Available online at www.sciencedirect.com
ScienceDirect
Journal of Empirical Finance 15 (2008) 549–566
Journal of EMPIRICAL FINANCE
www.elsevier.com/locate/jempfin

Box-Cox stochastic volatility models with heavy-tails and correlated errors

Xibin Zhang, Maxwell L. King *

Department of Econometrics and Business Statistics, Monash University, Clayton, Victoria, 3800, Australia

Accepted 8 May 2007
Available online 25 October 2007

Abstract

This paper presents a Markov chain Monte Carlo (MCMC) algorithm to estimate parameters and latent stochastic processes in the asymmetric stochastic volatility (SV) model, in which the Box-Cox transformation of the squared volatility follows an autoregressive Gaussian distribution and the marginal density of asset returns has heavy-tails. We employed the Bayes factor and the Bayesian information criterion (BIC) to examine whether the Box-Cox transformation of squared volatility is favored against the log-transformation. When applying the heavy-tailed asymmetric Box-Cox transformed SV model, three competing SV models and the *t*-GARCH(1,1) model to continuously compounded daily returns of the Australian stock index, we find that the Box-Cox transformation of squared volatility is strongly favored by Bayes factors and BIC against the log-transformation. While both criteria strongly favor the *t*-GARCH(1,1) model against the heavy-tailed asymmetric Box-Cox transformed SV model and the other three competing SV models, we find that SV models fit the data better than the *t*-GARCH(1,1) model based on a measure of closeness between the distribution of the fitted residuals and the distribution of the model disturbance. When our model and its competing models are applied to daily returns of another five stock indices, we find that in terms of SV models, the Box-Cox transformation of squared volatility is strongly favored against the log-transformation for the five data sets.

JEL classification: C12; C15; C52
Keywords: Bayes factors; GARCH; Leverage effect; Likelihood; Markov chain Monte Carlo; Particle filter

1. Introduction

The volatility of asset returns often exhibits a time-varying feature. One approach to modelling volatility is to employ the autoregressive conditional heteroskedasticity (ARCH) model developed by Engle (1982) or the generalized ARCH (GARCH) model by Bollerslev (1986). An alternative is the stochastic volatility (SV) model, in which the volatility is assumed to be a latent stochastic process. The SV model has received increased attention in the finance

* Corresponding author. Tel.: +61 3 99052449; fax: +61 3 99058039.
E-mail addresses: xibin.zhang@buseco.monash.edu.au (X. Zhang), max.king@adm.monash.edu.au (M.L. King).

0927-5398/$ - see front matter
doi:10.1016/j.jempfin.2007.05.002

literature, because it provides an alternative approach to the Black–Scholes option pricing formula (Hull and White, 1987). Taylor (1982, 1986) shows that the SV model is often formulated in terms of stochastic differential equations,

$$\begin{aligned} \mathrm{d}(\ln p_t) &= \alpha \mathrm{d}t + \sigma_t \mathrm{d}w_{1t}, \\ \mathrm{d}\left(\ln\sigma_t^2\right) &= \lambda\left(\xi - \ln\sigma_t^2\right)\mathrm{d}t + \sigma_w \mathrm{d}w_{2t}, \end{aligned} \tag{1}$$

where p_t is the price of an asset at time t and $(w_{1t}, w_{2t})'$ is a bivariate standard Brownian motion. The correlation between $\mathrm{d}w_{1t}$ and $\mathrm{d}w_{2t}$, denoted by ρ=corr($\mathrm{d}w_{1t}$, $\mathrm{d}w_{2t}$), captures the leverage effect, which refers to the asymmetric behaviour that price movements are negatively correlated with volatility and is often observed in returns of equity prices (see, e.g., Nelson, 1990; Gallant et al., 1992, 1993; Campbell and Kyle, 1993; Engle and Ng, 1993). The empirical version of this model is typically formulated in discrete time as

$$\begin{aligned} y_t &= \sigma_t\, \varepsilon_t, \\ \ln\sigma_{t+1}^2 &= \mu + \phi\left(\ln\sigma_t^2 - \mu\right) + \sigma_u u_{t+1}, \end{aligned} \tag{2}$$

where y_t is the continuously compounded return, $\varepsilon_t \sim N(0,1)$, $u_t \sim N(0,1)$, $\ln\sigma_1^2 \sim N(0, \sigma_u^2/(1-\phi^2))$, and the correlation between ε_t and u_{t+1}, denoted by ρ=corr(ε_t, u_{t+1}), captures the leverage effect.[1] To reflect the asymmetric correlation between errors in the mean and volatility equations, this model is often called the asymmetric log-normal SV (LSV) model, which was set up based on models of Clark (1973) and Tauchen and Pitts (1983) and was first documented by Taylor (1982).

The discrete-time LSV model specifies the logarithmic squared volatility as an autoregressive Gaussian process, while there are some other specifications of the volatility process (see, e.g., Hull and White, 1987; Stein and Stein, 1991; Heston, 1993; Andersen, 1994; Jacquier et al., 1994; Eraker et al., 2003). Yu et al. (2006) present an extension to the log-normal specification of squared volatilities, in which the Box-Cox transformation of the squared volatility is assumed to follow an autoregressive Gaussian distribution. Ignoring the leverage effect, they developed an MCMC algorithm to sample parameters and volatilities. They applied their model to daily returns of the dollar/pound exchange rate and found that the 90% Bayesian credible interval for the Box-Cox transformation parameter does not cover zero. This is significant because a value of zero is equivalent to the logarithmic transformation of squared volatilities. Applying the asymmetric Box-Cox transformed SV model with heavy tails to daily returns of six stock indices, we find strong evidence supporting the Box-Cox transformation of squared volatility.

The introduction of the Box-Cox transformation of squared volatilities in a SV model was motivated by a similar transformation proposed by Higgins and Bera (1992), who applied the Box-Cox power transformation to conditional variance in an ARCH model. The purpose of such a transformation is to allow for skewness in the marginal distribution of the squared volatility, because it is inappropriate to assume this distribution to be symmetric. The log-transformation of squared volatility has a similar purpose as the Box-Cox transformation. However, as we discuss in Section 4.4, the latter can make the tails of the density of squared volatility thinner than the former, especially for the left tail. The Box-Cox transformation has an additional advantage that it allows the volatility process itself to choose a transformation parameter. It seems that the Box-Cox transformation is a useful extension to the log-transformation for squared volatilities in SV models.

The sampling algorithm developed by Yu et al. (2006) is limited in two aspects. First, it does not incorporate the leverage effect, which is often observed in the distribution of equity returns (see, e.g., Eraker et al., 2003; Jacquier et al., 2004). Jacquier et al. (2004) point out that the leverage effect often induces skewness in the marginal distribution of returns on asset prices. This finding is consistent with the non-parametric evidence found by Gallant et al. (1997). Second, the sampling algorithm of Yu et al. (2006) does not take account of the heavy-tailed feature in the marginal distribution of asset returns. Eraker et al. (2003) argue that asset returns often experience significant shocks, which result in a heavy-tailed marginal distribution. This paper provides remedies to these two problems.

In this paper, we incorporate the leverage effect into the Box-Cox transformed SV model and develop a fully specified posterior density of parameters and transformed volatilities. To incorporate the heavy-tailed feature into the marginal distribution of errors in the mean equation, we assume that the errors follow a mixture of distributions using a particular hierarchical prior. Such a mixture turns into a Student t distribution and has been widely used to capture the heavy-tailed feature in the marginal distribution of asset returns. A similar treatment of such a heavy-tailed distribution can be found in Geweke (1993) and Jacquier et al. (2004). A sampling algorithm is presented to estimate this latent

[1] Taylor (1994) shows that the correlation between ε_t and u_{t+1} captures the leverage effect, and empirical evidence can be found in Ghysels et al. (1996), Harvey and Shephard (1996) and Bollerslev and Zhou (2005).

process. In addition to the estimation procedure, we employ Bayes factors, BIC and a closeness measure to examine whether the heavy-tailed asymmetric Box-Cox transformed SV model is favored against several competing SV models and the GARCH(1,1) model with Student t errors.

This paper aims to develop an MCMC algorithm, in which both the leverage effect and the heavy-tailed marginal distribution can be incorporated. The rest of the paper is organized as follows. Section 2 provides a brief description of the asymmetric Box-Cox SV model, the fully specified posterior density, conditional densities, and sampling algorithms designed to sample parameters and volatilities. In Section 3, we discuss the computation of Bayes factors in the asymmetric Box-Cox transformed SV model with heavy tails, where a particle filter is employed to compute the likelihood numerically. Section 4 presents an application of heavy-tailed asymmetric Box-Cox transformed SV models and several competing models to daily returns of the Australian stock index. In Section 5, the heavy-tailed asymmetric Box-Cox transformed SV model and several competing models are applied to daily returns of some international stock indices. Section 6 concludes the paper.

2. MCMC in the Box-Cox transformed SV model

2.1. Basic Box-Cox transformed SV model

The SV model proposed by Yu et al. (2006) assumes that the Box-Cox transformation of squared volatility follows an autoregressive Gaussian process,

$$\begin{aligned} y_t &= \sigma_t\ \varepsilon_t, \\ h\left(\sigma_{t+1}^2, \delta\right) &= \mu + \phi\left[h\left(\sigma_t^2, \delta\right) - \mu\right] + \sigma_u u_{t+1}, \end{aligned} \tag{3}$$

where $\varepsilon_t \sim N(0,1)$, $u_t \sim N(0,1)$, and $h(\sigma_t^2, \delta)$ is defined by

$$h(x, \delta) = \begin{cases} \left(x^\delta - 1\right)/\delta & \text{if } \delta \neq 0 \\ \ln x & \text{if } \delta = 0 \end{cases}. \tag{4}$$

This model is called the basic BCSV model hereafter. Let $\alpha_t = h(\sigma_t^2, \delta)$ denote the Box-Cox transformed squared volatility. The model can be equivalently expressed as

$$\begin{aligned} y_t &= \sqrt{g(\alpha_t, \delta)}\ \varepsilon_t, \\ \alpha_t &= \mu + \phi(\alpha_{t-1} - \mu) + \sigma_u u_t, \end{aligned} \tag{5}$$

where

$$g(\alpha_t, \delta) = \begin{cases} (1 + \delta\alpha_t)^{1/\delta} & \text{if } \delta \neq 0 \\ \exp(\alpha_t) & \text{if } \delta = 0 \end{cases}, \tag{6}$$

which is denoted as g_t hereafter.

