than the student-t model. Overall, however, the simple Gaussian SV model with leverage is preferred for these data.

6   Conclusion

In this paper we have extended the Kim, Shephard, and Chib (1998) approach to SV models with leverage. This approach starts with the joint distribution of $\log \epsilon_t^2, \eta_t | sign(y_t)$ which is then approximated by a suitably constructed ten-component mixture of bivariate normal distributions. We show that this approach, 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 approach 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 Kim, Shephard, and Chib (1998) 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.

7   Acknowledgment

References

Barndorff-Nielsen, O. E. and N. Shephard (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. and N. Shephard (2003). Impact of jumps on returns and realised variances: econometric analysis of time-deformed Lévy processes. Unpublished paper: Nuffield College, Oxford.

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

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

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

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

Chernov, M., A. R. Gallant, E. Ghysels, and G. Tauchen (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*, 1313–21.

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

Chib, S. and E. Greenberg (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistican 49*, 327–35.

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

Cont, R. and P. Tankov (2004). *Financial Modelling with Jump Processes.* London: Chapman and Hall.

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

Diebold, F. X., T. A. Gunther, and T. S. Tay (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.* London: Timberlake Consultants Press.

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

Durbin, J. and S. J. Koopman (2001). *Time Series Analysis by State Space Methods.* Oxford: Oxford University Press.

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

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

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

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

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

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

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

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

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

Li, H., M. Wells, and L. Yu (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. and P. Schotman (1998). An empirical application of stochastic volatility models. *Journal of Applied Econometrics 16*, 333–59.

Nelson, D. B. (1988). The time series behaviour of stock market volatility and returns. Unpublished Ph.D.: MIT.

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

Pitt, M. K. and N. Shephard (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*. Forthcoming.

Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics 23*, 470–2.

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

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

Shephard, N. (2004). *Stochastic Volatility: Selected Readings.* Oxford: Oxford University Press. Forthcoming.

Smith, J. Q. (1985). Diagnostic checks of non-standard time series models. *Journal of Forecasting 4*, 283–91.

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

West, M. and J. Harrison (1997). *Bayesian Forecasting and Dynamic Models* (2 ed.). New York: Springer-Verlag.

Yu, J. (2004). On leverage in a stochastic volatility model. *Journal of Econometrics*. Forthcoming.