Statistics and Its Interface Volume 7 (2014) 487–502

Bayesian inference for stochastic volatility models using the generalized skew-t distribution with applications to the Shenzhen Stock Exchange returns*

Carlos A. Abanto-Valle, Caifeng Wang, Xiaojing Wang†, Fei-Xing Wang, and Ming-Hui Chen

In this paper, we propose a new stochastic volatility model based on a generalized skew-Student-t distribution for stock returns. This new model allows a parsimonious and flexible treatment of the skewness and heavy tails in the conditional distribution of the returns. An efficient Markov chain Monte Carlo (MCMC) sampling algorithm is developed for computing the posterior estimates of the model parameters. Value-at-Risk (VaR) and Expected Shortfall (ES) forecasting via a computational Bayesian framework are considered. The MCMC-based method exploits a skew-normal mixture representation of the error distribution. The proposed methodology is applied to the Shenzhen Stock Exchange Component Index (SZSE-CI) daily returns. Bayesian model selection criteria reveal that there is a significant improvement in model fit to the SZSE-CI returns data by using the SV model based on a generalized skew-Student-t distribution over the usual normal and Student-t models. Empirical results show that the skewness can improve VaR and ES forecasting in comparison with the normal and Student-t models. We demonstrate that the generalized skew-Student-t tail behavior is important in modeling stock returns data.

Keywords and phrases: Bayesian predictive information criterion (BPIC), Deviance information criterion (DIC), Log predictive score criterion, Markov chain Monte Carlo, Non-Gaussian and nonlinear state space models, Expected Shortfall, Value-at-Risk.

1. INTRODUCTION

Stochastic volatility (SV) models have been considered as useful tools for modeling time-varying variances. Volatility prediction is important mainly in financial applications, including value-at-risk (VaR) estimation and other risk practices, where policymakers or stockholders are constantly facing decision problems that usually depend on measures of volatility and risk. It is a well known fact that financial returns from market variables are characterized by nonnormality. The empirical distribution is more peaked, has heavier tails than the normal distribution, and is often skewed. These properties are crucial not only for describing the return distributions, but also for asset allocation, option pricing, forecasting, and risk management.

Discrete-time formulations of SV models were introduced by Tauchen and Pitts [62] and Taylor [63]. These models directly connect to the type of diffusion processes used in asset-pricing theory in finance [49] and capture the main empirical properties often observed in the daily series of financial returns [16] in a more appropriate way. Therefore, the discrete-time formulations of SV models have emerged as an alternative to generalized autoregressive conditional heteroscedasticity (GARCH) models of Bollerslev [12].

In literature, the basic SV model with a conditional normal distribution for stock returns has been extensively studied. From a Bayesian standpoint, several MCMC-based algorithms have been suggested to estimate the SV model. For example, Jacquier et al. [40] used the single-move Gibbs sampling within the Metropolis-Hastings algorithm to sample from the log-volatilities. Kim et al. [43], Mahieu and Schotman [48], and among others approximated the distribution of log-squared returns with a discrete mixture of several normal distributions, allowing for jointly drawing the components of the whole vector of log-volatilities. Shephard and Pitt [58] and Watanabe and Omori [66] suggested the use of random blocks, which contains some of the components of the log-volatilities, to effectively reduce the autocorrelation. However, in all of these, the normal distribution was assumed as the basis for the parameter inference.

Unfortunately, the basic SV model with a conditional normal distribution for the returns is too restrictive to model the usual leptokurtosis observed in financial return series [See 46, 22, 39, 2, among others]. To account for the lep-

*Dr. Carlos A. Abanto-Valle's research was supported by CNPq grant 481159/2013-4. Dr. Caifeng Wang's research was support by Research Institute foundation of USTB No. 39390004. Dr. Ming-Hui Chen's research was partially supported by NIH grant # GM 70335.
†Corresponding author.

tokurtosis, the SV model with Student-t is the most popular. Chib et al. [22], Jacquier et al. [39] and Abanto-Valle et al. [2] exploited the well-known fact that the Student-t distribution can be expressed as a particular scale mixture of normal distributions. Alternatively, Choy et al. [24] represented it as a scale mixture of uniform (SMU) distributions. In addition, there are other distributions that have been considered to model heavy tails in the context of SV models. For instance, Liesenfeld and Jung [46] fitted the SV model with a general error distribution (GED). Choy and Chan [23] expressed the SV with a GED distribution as a SMU distribution. Asai [9] used the SV model with contaminated normal errors and compared it with the Student-t and the GED using MCMC-based algorithms.