Let $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)'$, $\mathbf{y} = (y_1, y_2, \ldots, y_n)'$, and $\theta = (\phi, \delta, \mu, \sigma_u)'$. The posterior density of $(\theta', \alpha')'$ is

$$\pi(\theta, \alpha|\mathbf{y}) \propto p(\mathbf{y}|\theta, \alpha) p(\alpha|\theta) p(\theta), \tag{7}$$

where $p(\mathbf{y}|\theta, \alpha)$ is the likelihood of $\mathbf{y}$ given $(\theta', \alpha')'$, $p(\alpha|\theta)$ is the density of α, and $p(\theta)$ is the prior density of θ. The sampling algorithm developed by Yu et al. (2006) involves breaking the joint posterior $\pi(\theta, \alpha|\mathbf{y})$ into two Gibbs' blocks denoted by $p(\theta|\alpha, \mathbf{y})$ and $p(\alpha|\theta, \mathbf{y})$, respectively. The sampling algorithm consists of sampling μ and σ_u^2 from the corresponding conditional posteriors, sampling ϕ and δ simultaneously through the random-walk Metropolis–Hastings algorithm, and sampling α via the single-move random-walk Metropolis–Hastings algorithm.

2.2. Asymmetric BCSV model

The basic BCSV model can be extended by allowing a nonzero correlation between ε_t and u_{t+1} that captures the leverage effect, and the model can be expressed as

$$\begin{aligned} y_t &= g_t^{1/2} \varepsilon_t, \\ \alpha_{t+1} &= \mu + \phi(\alpha_t - \mu) + \sigma_u u_{t+1}, \end{aligned} \tag{8}$$

where $\rho = \text{corr}(\varepsilon_t, u_{t+1})$, and g_t is defined in Eq. (6). Let $\theta = (\phi, \delta, \mu, \rho, \sigma_u)'$ represent the augmented parameter vector. To obtain the likelihood of $\mathbf{y}$ given $(\theta', \alpha')'$, we define

$$u_{t+1} = \rho\varepsilon_t + \sqrt{1-\rho^2}\eta_{t+1}, \tag{9}$$

for $t = 1, 2, \ldots, n-1$, where η_{t+1} is assumed to follow $N(0,1)$ and to be uncorrelated with ε_t. Eq. (9) shows that $\text{var}(u_{t+1}) = 1$ and $\text{cov}(u_{t+1}, \varepsilon_t) = \rho$. Substituting Eq. (9) into Eq. (8), we obtain

$$\alpha_{t+1} = \mu + \phi(\alpha_t - \mu) + \rho\sigma_u g_t^{-1/2} y_t + \sqrt{1-\rho^2}\sigma_u \eta_{t+1}.$$

When incorporating the leverage effect into the log-normal SV model, Jacquier et al. (2004) re-parameterized ρ and σ_u to $\varphi = \rho\sigma_u$ and $\tau^2 = (1-\rho^2)\sigma_u^2$. We follow such a re-parameterization and obtain

$$\begin{aligned} y_t &= \sqrt{g_t}\varepsilon_t, \\ \alpha_{t+1} &= \mu + \phi(\alpha_t - \mu) + \varphi g_t^{-1/2} y_t + \tau\eta_{t+1}, \end{aligned} \tag{10}$$

where $\alpha_1 \sim N(\mu, \tau^2/(1-\phi^2))$ and $y_n \sim N(0, g_n)$. Because ε_t and η_{t+1} are uncorrelated, the posterior of $(\theta', \alpha')'$ can be obtained in the same way as that in the basic BCSV model.

Model (10) is referred to as the asymmetric BCSV model hereafter.

2.3. Joint posterior of parameters and transformed volatilities

In order to obtain the posterior density of $(\theta', \alpha')'$, we need a prior for each parameter. Assume that $(\phi+1)/2 \sim \text{Beta}(\omega, \gamma)$ and $\tau^2 \sim \text{IG}(\zeta/2, S_\tau/2)$, which are, respectively, expressed explicitly as

$$\begin{aligned} p(\phi) &\propto \left(\frac{\phi+1}{2}\right)^{\omega-1}\left(1-\frac{\phi+1}{2}\right)^{\gamma-1}, \\ p(\tau^2) &\sim \left(\frac{1}{\tau^2}\right)^{\zeta/2+1} \exp\left\{-\frac{S_\tau/2}{\tau^2}\right\}, \end{aligned}$$

where ω, γ, ζ and S_τ are hyperparameters to be defined by users. The priors of the other parameters are, respectively, $\varphi|\tau^2 \sim N(\varphi_0, \tau^2/p_0)$, $\mu|\tau^2 \sim N(\mu_0, \tau^2/q_0)$, and $\delta \sim U(-2, 2)$ with φ_0, μ_0, p_0 and q_0 being hyperparameters.

The joint prior of θ is the product of these marginal priors. According to Eq. (7), the posterior density of $(\theta', \alpha')'$ is

$$\begin{aligned} \pi(\theta, \alpha|\mathrm{y}) \;&\propto\; p(\phi, \delta, \mu, \rho, \sigma) \times \prod_{t=1}^{n-1} p(y_t|\alpha_t, \theta) p(y_n|\theta) \times p(\alpha_1|\theta) \prod_{t=1}^{n-1} p(\alpha_{t+1}|\alpha_t, \theta) \\ &= p(\delta)(1+\phi)^{\omega-1/2}(1-\phi)^{\gamma-1/2}\left(\prod_{t=1}^{n} g_t^{-1/2}\right) \exp\left\{-\frac{1}{2}\sum_{t=1}^{n}\frac{y_t^2}{g_t}\right\}\left(\frac{1}{\tau^2}\right)^{(n+v+2)/2+1} \exp\left\{-\frac{\kappa}{2\tau^2}\right\}, \end{aligned} \tag{11}$$

where

$$\kappa = \left(1-\phi^2\right)(\alpha_1 - \mu)^2 + \sum_{t=1}^{n-1}\left(\alpha_{t+1} - \mu - \phi(\alpha_t - \mu) - \varphi g_t^{-1/2} y_t\right)^2 + p_0(\varphi - \varphi_0)^2 + q_0(\mu - \mu_0)^2 + S_\tau.$$

After integrating out τ^2 from the joint posterior (11), we obtain the logarithmic posterior density of $(\phi, \delta, \mu, \phi, \alpha')'$ given $\mathbf{y}$,

$$\begin{aligned} \ln \pi(\phi, \delta, \mu, \varphi, \alpha|\mathbf{y}) = \ln\; & p(\delta) + (\omega - 1/2)\ln(1+\phi) + (\gamma - 1/2)\ln(1-\phi) \\ & -\frac{1}{2}\sum_{t=1}^{n} \ln(g_t) - \frac{1}{2}\sum_{t-1}^{n}\frac{y_t^2}{g_t} - \frac{n+\zeta+2}{2}\ln(\kappa/2), \end{aligned} \tag{12}$$

while the conditional posterior of τ^2 is the inverted gamma density, IG($(n+\zeta+2)/2$, $\kappa/2$). In the appendix of Zhang and King (2004), a different routine to obtain the joint posterior of $(\theta', \alpha')'$ was presented. However, these two approaches result in the same posterior density.

2.4. Conditional posteriors

Once the posterior of $(\theta', \alpha')'$ is obtained, we can use the Gibbs sampler to sample each component of $(\theta', \alpha')'$ conditional on the other components. However, the mixing speed will generally be slow. If conditional posteriors of some parameters can be obtained, these parameters can be sampled, respectively, from their conditional posteriors directly. As a consequence, the overall mixing performance will be improved (see, e.g., Johannes and Polson (in press), for a discussion on sampling techniques based on conditional posteriors).

2.4.1. Conditional posterior of τ^2

As τ^2 can be integrated out from the joint posterior (11), the conditional posterior density of τ^2 is

$$\tau^2 \sim \mathrm{IG}\left(\frac{n+\zeta+2}{2}, \frac{\kappa}{2}\right). \tag{13}$$

Given the other parameters and α, we can sample τ^2 directly from its conditional posterior.

2.4.2. Conditional posterior of φ

For sampling ϕ based on Eq. (11), we found that the conditional posterior of ϕ is (up to a normalizing constant)

$$\begin{aligned} & p(\varphi|\tau^2, \phi, \delta, \mu, \alpha, \mathbf{y}) \\ \propto \quad & \exp\left\{-\frac{1}{2\tau^2}\left[\sum\nolimits_{t=1}^{n-1}\left((\alpha_{t+1}-\mu)-\phi(\alpha_t-\mu)-\varphi g_t^{-1/2}y_t\right)^2+p_0(\varphi-\varphi_0)^2\right]\right\} \\ = \quad & \exp\left\{-\frac{1}{2\tau^2}\left[a_{11}\varphi^2-2a_{12}\varphi+a_{22}+p_0\left(\varphi^2-2\varphi_0\varphi+\varphi_0^2\right)\right]\right\} \\ \propto \quad & \exp\left\{-\frac{1}{2\tau^2/(a_{11}+p_0)}\left(\varphi^2-2\frac{a_{12}+\varphi_0 p_0}{a_{11}+p_0}\varphi\right)\right\}, \end{aligned}$$

where

$$a_{11} = \sum_{t=1}^{n-1} y_t^2/g_t, \quad a_{12} = \sum_{t=1}^{n-1}[\alpha_{t+1}-\mu-\phi(\alpha_t-\mu)]y_t/\sqrt{g_t}.$$

Hence the conditional posterior of φ is the Gaussian density,

$$\varphi \sim N\left(\frac{a_{12}+\varphi_0 p_0}{a_{11}+p_0}, \frac{\tau^2}{a_{11}+p_0}\right), \tag{14}$$

based on which φ can be sampled directly, given the other parameters and latent volatilities. Once τ^2 and φ are sampled, respectively, from their conditional posteriors, we can calculate ρ and σ through $\sigma^2=\varphi^2+\tau^2$ and $\rho=\varphi/\sigma$.

2.4.3. Conditional posterior of μ

For sampling μ based on Eq. (11), we found that the conditional posterior of μ is (up to a normalizing constant)

$$\begin{aligned} & p(\mu|\tau^2, \phi, \delta, \varphi, \alpha, \mathbf{y}) \\ \propto \quad & \exp\left\{-\frac{1}{2\tau^2}\left[\left(1-\phi^2\right)(\alpha_1-\mu)^2+\sum\nolimits_{t=1}^{n-1}(b_{t+1}-(1-\phi)\mu)^2+q_0(\mu-\mu_0)^2\right]\right\} \\ \propto \quad & \exp\left\{-\frac{(n-1)(1-\phi)^2+\left(1-\phi^2\right)q_0}{2\tau^2}\left(\mu^2-2\frac{\left(1-\phi^2\right)\alpha_1+(1-\phi)\sum b_{t+1}+q_0\mu_0}{(n-1)(1-\phi)^2+\left(1-\phi^2\right)+q_0}\mu\right)\right\}, \end{aligned}$$

where $b_{t+1}=\alpha_{t+1}-\phi\alpha_t-\varphi g_t^{-1/2}y_t$ for $t=1, 2,\dots, n-1$. Then the conditional posterior of μ is the Gaussian density with mean and variance defined, respectively, by

$$\mu_* = \frac{(1-\phi^2)\alpha_1 + (1-\phi)\sum_{t=1}^{n-1} b_{t+1} + q_0\mu_0}{(n-1)(1-\phi)^2 + (1-\phi^2) + q_0},$$
$$\sigma_*^2 = \frac{\tau^2}{(n-1)(1-\phi)^2 + (1-\phi^2) + q_0}. \tag{15}$$

Hence μ can be sampled directly from $N(\mu_*, \sigma_*^2)$, given the other parameters and α.

2.4.4. Sampling ϕ and δ

In order to sample ϕ and δ, we use the random-walk Metropolis–Hastings algorithm, in which the proposal density is the standard normal and the acceptance probability is calculated based on the joint posterior (12).

2.4.5. Sampling α

We use the single-move random-walk Metropolis–Hastings algorithm to sample components of α sequentially, where the acceptance probability is calculated based on the joint posterior (11). Let $\alpha\backslash_k$ denote α with its kth component deleted, and $\pi(\alpha_k)$ denote the posterior of α_k conditional on $(\theta, \alpha\backslash_k, \mathbf{y})$, for $k=1, 2,\dots, n$. We found that $\pi(\alpha_k)$ has the following simple expressions.[2]

$$\pi(\alpha_1) \propto -\frac{\ln g_1}{2} - \frac{y_1^2}{2g_1} - \frac{1}{2\tau^2}\left[(1-\phi^2)(\alpha_1-\mu)^2 + \left(\alpha_2-\mu-\phi(\alpha_1-\mu)-\varphi g_1^{-1/2}y_1\right)^2\right],$$
$$\pi(\alpha_n) \propto -\frac{\ln g_n}{2} - \frac{y_n^2}{2g_n} - \frac{1}{2\tau^2}\left[\alpha_n-\mu-\phi(\alpha_{n-1}-\mu)-\varphi g_{n-1}^{-1/2}y_{n-1}\right]^2, \tag{16}$$
$$\pi(\alpha_k) \propto -\frac{\ln g_k}{2} - \frac{y_k^2}{2g_k} - \frac{1}{2\tau^2}\sum_{t=k-1}^{k}\left[\alpha_{t+1}-\mu-\phi(\alpha_t-\mu)-\varphi g_t^{-1/2}y_t\right]^2,$$

for $k=2, 3\dots, n-1$.

In summary, the sampling procedure of $(\theta', \alpha')'$ is as follows:

- sample (ϕ, δ) from their joint posterior (12) using the random-walk Metropolis-Hastings algorithm;
- sample τ^2 directly from the inverted Gamma density given by Eq. (13);
- sample φ directly from the Gaussian density given by Eq. (14);
- compute ρ and σ^2 based on τ^2 and φ;
- sample μ directly from the Gaussian density with mean and variance defined in Eq. (15);
- sample α using the single-move random-walk Metropolis–Hastings algorithm with the acceptance probability computed through Eq. (16).

2.5. Heavy-tailed departure from normality

In the context of SV models, heavy tails are often observed in the marginal distribution of errors in the mean equation (see, e.g., Gallant, Hsieh and Tauchen, 1997). Motivated by the heavy-tailed asymmetric LSV model by Jacquier et al. (2004), we set up the following model,

$$y_t = g_t^{1/2}\lambda_t^{1/2}\varepsilon_t,$$
$$\alpha_{t+1} = \mu + \phi(\alpha_t-\mu) + \sigma_u u_{t+1}, \tag{17}$$

[2] In the literature on SV models involving MCMC simulation, there are alternative algorithms for sampling α, such as the block-wise Metropolis-Hastings algorithm presented in Kim, Shephard and Chib (1998) and the accept/reject Metropolis-Hastings algorithm in Jacquier et al. (2004). However, we have been unable to modify these algorithms to incorporate the Box-Cox transformation involved in our model.

where $\varepsilon_t \sim N(0, 1)$, $u_{t+1} \sim N(0, 1)$, $\mathrm{cov}(\varepsilon_t, u_{t+1}) = \rho$, and $\lambda_t \sim \mathrm{IG}(v/2, v/2)$, which is equivalent to the assumption that v/λ_t follows a χ^2 distribution with v degrees of freedom.

This assumption implies that the marginal distribution of $\sqrt{\lambda_t \varepsilon_t}$ is Student t with v degrees of freedom. The density of λ_t is

$$p(\lambda_t|v) = \frac{(v/2)^{v/2}}{\Gamma(v/2)} \left(\frac{1}{\lambda_t}\right)^{v/2+1} \exp\left\{-\frac{v/2}{\lambda_t}\right\}.$$

The heavy-tailed asymmetric BCSV model has a parameter vector $(v, \theta')'$ and two latent stochastic processes denoted by α and $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_n)'$, respectively. Consider simulating from the joint posterior density $\pi(\theta, \alpha, v, \lambda|\mathbf{y})$ by successive sampling through conditional posteriors $p(\theta, \alpha|v, \lambda, \mathbf{y})$ and $p(\lambda, v|\theta, \alpha, \mathbf{y})$. The algorithm for drawing each set of conditionals is as follows.

2.5.1. Sampling θ and α

Given λ, the mean equation of the heavy-tailed asymmetric BCSV model is

$$y_t^* = y_t \lambda_t^{-1/2} = g_t^{1/2} \varepsilon_t.$$

The sampling algorithm for the asymmetric BCSV model discussed in Section 2.4 applies directly.

2.5.2. Sampling λ and v

Given $(v, \theta', \alpha')'$, the posterior of λ is (up to a normalizing constant)

$$\begin{aligned} p(\lambda|v, \theta, \alpha, \mathbf{y}) &= \prod_{t=1}^{n} p(\lambda_t|y_t, \alpha_t, v) \propto \prod_{t=1}^{n} p(y_t|\lambda_t, v) p(\lambda_t|v) \\ &= \frac{1}{\sqrt{2\pi\lambda_t}} \exp\left\{-\frac{y_t^2/g_t}{2\lambda_t}\right\} \frac{(v/2)^{v/2}}{\Gamma(v/2)} \left(\frac{1}{\lambda_t}\right)^{v/2+1} \exp\left\{-\frac{v/2}{\lambda_t}\right\} \\ &\propto \left(\frac{1}{\lambda_t}\right)^{(v+1)/2+1} \exp\left\{-\frac{y_t^2/g_t + v}{2\lambda_t}\right\}. \end{aligned} \tag{18}$$

Hence we can sample λ_t directly from the inverted Gamma distribution,

$$\lambda_t \sim \mathrm{IG}\big((v+1)/2, \big(y_t^2/g_t + v\big)/2\big), \quad \text{for } t = 1, 2, \ldots, n.$$

When sampling v, we set its prior density as the Gaussian density with known mean μ_v and variance σ_v^2.[3] Given λ, the posterior of v is (up to a normalizing constant)

$$p(v|\lambda) \propto \frac{1}{\sqrt{2\pi\sigma_v^2}} \exp\left\{-\frac{(v-\mu_v)^2}{2\sigma_v^2}\right\} \frac{(v/2)^{v/2}}{\Gamma(v/2)} \left(\frac{1}{\lambda_t}\right)^{v/2+1} \exp\left\{-\frac{v/2}{\lambda_t}\right\}, \tag{19}$$

and v can be sampled using the random-walk Metropolis–Hastings algorithm, in which the proposal density is the standard Gaussian density and the acceptance probability is computed through Eq. (19).

In order to capture the heavy-tailed feature in the marginal distribution of errors in the mean equation, we assumed that the errors follow a mixture of distributions using a particular hierarchical prior. Such a mixture turns into a Student t distribution and has been widely used to capture the heavy-tailed feature in the marginal distribution of asset returns. It is important to note that the heavy-tailed feature of the marginal distribution of errors can be captured through other mixtures of normal densities, and the Gibbs sampling procedure is capable of handling these mixtures.