Besides, the empirical evidence on skewness in the distribution of financial returns is well documented in literature [37, 38, 50, 41, 20]. Corrado and Su [26] suggested that fat tails and asymmetry jointly determine the so-called "volatility smile" in option pricing using the Black-Scholes approach and the explicit account of them improves accuracy in option pricing, Chunhachinda et al. [25] showed that the introduction of skewness significantly affected the construction of the optimal portfolio, Mittnik and Paolella [50] argued that skewness and heavy tails should be taken into account in Value-at-Risk forecasts. Thus, Hansen [36] considered skewness in a GARCH model using skew-Student-t distribution errors allowing for both skewness and heavy tails to co-exist in a time-varying volatility setup. In recent years, there are more developments focusing on the skewness and heavy-tails for financial returns in the class of SV models. For example, Cappuccio et al. [14, 15] used the skew-GED and Tsiotas [65] applied the skew-Student-t distribution to model skewness and heavy tails in the conditional distribution of the returns. In their MCMC sampling algorithms, the log-volatilities are drawn using an inefficient single-move algorithm. Recently, Nakajima and Omori [52] introduced the generalized hyperbolic (GH) skew Student's t as the distribution of the returns and represented the error distribution as a normal variance-mean mixture with an inverse gamma distribution being the mixing distribution.

Value-at-Risk (VaR) has become a benchmark for measuring financial risk because it represents the market risk as one number: the maximum loss expected on an investment over a given time period at specific level of confidence. One drawback is that VaR measure is not sensitive to the shape of the loss distribution in the tails. Artzner et al. [7, 8] proposed an alternative coherent measure, called expected shortfall (ES), which gives the expected loss (magnitude) conditional on exceeding a VaR threshold.

In this paper, in order to account for skewness and heavy tails simultaneously, we extend the SV model by assuming a generalized skew-t (GST) distribution introduced by Kim et al. [42] and hence term it as the SV-GST. We develop an empirical framework for the posterior estimation of the SV model using efficient MCMC and sequential Monte Carlo (SMC) procedures. This includes a detailed sampling procedure, volatility filtering, convergence diagnostics, and model comparison. VaR and ES are estimated by simulation using the MCMC output. The data used in this paper is the SZSE-CI, which is an index tracking 40 securities traded on the Shenzhen Stock Exchange. A more detailed description about this data set is given in Section 5. Our preliminary data analysis shows that this data set exhibits certain interesting features in the stock returns, such as volatility clustering and excess kurtosis and skewness.

The rest of the paper is organized as follows. Section 2 gives a brief review of the GST distribution, including some of its properties. Section 3 presents the SV model with the GST distribution as well as the Bayesian estimation procedure using MCMC methods. We discuss some technical details about Bayesian model selection in Section 4. In Section 5, we carry out a detailed analysis of the SZSE-CI data. We conclude the paper with a brief discussion in Section 6.

2. THE GENERALIZED SKEW-T DISTRIBUTION

We first introduce some notation that will be used throughout the paper, then briefly review the generalized skew-t (GST) distribution [42], and finally discuss the related properties of this distribution.

A univariate random variable X follows a scalar GST distribution, $X \sim \mathcal{GST}(\zeta, \omega^2, \lambda, \nu_1, \nu_2)$, if it has the following stochastic representation

$$X = \zeta + U^{-1/2}\omega\delta W + U^{-1/2}\omega(1-\delta^2)^{1/2}\varepsilon, \tag{1}$$

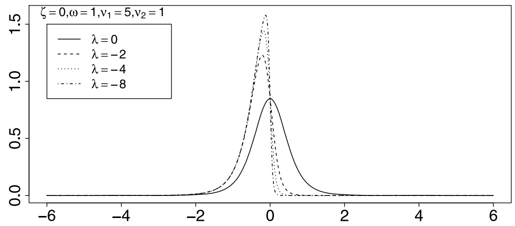
where ζ, λ and ν_1 denote the location, asymmetry, and shape parameters, respectively. The scale parameters are denoted by ω^2 and ν_2, $W \sim \mathcal{N}_{(0,\infty)}(0,1)$, $\varepsilon \sim \mathcal{N}(0,1)$, and $U \sim \mathcal{G}(\nu_1/2, \nu_2/2)$ are independent. We use $\mathcal{N}_{(a,b)}(\cdot,\cdot)$, $\mathcal{N}(\cdot,\cdot)$, and $\mathcal{G}(\cdot,\cdot)$ to denote the truncated normal in the (a,b) interval, the normal distribution, and the Gamma distribution, respectively. We use the notation $\mathcal{G}(a,b)$ to indicate a gamma distribution with mean a/b. Moreover, we write $\delta = \lambda/\sqrt{1+\lambda^2}$.