[3] Geweke (1993) uses $\psi \exp(-\psi v)$ as the prior of the degree-of-freedom parameter in the linear regression model with t distributed errors. A reasonable prior for v should prevent v from getting too large during MCMC iterations, because the update of v has a negligible effect when v is already large. The purpose of such a prior is to put a low prior probability on the "problematic" region, where the likelihood is flat. One can also use $p(v) \propto 1/(1+v^2)$ employed by Bauwens and Lubrano (1998).

3. Bayes factors for model comparison

As the widely documented heavy-tailed asymmetric LSV model is nested in the heavy-tailed asymmetric BCSV model by fixing $\delta=0$, we are interested in examining whether the Box-Cox transformation of squared volatility is preferable to the log-transformation. Bayes factors can be used for such a comparison.

Spiegelhalter and Smith (1982) show that the Bayes factor for a model $\mathcal{M}_0$ against a model $\mathcal{M}_1$ is defined by

$$B_{01} = m(\mathbf{y}|\mathcal{M}_0)/m(\mathbf{y}|\mathcal{M}_1),$$

where $m(\mathbf{y}|\mathcal{M}_0)$ and $m(\mathbf{y}|\mathcal{M}_1)$ are marginal likelihoods under $\mathcal{M}_0$ and $\mathcal{M}_1$, respectively.

Let $\Theta=(v, \theta')'$ denote the parameter vector of the heavy-tailed asymmetric BCSV model. Newton and Raftery (1994) show that the marginal likelihood is defined as

$$m(\mathbf{y}|\mathcal{M}) = \int f(\mathbf{y}|\Theta,\mathcal{M})p(\Theta|\mathcal{M})d\Theta, \tag{20}$$

where $f(\mathbf{y}|\Theta, \mathcal{M})$ is the likelihood of $\mathbf{y}$ given Θ under model $\mathcal{M}$, and $p(\Theta|\mathcal{M})$ is the prior of Θ. As the solution to this integral requires high-dimensional integration, Chib (1995) suggests that the marginal likelihood be expressed as

$$m(\mathbf{y}|\mathcal{M}) = \frac{f(\mathbf{y}|\Theta,\mathcal{M})p(\Theta|\mathcal{M})}{\pi(\Theta|y,\mathcal{M})},$$

where $\pi(\Theta|\mathbf{y}, \mathcal{M})$ is the posterior density of Θ given $\mathbf{y}$ and can be estimated at a given value of Θ.

In SV models, the posterior of Θ given $\mathbf{y}$ under a model $\mathcal{M}$ is generally expressed as

$$\pi(\Theta|\mathbf{y}) = \int\pi(\Theta|\mathbf{y},\alpha)p(\alpha|\mathbf{y})d\alpha,$$

where $\pi(\Theta|\mathbf{y}, \alpha)$ is the posterior of Θ given $(\mathbf{y}, \alpha)$ as expressed in Eq. (11). Chib (1995) shows that at a given value of Θ, an appropriate Monte Carlo estimate of $\pi(\Theta|\mathbf{y})$ is

$$\hat{\pi}(\Theta|\mathbf{y}) = \frac{1}{N}\sum_{i=1}^{N} \pi\left(\Theta|\mathbf{y},\alpha^{(i)}\right),$$

where $\{\alpha^{(i)}: i=1, 2,\ldots, N\}$ are draws recorded from MCMC iterations.[4] Then the logarithmic marginal likelihood can be estimated as

$$\ln\ \hat{m}(\mathbf{y}) = \ln\ f(\mathbf{y}|\Theta) + \ln p(\Theta) - \ln\ \hat{\pi}(\Theta|\mathbf{y}).$$

As $f(\mathbf{y}|\Theta)$ is analytically intractable, it should be either estimated or approximated numerically.

Kim et al. (1998) indicate that for a given value of Θ, the likelihood can be approximated through a particle filter. Let $\mathbf{y}_t=(y_1, y_2,\ldots, y_t)'$, for $t=1, 2,\ldots, n$. Kim et al. (1998) show that by successive conditioning, the log-likelihood of $\mathbf{y}$ given Θ can be expressed as

$$\ln\ L(\mathbf{y}|\Theta) = \ln L(\mathrm{y}_1|\Theta) + \sum_{t=2}^{n}\ \ln L(\mathrm{y}_t|\mathrm{y}_{t-1},\Theta), \tag{21}$$

and that the likelihood function can be approximated though the particle filter algorithm, whose objective is to obtain a sample of draws of α_t conditional on α_{t-1} given θ and a sample of draws $(\alpha_{t-1}^{(1)}, \alpha_{t-1}^{(2)},\ldots, \alpha_{t-1}^{(M)})$. We draw latent variables according to the following equations,

$$\begin{aligned}
&\alpha_1^{(j)} \sim N\left(0, (1-\rho^2)\sigma_u^2/\left(1-\phi^2\right)\right),\\
&\alpha_{t+1}^{(j)} \sim N\left(\mu + \phi\left(\alpha_t^{(j)} - \mu\right), \sigma_u^2\right),\\
&\lambda_t^{(j)} \sim \mathrm{IG}(v/2, v/2),
\end{aligned}$$

[4] Kim et al. (1998) present an alternative approach to the estimation of $p(\Theta|\mathbf{y})$. When deriving a posterior sample of Θ recorded from MCMC iterations, one can estimate $\pi(\Theta|\mathbf{y})$ through multivariate kernel density estimation computed at a given value of Θ. This kernel method does not require extensive computation and is very easy to implement.

for $j=1,2,\ldots,M$, and $t=1, 2,\ldots, n-1$.[5] Given $\alpha_t^{(j)}$, $\alpha_{t+1}^{(j)}$ and $\lambda_t^{(j)}$, we have

$$y_t \sim N\left(\frac{\rho}{\sigma_u}\sqrt{g_t^{(j)}\lambda_t^{(j)}}\left[\alpha_{t+1}^{(j)} - \mu - \phi\left(\alpha_t^{(j)} - \mu\right)\right], g_t^{(j)}\lambda_t^{(j)}\left(1-\rho^2\right)\right), \tag{22}$$

for $t=1, 2,\ldots, n-1$, and $y_n \sim N(0, g_n^{(j)}\lambda_n^{(j)}(1-\rho^2))$, where $g_t^{(j)}$ is computed according to Eq. (6) with α_t replaced by $\alpha_t^{(j)}$. According to the particle filter, the estimate of $L(y_t|\mathbf{y}_{t-1}, \Theta)$ is approximated by

$$\hat{L}(y_t|\mathbf{y}_{t-1}, \Theta) = \frac{1}{M}\sum_{j=1}^{M} p\left(y_t|\lambda_t^{(j)}, \alpha_t^{(j)}, \alpha_{t+1}^{(j)}\right), \tag{23}$$

where $p(y_t|\lambda_t^{(j)},\alpha_t^{(j)},\alpha_{t+1}^{(j)})$ is the density of y_t conditional on $\lambda_t^{(j)}$, $\alpha_t^{(j)}$, $\alpha_{t+1}^{(j)}$ and is given by Eq. (22). Hence the likelihood of $\mathbf{y}$ given Θ can be approximated by substituting Eq. (23) into Eq. (21), and the logarithmic marginal likelihood can be approximated by

$$\ln \hat{m}(\mathbf{y}|\mathcal{M}) = \ln \hat{f}(\mathbf{y}|\theta, \mathcal{M}) + \ln p(\theta|\mathcal{M}) - \ln \hat{\pi}(\theta|\mathbf{y}, \mathcal{M}). \tag{24}$$

The Bayes factor for model $\mathcal{M}_0$ against an alternative model $\mathcal{M}_1$ can be estimated as

$$\hat{B}_{01} = \exp\{\ln\ \hat{m}(\boldsymbol{y}|\mathcal{M}_0) - \ln\ \hat{m}(\boldsymbol{y}|\mathcal{M}_1)\}. \tag{25}$$

The value of Θ in the computation of $\ln \hat{m}(\mathbf{y})$ can be arbitrary. However, we follow Chib's (1995) suggestion of using the posterior averages of Θ obtained through MCMC iterations under $\mathcal{M}_0$ and $\mathcal{M}_1$, respectively.

4. Application to the Australian stock index

In this section, we present an application of the heavy-tailed asymmetric BCSV model to daily returns of the Australian All Ordinaries stock index, whose historical data were down-loaded from yahoo.com. As required by the construction of SV models, the continuously compounded daily returns were mean-corrected and variance-scaled. The data set contains 1508 observations from the 2nd of January 2000 to the 30th of December 2005, excluding weekends and holidays.

4.1. Estimation of the heavy-tailed BCSV model

We applied the proposed sampling algorithm to the heavy-tailed asymmetric BCSV model of daily returns of the index. The hyperparameters required in the joint prior density were set, respectively, to $\omega=20$, $\gamma=1.5$, $\zeta=2$, $S_\tau=0.01$, $\varphi_0=0$, $\mu_0=0$, $p_0=0.5$ and $q_0=0.2$. The prior density of δ is the uniform density over $(-2, 2)$, and the prior density of v is a normal density with mean 20 and variance 25.[6]

The burn-in period contained 20,000 iterations, and the recorded period contained 200,000 iterations. To measure the mixing performance, we calculated the batch-mean standard error (BMSE) for each component of $(v, \theta', \alpha', \lambda')'$, where the number of batches was 50 and there were 4000 draws in each batch (see, e.g., Roberts, 1996; Tse et al., 2004). The ergodic average (or posterior mean) of the retained draws acts as an estimator of $(v, \theta', \alpha', \lambda')'$.

The sampling algorithm is the same as the one reported by Zhang and King (2004), who showed that the sampler has mixed very well. When we finished a sampling procedure, we did check the mixing performance of the sampling algorithm. However, in order to save space, we will not report the graphs and statistics used by Zhang and King (2004) for the purpose of examining the mixing performance of simulated chains. The sample consisting of draws retained from MCMC iterations is referred to as the posterior sample, whose mean value is used as an estimate of $(v, \theta', \alpha', \lambda')'$ and is presented in the second column of Table 1. The estimates of (σ_t, λ_t), for $t=1, 2,\ldots, n$, are presented in Fig. 1.

---

[5] The particle filter algorithm presented in Kim et al. (1998) is applicable to LSV models. We need to modify the algorithm and make it applicable to the heavy-tailed BCSV model.

[6] The empirical evidence found by Jacquier et al. (2004) showed that the posterior mean of v for most return series is between 10 and 30, so we let the prior be centered at 20. Zhang and King (2004) choose the prior of v to be proportional to $\psi \exp(-\psi n)$ with $\psi=0.2$.

Table 1
Estimates of parameters obtained from daily returns of the Australian stock index through the heavy-tailed asymmetric BCSV model and its competing models

Parameters	model (17)	model (26)	model (27)	model (28)	model (29)
ϕ	0.95692	0.94968	0.96088	0.97703	$a_0=0.02156$
δ	−0.46914	–	−0.35669	−0.53609	$a_1=0.07376$
μ	−0.94660	−0.69531	−0.75067	−0.77191	$b_1=0.85913$
ρ	−0.43346	−0.44474	−0.63084	–	$v=8.62873$
σ_u	0.25621	0.21317	0.22000	0.19833	
v	22.72435	22.11213	–	22.57576	

4.2. Estimation of several competing models

4.2.1. Heavy-tailed asymmetric LSV model

In order to investigate whether the heavy-tailed asymmetric BCSV model yields very different results from a model without the Box-Cox transformation, we fixed $\delta=0$ in the heavy-tailed asymmetric BCSV model, which turns to the heavy-tailed asymmetric LSV model expressed as follows.

$$\begin{aligned} y_t &= \exp\,(\alpha_t/2)\lambda_t^{1/2}\varepsilon_t, \\ \alpha_{t+1} &= \mu + \phi(\alpha_t - \mu) + \sigma_u u_{t+1}, \end{aligned} \tag{26}$$

where $\varepsilon_t \sim N(0, 1)$, $u_{t+1} \sim N(0, 1)$, $\mathrm{cov}(\varepsilon_t, u_{t+1})=\rho$ and $\lambda_t \sim \mathrm{IG}(v/2, v/2)$.

Using the same prior densities of parameters as those used for Eq. (17) and fixing $\delta=0$, we applied the sampling algorithm to the heavy-tailed asymmetric LSV model with the same data set. A summary of the posterior average of the parameter vector is presented in the third column of Table 1, where the estimate of μ is obviously different from that estimated through Eq. (17), and the estimated σ_u is smaller than that estimated through Eq. (17).

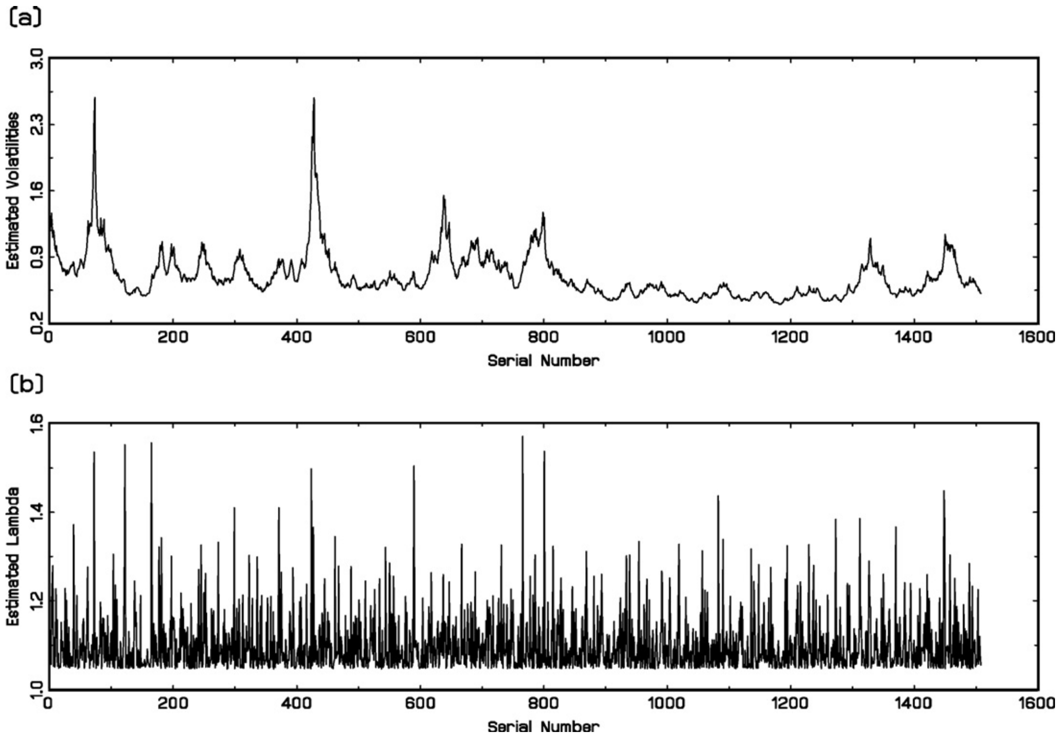

Fig. 1. The estimated volatilities and λ_t's obtained from daily returns of the Australian stock index.

4.2.2. Asymmetric BCSV model

In order to investigate whether the heavy-tailed asymmetric BCSV model yields very different results from a model with normal errors, rather than the heavy-tailed errors, we fixed $\lambda_t=1$, for $t=1, 2,\ldots, n$, in the heavy-tailed asymmetric BCSV model, which becomes the asymmetric BCSV model expressed as follows.

$$\begin{aligned} y_t &= g_t^{1/2}\varepsilon_t, \\ \alpha_{t+1} &= \mu + \phi(\alpha_t - \mu) + \sigma_u u_{t+1}, \end{aligned} \tag{27}$$

where $\varepsilon_t \sim N(0, 1)$, $u_{t+1} \sim N(0, 1)$, and $\mathrm{cov}(\varepsilon_t, u_{t+1})=\rho$.

Using the same prior densities of parameters as those used in Eq. (17) and fixing $\lambda_t=1$, for $t=1, 2,\ldots, n$, we applied the sampling algorithm to the asymmetric BCSV model with the same data set. The fourth column of Table 1 presents the posterior average of the parameter vector, which indicates that the estimates of ρ, σ_u and μ are obviously different from those estimated through Eq. (17), in particular, the estimated ρ is much larger in magnitude than that estimated through Eq. (17).

4.2.3. Heavy-tailed BCSV model

The correlation coefficient between the regression errors and volatility errors captures the so-called leverage effect. In order to investigate the effect of allowing for such a correlation on the estimation of parameters and volatilities, we let $\rho=0$ in the heavy-tailed asymmetric BCSV model, which becomes the following model,

$$\begin{aligned} y_t &= g_t^{1/2}\lambda_t^{1/2}\varepsilon_t, \\ \alpha_{t+1} &= \mu + \phi(\alpha_t - \mu) + \sigma_u u_{t+1}, \end{aligned} \tag{28}$$

where $\varepsilon_t \sim N(0, 1)$, $u_{t+1} \sim N(0, 1)$, $\mathrm{cov}(\varepsilon_t, u_{t+1})=0$ and $\lambda_t \sim \mathrm{IG}(v/2, v/2)$.

Using the same prior densities of $(\phi, \delta, \mu, \sigma^2, v)$ as those used in Eq. (17) and fixing $\rho=0$, we applied the sampling algorithm to the heavy-tailed BCSV model with the same data set. The posterior average of the parameter vector is presented in the fifth column of Table 1.

4.2.4. GARCH model with Student t errors

An obvious alternative to SV models is the GARCH model, so it is of interest to see how it compares to our heavy-tailed asymmetric BCSV model. As the GARCH(1,1) specification with Student *t* errors is one of the most popular GARCH models in empirical studies, we use this model for comparison purposes. Consider a GARCH(1,1) model given by

$$\begin{aligned} y_t &= h_t^{1/2}\varepsilon_t, \\ h_t &= a_0 + a_1 y_{t-1}^2 + b_1 h_{t-1}, \end{aligned} \tag{29}$$

where ε_t, for $t=1, 2,\ldots, n$, are independent and identically distributed as *t* distribution with v degrees of freedom. Eq. (29) is referred to as the *t*-GARCH(1,1) model hereafter. We assume that $a_0 \geq 0$, $a_1 \geq 0$, $b_1 \geq 0$ and $a_1+b_1<1$ to ensure positive h_t. Given (a_0, a_1, b_1, v), the likelihood of $(y_1, y_2,\ldots, y_n)'$ is expressed as

$$l(y_1, y_2, \ldots, y_n | a_0, a_1, b_1, v) = \prod_{t=1}^{n} \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{v\pi}} \prod_{t=1}^{n} \frac{1}{\sqrt{h_t}} \left(1 + \frac{y_t^2}{v h_t}\right)^{(v+1)/2}. \tag{30}$$

We assume that the prior of a_0 is the log-normal density with mean zero and variance one, while the priors of a_1 and b_1 are the uniform density on (0,1). The prior of v is assumed to be $N(0, 1)$. The joint prior density of a_0, a_1, a_2 and v is the product of these prior densities, and the posterior density of $(y_1, y_2,\ldots, y_n)'$ is proportional to the product of likelihood and the joint prior. We employed the random-walk Metropolis–Hastings algorithm to derive a sample of retained draws of (a_0, a_1, b_1, v). The estimate of (a_0, a_1, b_1, v) is presented in the last column of Table 1.

According to the estimation results given in Table 1, we found that the estimated values of σ_u in three competing stochastic volatility models are obviously smaller than the estimated value of σ_u in the heavy-tailed asymmetric BCSV model. However, we are not sure which estimate is reasonable at the present stage.