From (1), we have

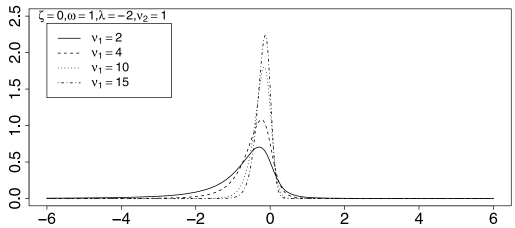
$$E(X) = \zeta + \sqrt{\frac{2}{\pi}}k_1\omega\delta, \tag{2}$$

$$V(X) = \omega^2 k_2 - \frac{2}{\pi}k_1^2\omega^2\delta^2, \tag{3}$$

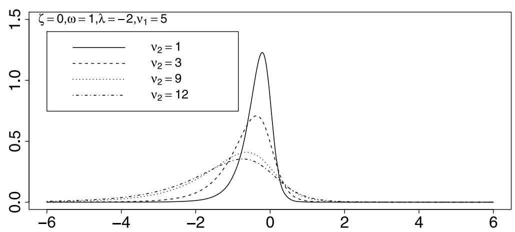
where $k_m = E(U^{-m/2})$ for $m = 1, 2$, and $E(\cdot)$ and $V(\cdot)$ denote the expected value and variance, respectively. Special cases of the GST include the skew-Student-t ($\nu_1 = \nu_2$), the generalized Student-t ($\lambda = 0$) and the traditional symmetric Student-t ($\lambda = 0, \nu_1 = \nu_2$). Moreover, the GST can capture left-tailed or negative skewness when $\lambda < 0$, and right-tailed or positive skewness when $\lambda > 0$.



(a) $\zeta = 0, \omega = 1, \nu_1 = 5, \nu_2 = 1$ (fixed), $\lambda = 0, -2, -4, -8$.



(b) $\zeta = 0, \omega = 1, \lambda = -2, \nu_2 = 1$ (fixed), $\nu_1 = 2, 4, 10$ and 15.



(c) $\zeta = 0, \omega = 1, \lambda = -2, \nu_1 = 5$ (fixed), $\nu_2 = 1, 3, 9, 12$ and 15.

Figure 1. The GST distribution.

In order to interpret the parameters (λ, ν_1, ν_2) in relation to skewness and heavyness of tails, some GST densities are plotted in Figure 1, considering several combinations of the parameter values λ, ν_1 and ν_2, with ζ and ω fixed at 0 and 1, respectively. In Figure 1 (left panel), the densities are drawn using $\lambda = 0, -2, -4, -8$ with ν_1 and ν_2 fixed at 5 and 1. As mentioned earlier, $\lambda = 0$ corresponds to a symmetric generalized Student-t density. We can see that a more negative value of λ implies a more negative skewness as well as heavier tails. Figure 1 (center panel) shows the densities for ν_1 at 2, 4, 10 and 15 with λ and ν_2 fixed at -2 and 1. We can see that as ν_1 becomes larger, the density becomes less skewed, and has lighter tails. Figure 1 (right panel) shows the densities for ν_2 at 1, 3, 9 and 12 with λ and ν_1 fixed at -2 and 5. We can see that as ν_2 becomes larger, the density becomes more skewed and has heavy tails. Hence, the skewness and heavyness of the tails of the distribution are jointly determined by the combination of the values of parameters λ, ν_1 and ν_2.

3. THE SV-GST MODEL

3.1 The model

In order to account for both the excess kurtosis and skewness in stock returns, we introduce the stochastic volatility model with generalized skew-t errors (SV-GST), which is defined as

$$\text{(4a)} \qquad y_t = \mathrm{e}^{\frac{h_t}{2}} \xi_t,$$

$$\text{(4b)} \qquad h_{t+1} = \mu + \varphi(h_t - \mu) + \sigma_\eta \eta_t,$$

where y_t and h_t are, respectively, the compounded return and the log-volatility at time t, μ, φ and σ_η^2 denote the drift, the persistence, and the variance of the volatilities process, respectively. We assume that $|\varphi| < 1$, i.e., the log-volatility process is stationary with the initial value $h_1 \sim \mathcal{N}(\mu, \frac{\sigma_\eta^2}{1-\varphi^2})$, moreover, $\xi_t \sim \mathcal{GST}(\zeta, \omega^2, \lambda, \nu_1, \nu_2)$ and $\eta_t \sim \mathcal{N}(0, 1)$ are independent.

To ensure model identifiability, we set $\nu_1 = \nu$ and $\nu_2 = 1$. The parameters, ζ and ω, are restricted in such a way that $E(\xi_t) = 0$ and $V(\xi_t) = 1$, because they imply the hypothesis of martingale of the return series. Thus, we have $\zeta = -\sqrt{\frac{2}{\pi}} k_1 \delta \omega$ and $\omega^2 = [k_2 - \frac{2}{\pi} k_1^2 \delta^2]^{-1}$, where $k_1 = \sqrt{\frac{1}{2}} \frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}$, $k_2 = \frac{1}{\nu - 2}$ and $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$.

The SV-GST defined by (4a) and (4b) can be written hierarchically using the stochastic representation of the GST distribution in (1) as follows:

$$\text{(5a)} \qquad y_t = (\zeta + \omega \delta W_t U_t^{-\frac{1}{2}}) \mathrm{e}^{\frac{h_t}{2}} + \mathrm{e}^{\frac{h_t}{2}} U_t^{-\frac{1}{2}} \omega (1 - \delta^2)^{\frac{1}{2}} \varepsilon_t,$$

$$\text{(5b)} \qquad h_{t+1} = \mu + \varphi(h_t - \mu) + \sigma_\eta \eta_t,$$

$$\text{(5c)} \qquad W_t \sim \mathcal{N}_{(0,\infty)}(0, 1),$$

$$\text{(5d)} \qquad U_t | \nu_1, \nu_2 \sim \mathcal{G}(\frac{\nu_1}{2}, \frac{\nu_2}{2}),$$

where ε_t and η_t are mutually independent and normally distributed with zero mean and unit variance. In this setup, Equations (5a) and (5b) with $\lambda = 0$ (equivalently $\delta = 0$) and $U_t = 1, \forall t = 1, \ldots, T$, define the SV model with a normal distribution (SV-N). Equations (5a), (5b) and (5d) with $\lambda = 0$ define the the SV model with generalized Student-t distribution (SV-GT). Equations (5a), (5b) and (5d) with $\lambda = 0$ and $\nu_1 = \nu_2$ define the SV model with a Student-t distribution (SV-T). Finally, Equations (5a), (5b) and (5c) with $U_t = 1, \forall t = 1, \ldots, T$ yield the SV model with a skew normal distribution (SV-SN).

3.2 Parameter estimation via MCMC

Denote $\boldsymbol{\theta} = (\mu, \varphi, \sigma_\eta^2, \nu, \lambda)'$ as the full parameter vector of the SV-GST model, where ν is the degrees of freedom parameter vector associated with the mixture distribution and λ is the skewness parameter. Let $\mathbf{h}_{1:T} = (h_1, \ldots, h_T)'$ be the vector of the log volatilities, $\mathbf{U}_{1:T} = (U_1, \ldots, U_T)'$ and $\mathbf{W}_{1:T} = (W_1, \ldots, W_T)'$ be the mixing variables, and $\mathbf{y}_{1:T} = (y_1, \ldots, y_T)'$ be the information available up to time T. The Bayesian approach to estimate the parameters in the SV-GST model uses the data augmentation principle, in which $\mathbf{h}_{1:T}$, $\mathbf{W}_{1:T}$ and $\mathbf{U}_{1:T}$ are considered as latent variables. The joint posterior density of the parameters and latent unobservable variables can be written as

$$\begin{aligned} & p(\boldsymbol{\theta}, \mathbf{W}_{1:T}, \mathbf{U}_{1:T}, \mathbf{h}_{1:T} \mid \mathbf{y}_{1:T}) \\ \propto\ & p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta}, \mathbf{W}_{1:T}, \mathbf{U}_{1:T}, \mathbf{h}_{1:T}) \\ \times\ & p(\mathbf{h}_{1:T} \mid \boldsymbol{\theta}) p(\mathbf{W}_{1:T}) p(\mathbf{U}_{1:T} \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}), \end{aligned} \tag{6}$$