The graphs of volatility processes estimated by, respectively, the heavy-tailed asymmetric BCSV model and each competing SV model are presented in Fig. 2. It can be found from Fig. 2(a) that the heavy-tailed asymmetric BCSV and LSV models produce similar estimates of volatilities. However, Fig. 2(a) and (b) indicate that volatility process estimated by the heavy-tailed asymmetric BCSV model is different from that estimated by either the asymmetric BCSV model or the heavy-tailed BCSV model.

The marginal densities of the squared volatility implied by the four SV models are plotted in Fig. 3. The marginal density of the squared volatility implied by the heavy-tailed asymmetric BCSV model is obviously different from that implied by the heavy-tailed asymmetric LSV model. The left tail of the former is thicker than that of the latter, while the right tail of former is thinner. As what we will discuss in Section 4.4, a negative value of the Box-Cox transformation parameter allows for such a thick left tail in the marginal density of the squared volatility implied by the heavy-tailed asymmetric BCSV model.

The location of the marginal density implied by the asymmetric BCSV model with normal errors is obviously different from that of the heavy-tailed asymmetric BCSV model. When the correlation that captures the leverage effect is ignored, the marginal density implied by Eq. (28) has a higher peak than, and a different location from, that implied by the heavy-tailed asymmetric BCSV model.

4.3. A comparison of estimated models

Let $\mathcal{M}_0$ denote the heavy-tailed asymmetric BCSV model given by Eq. (17), and $\mathcal{M}_1$ denote any competing model. As discussed in Section 3, the Bayes factor is expressed as the ratio of marginal likelihoods of $\mathcal{M}_0$ and $\mathcal{M}_1$. In order to compute the logarithmic marginal likelihood given by Eq. (24), the posterior density of parameters is estimated by its kernel density obtained through the posterior sample, where bandwidths were chosen using the Bayesian method proposed by Zhang, King and Hyndman (2006).

Table 2 presents the values of Bayesian information criterion (BIC) or the Schwarz criterion computed from $\mathcal{M}_0$ and all competing models, as well as Bayes factors for $\mathcal{M}_0$ against competing models, respectively. According to Kass and Raftery's (1995) guidelines for interpreting Bayes factors, when the logarithmic Bayes factor is greater than 1 and less

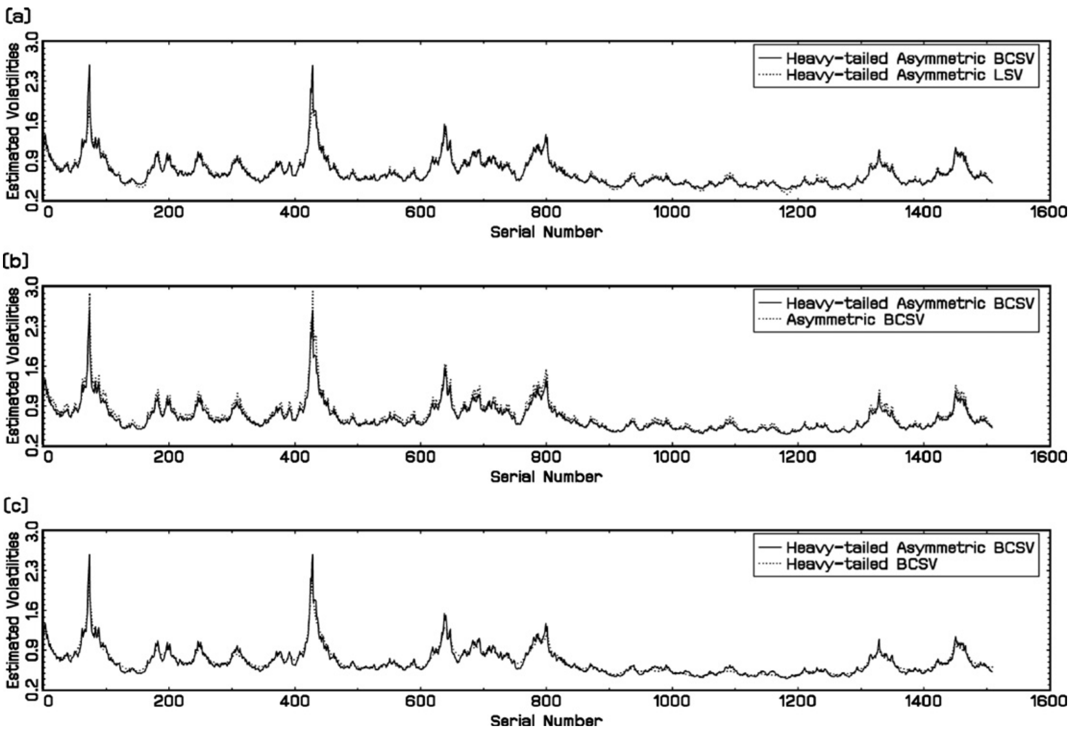

Fig. 2. Comparison of estimated volatilities for the Australian stock index.

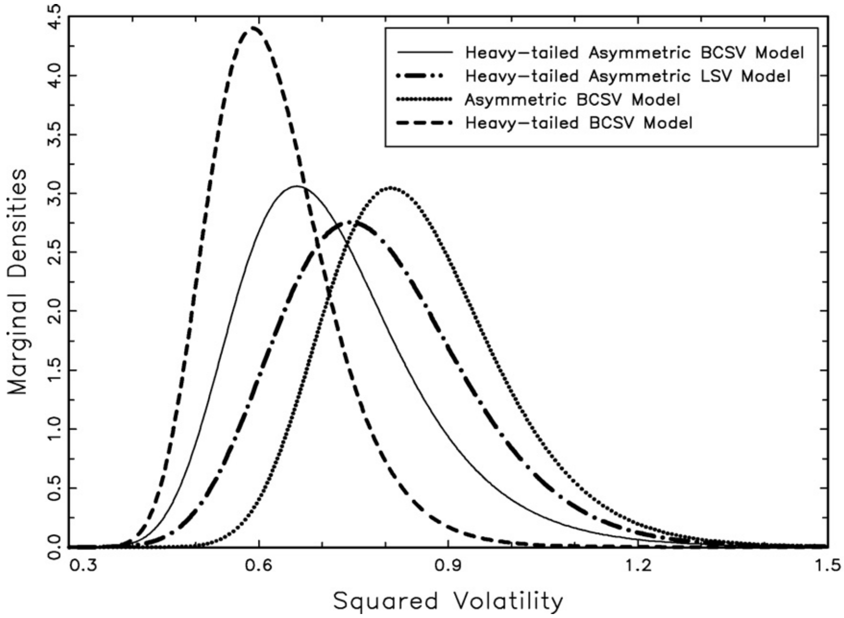

Fig. 3. Marginal densities of the squared volatility implied by SV models for the Australian stock index.

than 3, the null model is positively favored; when the logarithmic Bayes factor is greater than 3 and less than 5, the null model is strongly favored; and when the logarithmic Bayes factor is greater than 5, the null model is very strongly favored.

The logarithmic Bayes factor for the heavy-tailed asymmetric BCSV model against the heavy-tailed asymmetric LSV model given by Eq. (26) is 18.03, which is very large and very strongly in favor of the former model. This finding indicates that it is important to introduce a Box-Cox transformation to squared volatilities rather than the log-transformation in SV models. When the heavy-tailed asymmetric BCSV model is compared with the asymmetric BCSV model given by Eq. (27), we found that the logarithmic Bayes factor is 4.92, which is also a large value and strongly favors the former model. It suggests that incorporating heavy-tailedness into the error distribution in the mean equation is supported by the data. It is also consistent with the findings of Jacquier et al. (2004), who introduced a similar heavy-tailed assumption into the error distribution in their asymmetric LSV model.

The logarithmic Bayes factor of the heavy-tailed asymmetric BCSV model against the heavy-tailed BCSV model given by Eq. (28) is 21.44, another large value that very strongly favors the former model. This finding indicates the importance of incorporating the leverage effect into the BCSV model. However, the logarithmic Bayes factor for the heavy-tailed asymmetric BCSV model against the t-GARCH(1,1) model is −66.26, which very strongly favors the latter model. Kim et al. (1998) find similar results when applying the LSV model and the t-GARCH(1,1) model to returns of some currency prices.

In terms of Schwarz criterion, the BIC value of the asymmetric BCSV model given by Eq. (27) is 3565.59, the smallest value in all four SV models. The BIC value of the heavy-tailed asymmetric BCSV model is 3573.01, which is the second smallest value in all four SV models. Thus, BIC favors the asymmetric BCSV model against the heavy-tailed asymmetric BCSV model. This result is not surprising, because BIC tends to put a heavy penalty on model complexity. However, the BIC values of the other two competing SV models are much larger than those of the

Table 2
BIC of each model, logarithmic Bayes factors of the heavy-tailed asymmetric BCSV model against competing models and ψ computed through daily returns of the Australian stock index

	model (17)	model (26)	model (27)	model (28)	model (29)
BIC	3573.01	3607.99	3565.59	3612.09	3405.09
Bayes factors	–	18.03	4.92	21.44	−66.26
Ψ	0.3576	0.3986	0.4092	0.2362	3.6781

asymmetric BCSV model and the heavy-tailed asymmetric BCSV model. Hence, this finding also highlights the importance of the Box-Cox transformation of the squared volatility. However, the t-GARCH(1,1) model has a much smaller BIC value than all four SV models. This result is consistent with what was revealed by Bayes factors.

A further goodness-of-fit measure is how well the model predicts the distribution of its fitted residuals. Let m_i, for $i=1,2,3,4$, denote the mean, variance, skewness and kurtosis, respectively, of the model disturbances (ε_t), and let $\hat{m}_i$ denote the estimate of m_i obtained from the fitted residuals. To measure the closeness of the observed residual distribution to the model disturbance distribution, we use the statistic

$$\Psi = \sum_{i=1}^{4} (\hat{m}_i - m_i)^2. \tag{31}$$

Table 2 presents the closeness measures computed, respectively, from the five models. The closeness measures of the SV models are of a similar magnitude, which is much smaller than the closeness measure computed from the t-GARCH(1,1) model. Hence, our model and the other three SV models fit the data better than the t-GARCH(1,1) model in the sense that the residuals predicted by the fitted SV models are closer to the hypothesized disturbance distribution.