where $p(\boldsymbol{\theta})$ is the prior distribution. Since the posterior density $p(\boldsymbol{\theta}, \mathbf{W}_{1:T}, \mathbf{U}_{1:T}, \mathbf{h}_{1:T} \mid \mathbf{y}_{1:T})$ is analytically intractable, we first sample the parameters $\boldsymbol{\theta}$ and then draw the latent variables $\mathbf{W}_{1:T}$, $\mathbf{U}_{1:T}$ and $\mathbf{h}_{1:T}$ from the posterior distribution using the Gibbs sampling algorithm. The sampling scheme is described in Algorithm 3.1. Sampling the log-volatilities $\mathbf{h}_{1:T}$ in Step 5 of Algorithm 3.1 is the most difficult task due to the nonlinear setup in the observational equation (5a). In order to avoid the high correlations due to the Markovian structure of the h_t's, in the next subsection we develop a multi-move block sampler to sample $\mathbf{h}_{1:T}$ by blocks (Shephard and Pitt 58, Watanabe and Omori 66, Abanto-Valle et al. 2). Details on the full conditionals of $\boldsymbol{\theta}$ and the latent variables $\mathbf{U}_{1:T}$ and $\mathbf{W}_{1:T}$ are given in Appendix A.

Algorithm 3.1.

- *Step 1.* Set $i = 0$ and get starting values for the parameters $\boldsymbol{\theta}^{(i)}$ and the latent quantities $\mathbf{W}_{1:T}^{(i)}$, $\mathbf{U}_{1:T}^{(i)}$ and $\mathbf{h}_{1:T}^{(i)}$.
- *Step 2.* Generate $\boldsymbol{\theta}^{(i+1)}$ in turn from its full conditional distribution, given $\mathbf{y}_{1:T}$, $\mathbf{h}_{1:T}^{(i)}, \mathbf{W}_{1:T}^{(i)}$ and $\mathbf{U}_{1:T}^{(i)}$.
- *Step 3.* Draw $\mathbf{W}_{1:T}^{(i+1)} \sim p(\mathbf{W}_{1:T} \mid \boldsymbol{\theta}^{(i)}, \mathbf{U}_{1:T}^{(i)}, \mathbf{h}_{1:T}^{(i)}, \mathbf{y}_{1:T})$.
- *Step 4.* Draw $\mathbf{U}_{1:T}^{(i+1)} \sim p(\mathbf{U}_{1:T} \mid \boldsymbol{\theta}^{(i+1)}, \mathbf{W}_{1:T}^{(i+1)}, \mathbf{h}_{1:T}^{(i)}, \mathbf{y}_{1:T})$.
- *Step 5.* Generate $\mathbf{h}_{1:T}$ by blocks as:
  - i) For $l = 1, \ldots, K$, the knot positions are generated as k_l, the floor of $[T \times \{(l + u_l)/(K + 2)\}]$, where the $u_l's$ are independent realizations of the uniform random variable on the interval (0,1).
  - ii) For $l = 1, \ldots, K$, generate the block $h_{k_{l-1}+1:k_l-1}$ jointly conditional on $\mathbf{y}_{k_{l-1}:k_l-1}$, $\boldsymbol{\theta}^{(i+1)}$, $\mathbf{W}_{k_{l-1}+1:k_l-1}^{(i+1)}$, $\mathbf{U}_{k_{l-1}+1:k_l-1}^{(i+1)}$, $h_{k_{l-1}}^{(i)}$ and $h_{k_l}^{(i)}$.
  - iii) For $l = 1, \ldots, K$, draw $h_{k_l}^{(i+1)}$ conditional on $\mathbf{y}_{1:T}$, $\boldsymbol{\theta}^{(i)}$, $W_{k_l}^{(i+1)}$, $U_{k_l}^{(i+1)}$, $h_{k_l-1}^{(i+1)}$ and $h_{k_l+1}^{(i+1)}$.

6. Set $i = i + 1$ and return to *Step 2* until convergence is achieved.

The prior distributions of the parameters are specified as follows: $\mu \sim \mathcal{N}(\bar{\mu}, \sigma_\mu^2)$, $\varphi \sim \mathcal{N}_{(-1,1)}(\bar{\varphi}, \sigma_\varphi^2)$, and $\sigma_\eta^2 \sim \mathcal{IG}(\frac{T_0}{2}, \frac{M_0}{2})$, where $\mathcal{IG}(a, b)$ denotes an inverse gamma distribution with mean $b/(a-1)$. For ν, we assume a prior based on Fonseca et al. [31], which has the form

$$p(\nu) \propto \left(\frac{\nu}{\nu+3}\right)^{\frac{1}{2}} \left\{\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right\}^{\frac{1}{2}}, \tag{7}$$

where $\psi(a) = \frac{d\{\log \Gamma(a)\}}{da}$ and $\psi'(a) = \frac{d\{\psi(a)\}}{da}$ are the digamma and trigamma functions, respectively. To the skewness parameter, we assume that $\lambda \sim t_{0.5}(0, \frac{\pi^2}{4})$, a Jeffreys' prior suggested by Bayes and Branco [10], where $t_c(a, b)$ denotes the Student-t distribution with location a, scale b, and c degrees of freedom.

3.3 Forecasting returns, volatility, Value-at-Risk and Expected Shortfall

The K-step ahead prediction densities can be calculated using the composition method via the following recursive procedure:

$$\begin{aligned} & p(y_{T+K} \mid \mathbf{y}_{1:T}) \\ = & \int \Big[p(y_{T+K} \mid U_{T+K}, W_{T+K}, h_{T+K}) p(W_{T+K} \mid \boldsymbol{\theta}) \\ & \times p(U_{T+K} \mid \boldsymbol{\theta}) p(h_{T+K} \mid \boldsymbol{\theta}, \mathbf{y}_{1:T}) \\ & \times p(\boldsymbol{\theta} \mid \mathbf{y}_{1:T}) \Big] dh_{T+K} dW_{T+K} dU_{T+K} d\boldsymbol{\theta}, \\ & p(h_{T+K} \mid \boldsymbol{\theta}, \mathbf{y}_{1:T}) \\ = & \int p(h_{T+K} \mid \boldsymbol{\theta}, h_{T+K-1}) p(h_{T+K-1} \mid \boldsymbol{\theta}, \mathbf{y}_{1:T}) dh_{T+K-1}. \end{aligned}$$

Numerical evaluation of the last integral is straightforward. To initialize the recursion, we use $h_T^{(i)}$ and $\boldsymbol{\theta}^{(i)}$, for $i = 1, \ldots, N$, from the MCMC output. Given these N draws, we sample $h_{T+k}^{(i)}$ from $p(h_{T+k} \mid \boldsymbol{\theta}^{(i)}, h_{T+k-1}^{(i)})$, $W_{T+k}^{(i)}$ from $p(W_{T+k} \mid \boldsymbol{\theta}^{(i)})$, and $U_{T+k}^{(i)}$ from $p(U_{T+k} \mid \boldsymbol{\theta}^{(i)})$, for $i = 1, \ldots, N$ and $k = 1, \ldots, K$, by using (5b), (5c) and (5d), respectively. Finally, using (5a), we sample $y_{T+k}^{(i)}$ from $p(y_{T+k} \mid \boldsymbol{\theta}^{(i)}, W_{T+k}^{(i)}, U_{T+k}^{(i)}, h_{T+k}^{(i)})$, for $i = 1, \ldots, N$ and $k = 1, \ldots, K$.

To evaluate the performance of the model on VaR prediction, the likelihood ratio test introduced in Kupiec [45] is used to to test that the null hypothesis that the expected proportion of the number of "beyond VaR" or "violation" during the test periods is equal to α. The violation is formulated by $I_t(\alpha) = I[y_{T+1} < \widehat{\mathrm{VaR}}_t(\alpha)]$ for the left tail and $I_t(\alpha) = I[y_t > \widehat{\mathrm{VaR}}_t(\alpha)]$ for the right tail, where $I[.]$ is an indicator function and $\widehat{\mathrm{VaR}}_t(\alpha)$ is the estimated VaR at level α, which can be obtained by simulation using the k-step ahead densities described below [See 18, 30, for a detailed review]. Let x_α be the number of violations, that is, $x_\alpha = \sum_{t=T+1}^{T+m} I_t(\alpha)$ and $\hat{\alpha} = x_\alpha/m$. The unconditional test of Kupiec [45] is a likelihood ratio test with the χ_1^2-distributed test statistic defined as

$$LRuc = 2\{\log[\hat{\alpha}^{x_\alpha}(1-\hat{\alpha})^{m-x_\alpha}] - \log[\alpha^{x_\alpha}(1-\alpha)^{m-x_\alpha}]\}.$$