4.4. The role of the Box-Cox transformation parameter

It is important to note that the estimated δ is −0.46914, a negative value. To understand the effect of this type of Box-Cox transformation, we present graphs of $f(x)=\ln x$ and $f(x)=(x^{\delta}-1)/\delta$ with $\delta=-0.46914$ in Fig. 4(a) and (b). The Box-Cox transformation produces a much smaller value than the log-transformation for any given value of x between 0 and 0.3, while the former also produces a smaller value than the latter for any given value of x larger than 1.5. Moreover, the difference between these two transformations is larger when x is between 0 and 0.3 than that when x is above 1.5.

When the Box-Cox transformation and log-transformation are applied to SV models with x replaced by σ_t^2, the date-t squared volatility, it can be seen that the Box-Cox transformation can make the tails of volatility distribution

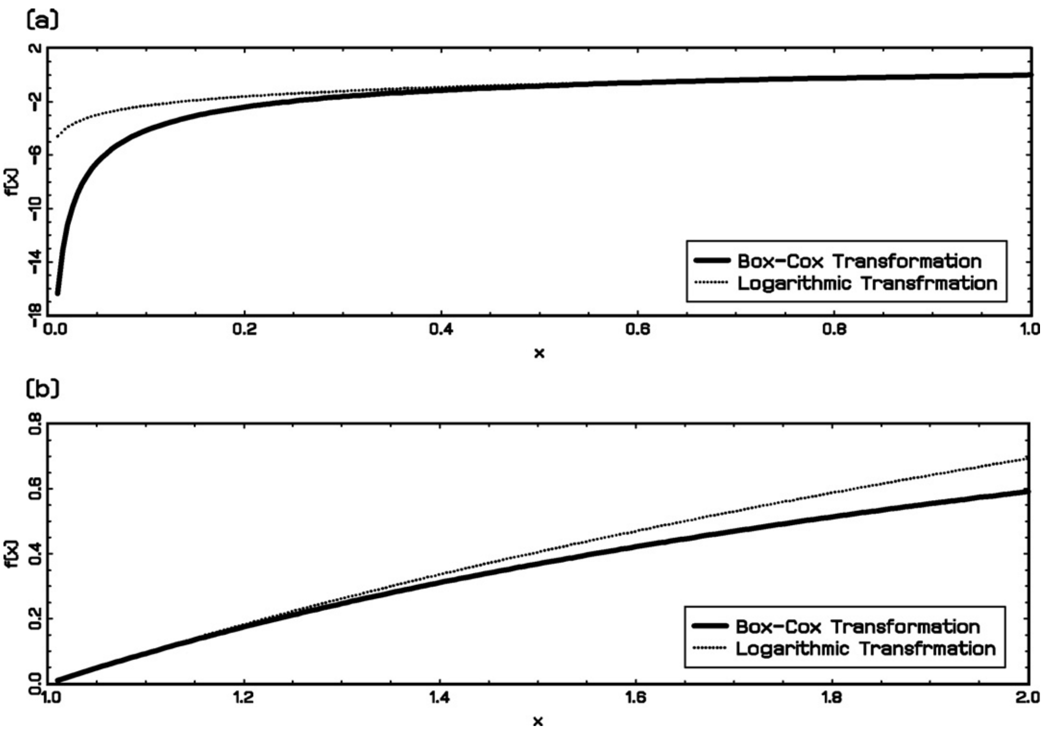

Fig. 4. A graphical comparison of the Box-Cox and logarithmic transformations.

thinner than the log-transformation, especially for the left tail. If the estimated δ is negative and is large in magnitude, it means that the marginal density of σ_t^2 has a thicker left tail than right tail, and both tails are thicker than tails of the Gaussian density. According to the graphical comparison between the marginal densities of σ_t^2 implied by Eq. (17) and Eq.(26) presented in Fig. 3, the former model allows for a thicker left tail in the marginal density of σ_t^2 than the latter.

In terms of the marginal density of σ_t^2, if the left tail is found to be thicker than the right tail, it means the empirical frequency of low volatility is higher than that of high volatility. Thus, the Box-Cox transformation of σ_t^2 captures the thick-tailed and asymmetric feature in the marginal distribution of σ_t^2. As the estimated value of δ is negative and is large in magnitude, we speculate that low volatilities might be more frequent than high volatilities in daily returns of the Australian stock index.

4.5. Sensitivity of priors

To examine the sensitivity of MCMC output to prior choices, we also chose the prior density of v as $\psi\ \exp(-\psi v)$ with $\psi=0.2$. When we applied the sampling algorithm to the daily returns of the Australian All Ordinaries stock index,

Table 3
Estimates of parameters obtained from daily returns of five international stock indices through the heavy-tailed asymmetric BCSV model and its competing models

Parameters	model (17)	model (26)	model (27)	model (28)	model (29)
DAX					
ϕ	0.98690	0.98589	0.98949	0.99323	$a_0=0.00803$
δ	−0.18660	–	−0.06531	−0.11092	$a_1=0.08829$
μ	−0.35403	−0.29789	−0.23644	−0.18610	$b_1=0.89171$
ρ	−0.53802	−0.53918	−0.81137	–	$v=17.89008$
σ_u	0.15919	0.15079	0.13662	0.11351	
v	25.32530	25.32590	–	25.07363	
DJIA					
ϕ	0.98694	0.98438	0.99104	0.99476	$a_0=0.01262$
δ	−0.37444	–	−0.23796	−0.54628	$a_1=0.07499$
μ	−0.45299	−0.35607	−0.33286	−0.22109	$b_1=0.89485$
ρ	−0.51588	−0.52873	−0.83017	–	$v=12.83781$
σ_u	0.15845	0.15205	0.13888	0.10382	
v	23.76762	23.72630	–	23.51314	
FTSE					
ϕ	0.98324	0.98125	0.98633	0.99180	$a_0=0.01127$
δ	−0.25771		−0.24841	−0.27650	$a_1=0.10337$
μ	−0.57297	−0.45370	−0.50905	−0.37075	$b_1=0.87030$
ρ	−0.54123	−0.53618	−0.85711	–	$v=17.16026$
σ_u	0.18222	0.17603	0.16221	0.13050	
v	24.93316	24.99851	–	24.74711	
Nikkei					
ϕ	0.97437	0.97550	0.97082	0.98854	$a_0=0.01478$
δ	0.13415		0.25566	0.19742	$a_1=0.06368$
μ	−0.20991	−0.23300	−0.12192	−0.14764	$b_1=0.90739$
ρ	−0.34698	−0.33642	−0.51374	–	$v=11.98270$
σ_u	0.13653	0.14065	0.14430	0.08582	
v	21.94532	22.02149	–	20.55171	
S&P500					
ϕ	0.98229	0.98043	0.98823	0.99316	$a_0=0.01152$
δ	−0.37028	–	−0.28185	−0.41001	$a_1=0.07608$
μ	−0.57535	−0.40557	−0.47091	−0.34721	$b1=0.89790$
ρ	−0.56618	−0.56261	−0.87168		$v=14.55547$
σ_u	0.17552	0.16210	0.14746	0.10289	
v	24.78086	24.41744	–	23.99362	

we found that the posterior averages of parameters and latent processes are quite similar to those obtained previously. The prior density of δ is noninformative, and we found no obvious changes in the MCMC output when using a flat Gaussian prior namely $N(0, 10)$.

As our heavy-tailed asymmetric BCSV model nests the heavy-tailed asymmetric LSV model by fixing $\delta=0$, one may wonder how the alternative priors used by Jacquier et al. (2004) for their SV model with heavy-tails works. Hence we assumed that $\varphi \sim N(0, 10)$, $\mu \sim (0, 100)$, $\phi \sim N(0, \tau^2/2)$ and $\tau^2 \sim \mathrm{IG}(1, 0.005)$. The prior of v is assumed to be the discrete uniform density on (3, 40), and the prior of δ is assumed to be the uniform density on $(-2, 2)$. Using these priors, we obtained the posterior estimate of the parameter vector under our model. We found that $\phi=0.96510$, $\delta=-0.53852$, $\mu=-1.03169$, $\rho=-0.47213$, $\sigma=0.23832$ and $v=38$, which are quite close to those obtained in Section 4.1 except for the estimated v. In addition, the estimates of volatilities are quite similar to those obtained in Section 4.1.

5. Applications to some international stock indices

In this section, we applied the heavy-tailed asymmetric BCSV model and four competing models given in Section 4 to daily returns of the DAX, Dow Jones Industrial Average (DJIA), FTSE, Nikkei 225 and S&P500 indices, respectively. These data sets were downloaded from yahoo.com, and the sample period is from the 2nd of January 2000 to the 30th of December 2005, excluding weekends and holidays. The continuously compounded daily returns were mean-corrected and variance-scaled.

Table 3 presents the estimates of parameters under different models for all five data sets. Under the heavy-tailed asymmetric BCSV model given by Eq. (17), the estimated values of δ are negative for DAX, DJIA, FTSE and S&P500, while the estimated δ is positive for Nikkei 225. In Table 4, we present the BIC values computed from all five models, and the logarithmic Bayes factors for the heavy-tailed asymmetric BCSV model against its competing models, for all five data sets.

The logarithmic Bayes factor for the heavy-tailed asymmetric BCSV model given by Eq. (17) against the heavy-tailed asymmetric LSV model given by Eq. (26) are very large positive values in all five data sets, therefore, we can conclude that the former model is strongly favored against the latter. This is strong evidence supporting the Box-Cox transformation of the squared volatility against the log-transformation. Moreover, in any data set, the BIC value computed from the former model is much smaller than that computed from the latter, therefore, the heavy-tailed asymmetric BCSV model is preferable.