The Expected Shortfall (ES) is formally defined via $ES_t(\alpha) = E[y_t|y_t < VaR_t(\alpha)]$ for the left tail and $ES_t(\alpha) = E[y_t|y_t > VaR_t(\alpha)]$ for the right tail. Following Aas and Haff [1] and Nakajima [51], we compute the measure developed by Embrechts et al. [29] for evaluating the performance of the predicted ES, denoted by $\widehat{ES_t(\alpha)}$. We define $\delta_t(\alpha) = y_t - \widehat{ES_t(\alpha)}$ as an excess of return. Let δ_α be the αth quantile $\{\delta_t(\alpha)\}_{t=T+1}^{T+m}$. Next, define $S_t(\alpha) = I[\delta_t(\alpha) < \delta_\alpha]$ for the left tail and $S_t(\alpha) = I[\delta_t(\alpha) > \delta_\alpha]$ for the right tail. Write $s_\alpha = \sum_{t=T+1}^{T+m} S_t(\alpha)$. The measure of Embrechts et al. [29] is given by $D(\alpha) = \frac{1}{2}(|D_1(\alpha)| + |D_2(\alpha)|$, where

$$D_1(\alpha) = \frac{1}{x_\alpha} \sum_{I_t(\alpha)=1} \delta_t(\alpha), \tag{8}$$

$$D_2(\alpha) = \frac{1}{s_\alpha} \sum_{S_t(\alpha)=1} \delta_t(\alpha). \tag{9}$$

As discussed in Aas and Haff [1] and Nakajima [51], $D_1(\alpha)$ is the standard back-testing measure for expected shortfall estimates. Its weakness is that it strongly depends on the VaR estimates without adequately reflecting the correctness of these values; $D_2(\alpha)$ is computed to correct this because $D_2(\alpha)$ measures an average difference between the return and the estimated ES for the α-level tail of that difference from all test periods. A smaller $D(\alpha)$ implies more precise prediction of ES.

4. BAYESIAN MODEL COMPARISON

In this section, we describe three Bayesian model selection criteria: the deviance information criterion [60, 11, 17], the Bayesian predictive information criterion (BPIC) [5, 6], and the log predictive score (LPS) [35, 34, 28]. The first one is directly obtained from the MCMC output. The others are obtained by using the predictive distribution at each time. The predictive distribution is evaluated numerically by using the auxiliary particle filtering method of Pitt and Shephard [54] described in Appendix C.

4.1 The deviance information criterion

Spiegelhalter et al. [60] introduced the deviance information criterion (DIC) defined as

$$\text{DIC} = -2E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})] + p_D, \tag{10}$$

where $E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}$ denotes the expectation taken with respect to the posterior distribution of $\boldsymbol{\theta}$ given the data $\mathbf{y}_{1:T}$. The second term p_D in (10) is the effective number of parameters, which measures the complexity of the model. Specifically, p_D is defined as twice the difference between the deviance evaluated at the posterior mean of the parameters and the posterior mean of the deviance:

$$p_D = 2[\log p(\mathbf{y}_{1:T} \mid \bar{\boldsymbol{\theta}}) - E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})]]. \tag{11}$$

As pointed out by Stone [61], Robert and Titterington [55], Celeux et al. [17] and Ando [6], the DIC suffers from some theoretical drawbacks. First, in the derivation of DIC, Spiegelhalter et al. [60, p. 604] assumed the specified parametric family of probability distributions that generate future observations encompassing the true model. This assumption may not always hold. Secondly, the observed data are used both to construct the posterior distribution and to compute the posterior mean of the expected log likelihood. Thus, the bias in the estimate of DIC tends to considerably underestimate the true bias. To overcome these theoretical problems in DIC, Ando [6] proposed the Bayesian predictive information criterion (BPIC) as an improved alternative of DIC.