The logarithmic Bayes factor for the heavy-tailed asymmetric BCSV model against the asymmetric BCSV model given by Eq. (27) are also positive values in all five data sets. This finding highlights the importance of the heavy-tailed assumption in the marginal density of asset returns. However, the former model has a larger BIC value than the latter for each data set, and on this criterion the latter should be favored. This result might be partly due to the fact that BIC

Table 4
BIC of each model, logarithmic Bayes factors of the heavy–tailed asymmetric BCSV model against competing models and Ψ computed through daily returns of five international stock indices

		model (17)	model (26)	model (27)	model (28)	model (29)
DAX	BIC	4239.31	4310.77	4224.53	4302.27	3754.21
	Bayes factors	–	35.00	2.04	34.43	−221.18
	Ψ	0.4995	0.5174	0.5398	0.3239	0.1862
DJIA	BIC	4183.87	4238.89	4181.15	4466.75	3879.29
	Bayes factors	–	26.87	7.78	144.19	−131.33
	Ψ	0.4163	0.4308	0.5289	0.2189	0.2395
FTSE	BIC	4143.47	4290.23	4133.79	4256.55	3423.81
	Bayes factors	–	73.82	6.14	58.78	−340.59
	Ψ	0.5160	0.5527	0.6357	0.3217	0.1696
Nikkei 225	BIC	4158.79	4177.55	4156.51	4235.25	4037.27
	Bayes factors	–	7.44	6.21	41.48	−46.96
	Ψ	0.2595	0.2662	0.1206	0.1565	0.3635
S&P500	BIC	4200.71	4279.55	4197.23	4333.05	3915.45
	Bayes factors	–	40.22	7.78	68.56	−129.34
	Ψ	0.5246	0.5259	0.6369	0.2573	0.0172

imposes a heavy penalty on model complexity, and partly due to the slightness of the advantage of the former model against the latter.

When comparing the heavy-tailed asymmetric BCSV model with the heavy-tailed BCSV model given by Eq. (28), we found that the logarithmic Bayes factors for the former model against the latter are very large positive values for all five data sets, therefore, the former is more preferable than the latter. This conclusion can be also derived according to BIC.

When the heavy-tailed asymmetric BCSV model is compared with the t-GARCH(1,1) model, the logarithmic Bayes factor for the former model against the latter are negative and large in magnitude for each data set. Clearly, Bayes factors strongly favor the latter against the former, and so does BIC. With respect to the measure of closeness of the observed residual distribution to the model disturbance distribution given by Eq. (31), we found that our model is slightly better than the t-GARCH(1,1) model for the Nikkei 225 data set and is slightly poorer than the t-GARCH(1,1) for the DAX, DJIA, FTSE and S&P500 data sets.

6. Conclusion

We present an MCMC algorithm for sampling parameters and latent stochastic processes of the heavy-tailed asymmetric BCSV model. The widely used log-transformation of squared volatilities is nested in the Box-Cox transformation. Bayes factors, BIC and a measure of closeness of the residual distribution to the model disturbance distribution have been used to examine whether the Box-Cox transformation of squared volatility is favored against the log-transformation. When applying the heavy-tailed asymmetric BCSV model and three competing SV models to continuously compounded daily returns of the Australian stock index, we have found significant evidence supporting the Box-Cox transformation of squared volatility against the log-transformation. While the t-GARCH(1,1) model is strongly favored by Bayes factors and BIC against the four SV models discussed in this paper, the four SV models fit the data better than the t-GARCH(1,1) model in the sense that residuals predicted by the fitted model are closer to the hypothesized disturbance distribution.

When applying the heavy-tailed asymmetric BCSV model and four competing models to continuously compounded daily returns of another five stock indices, we have found significant evidence supporting the Box-Cox transformation of the squared volatility against the log-transformation in terms SV models. Even though the t-GARCH(1,1) model is strongly favored by Bayes factors and BIC against four SV models, we have found that SV models fit daily returns of Nikkei 225 better than the t-GARCH(1,1) model based on the closeness of the fitted model residuals to error distribution of the model. It does appear there are circumstances, in which the heavy-tailed asymmetric Box-Cox transformed SV model has some advantages over alternative models.

Acknowledgements

We wish to thank the Editor and two anonymous referees for their very insightful comments that have led to an improved paper. We extend our sincere thanks to Siddhartha Chib, Eric Jacquier, Neil Shephard, Zhenlin Yang and Jun Yu for helpful discussions through emails. We greatly appreciate comments from Tim Bollerslev, Robert Brooks, Gael Martin and Vance Martin, the computational support from the Victorian Partnership for Advanced Computing, and financial support from the Australian Research Council.

References

Andersen, T., 1994. Stochastic autoregressive volatility: a framework for volatility modelling. Mathematical Finance 4, 75–102.

Bauwens, L., Lubrano, M., 1998. Bayesian inference on GARCH models using the Gibbs sampler. Econometrics Journal 1, C23–C26.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics 31, 307–327.

Bollerslev, T., Zhou, H., 2005. Volatility puzzles: a simple framework for gauging return-volatility regressions. Journal of Econometrics 131, 123–150.

Campbell, J.Y., Kyle, A.S., 1993. Smart money, noise trading and stock price behaviour. Review of Economic Studies 60, 1–34.

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

Clark, P.K., 1973. A subordinated stochastic process model with finite variance for speculative price. Econometrica 41, 135–155.

Engle, R.F., 1982. Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inνation. Econometrica 50, 987–1007.

Engle, R.F., Ng, V.K., 1993. Measuring and testing the impact of news on volatility. The Journal of Finance 48, 1749–1801.

Eraker, B., Johannes, M.S., Polson, N.G., 2003. The impact of jumps in volatility and returns. The Journal of Finance 58, 1269–1300.

Gallant, A.R., Tauchen, G., 1996. Which moments to match? Econometric Theory 12, 657–681.
Gallant, A.R., Rossi, P.E., Tauchen, G., 1992. Stock price and volume. Review of Financial Studies 5, 199–242.
Gallant, A.R., Rossi, P.E., Tauchen, G., 1993. Nonlinear dynamic structures. Econometrica 61, 871–907.
Gallant, A.R., Hsieh, D., Tauchen, G., 1997. Estimation of stochastic volatility models with diagnostics. Journal of Econometrics 81, 159–192.
Geweke, J., 1993. Bayesian treatment of the independent Student-*t* linear model. Journal of Applied Econometrics 8, S19–S40.
Ghysels, E., Harvey, A.C., Renault, E., 1996. In: Maddala, G.S., Rao, C.R. (Eds.), Stochastic volatility. Handbook of Statistics, vol. 14. North-Holland, Amsterdam, pp. 119–191.
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.
Heston, S.L., 1993. A closed-form solution for options with stochastic volatility, with application to bond and currency options. Review of Financial Studies 6, 327–343.
Higgins, M.L., Bera, A.K., 1992. A class of nonlinear ARCH models. International Economic Review 33, 137–158.
Hull, J., White, A., 1987. The pricing of options on assets with stochastic volatilities. Journal of Finance 42, 281–300.
Jacquier, E., Polson, N.G., Rossi, P.E., 1994. Bayesian analysis of stochastic volatility models. Journal of Business and Economic Statistics 12, 371–389.
Jacquier, E., Polson, N.G., Rossi, P.E., 2004. Bayesian analysis of stochastic volatility models with fat-tails and correlated errors. Journal of Econometrics 122, 185–212.
Johannes, M.S., Polson, N.G., in press. MCMC methods for financial econometrics. In: Aït-Sahalia, Y., Hansen, L. (Eds.), Handbook of Financial Econometrics.
Kass, R.E., Raftery, A.E., 1995. Bayes factors. Journal of the American Statistical Association 90, 773–795.
Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361–393.
Nelson, D., 1990. ARCH models as diffusion approximations. Journal of Econometrics 45, 7–38.
Newton, M.A., Raftery, A.E., 1994. Approximation Bayesian inference by the weighted likelihood bootstrap (with discussion). Journal of the Royal Statistical Society, Series B 56, 3–48.
Roberts, G.O., 1996. Markov chain concepts related to sampling algorithms. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), Markov Chain Monte Carlo in Practice. Chapman & Hall, London, pp. 45–57.
Spiegelhalter, D.J., Smith, A.F.M., 1982. Bayes factors for linear and log-linear models with vague prior information. Journal of the Royal Statistical Society, Series B 44, 377–387.
Stein, E.M., Stein, J.C., 1991. Stock price distributions with stochastic volatility: an analytical approach. Review of Financial Studies 4, 727–752.
Tauchen, G., Pitts, M., 1983. The price variability-volume relationship on speculative markets. Econometrica 51, 485–505.
Taylor, S.J., 1982. In: Anderson, O.D. (Ed.), Financial returns modelled by the product of two stochastic processes A study of the daily sugar prices 1961–75. Time Series Analysis: Theory and Practice, vol. 1. North-Holland, Amsterdam, pp. 203–226.
Taylor, S.J., 1986. Modelling Financial Time Series. Wiley, Chichester.
Taylor, S.J., 1994. Modelling stochastic volatility: a review and comparative study. Mathematical Finance 4, 183–204.
Tse, Y.K., Zhang, X., Yu, J., 2004. Estimation of hyperbolic discussion using the Markov chain Monte Carlo method. Quantitative Finance 4, 158–169.
Yu, J., Yang, Z., Zhang, X., 2006. A class of nonlinear stochastic volatility models and its implication for pricing currency options. Computational Statistics and Data Analysis 51, 2218–2231.
Zhang, X., King, M.L., 2004, Box-Cox stochastic volatility models with heavy-tails and correlated errors. Working Paper Series 26/04, Department of Econometrics and Business Statistics, Monash University. (http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/2004/wp26-04.pdf)
Zhang, X., King, M.L., Hyndman, R.J., 2006. A Bayesian approach to bandwidth selection for multivariate kernel density estimation. Computational Statistics and Data Analysis 50, 3009–3021.