4.2 The Bayesian predictive information criterion

Ando [5, 6] introduced BPIC, which is defined as

$$BPIC = -2E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log\{p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})\}] + 2T\hat{b}, \tag{12}$$

where $\hat{b}$ is given by

$$\begin{aligned} \hat{b} \approx & \frac{1}{T}\Big\{E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[\log\{p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta})p(\boldsymbol{\theta})\}] \\ & - \log[p(\mathbf{y}_{1:T} \mid \hat{\boldsymbol{\theta}})p(\hat{\boldsymbol{\theta}})] + \text{tr}\{J_T^{-1}(\hat{\boldsymbol{\theta}})I_T(\hat{\boldsymbol{\theta}})\} + 0.5q\Big\}, \end{aligned} \tag{13}$$

q is the dimension of $\boldsymbol{\theta}$, $E_{\boldsymbol{\theta}|\mathbf{y}_{1:T}}[.]$ denotes the expectation with respect to the posterior distribution, $\hat{\boldsymbol{\theta}}$ is the posterior mode, and

$$I_T(\hat{\boldsymbol{\theta}}) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial \ell_T(y_t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\frac{\partial \ell_T(y_t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right)\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

$$J_T(\hat{\boldsymbol{\theta}}) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\partial^2 \ell_T(y_t,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}'}\right)\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}$$

with $\ell_T(y_t,\boldsymbol{\theta}) = \log p(y_t \mid \mathbf{y}_{1:t-1},\boldsymbol{\theta}) + \log p(\boldsymbol{\theta})/T$.

4.3 The log predictive score criterion

Scoring rules provide summary measures for the evaluation of probabilistic forecast by assigning a numerical score based on the predictive distribution and on the event or value that materializes. The fit of the models studied here will be assessed using log predictive scores [35, 34, 28]. The average log predictive score for the one-step ahead prediction is given by

$$\text{LPS} = -\frac{1}{T}\sum_{t=1}^{T}\log p(Y_t \mid \mathbf{Y}_{1:t-1},\hat{\boldsymbol{\theta}}), \tag{14}$$

where $\mathbf{Y}_{1:t-1} = (Y_1,\ldots,Y_{t-1})'$, $\hat{\boldsymbol{\theta}}$ is an estimate of the model parameters and $p(Y_t \mid \mathbf{Y}_{1:t-1},\hat{\boldsymbol{\theta}})$ is the one-step ahead predictive density. The smaller the LPS value, the better the model fits the data. In the application, we use $\hat{\boldsymbol{\theta}}$ as being the posterior mean obtained from the MCMC output. Although an analytical evaluation of $p(Y_t \mid \mathbf{Y}_{1:t-1}, \hat{\boldsymbol{\theta}})$ is not possible, this predictive density can be evaluated numerically by using the auxiliary particle learning (APF) method [54], which is described in Appendix C.

5. ANALYSIS OF THE SZSE-CI DATA

5.1 The data

In this section, we carry out a detailed analysis of the daily closing prices of the Shenzhen Stock Exchange Component Index (SZSE-CI). The SZSE-CI is a capitalization weighted index, which is composed with the 40 top companies that issue A-shares on SZSE. The base is 1,000 and the base day is July 20, 1994. The SZSE regularly inspects the performance of the component shares, timely replaces that of lower performance. The replacement will not be too frequent and it is usually in January, March and September each year.

The period for the SZSE-CI we consider is from February 16, 2005 to December 12, 2012, which yields 1,956 observations. The data set was obtained from the Yahoo finance website, available to download at http://finance.yahoo.com. Throughout, we work with the mean corrected returns computed as

$$y_t = 100 \times \{(\log P_t - \log P_{t-1}) - \frac{1}{T}\sum_{j=1}^{T}(\log P_j - \log P_{j-1})\},$$

where P_t is the closing price on day t.

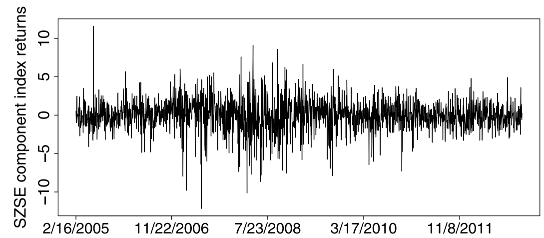
Table 1 summarizes descriptive statistics for the corrected compounded returns with the time series plot in Figure 2. For the returns series, the sample mean, standard deviation, skewness and kurtosis were 0.00, 1.98, -0.32 and 5.38, respectively. Note that the kurtosis of the returns is > 3 and the skewness is slightly below zero. These evidences imply that the daily SZSE-CI returns exhibit a departure from the underlying normality assumption. Thus, we reanalyze this data set with the aim of providing a robust inference by using the generalized skew-t distribution (GST). In our analysis, we fit the SV-N, SV-T, SV-GT, SV-SN, SV-ST and SV-GST models to the SZSE-CI data. All the calculations were carried out by running the stand alone codes developed by the authors using an open source C++ library for statistical computation along with the Scythe statistical library [53], which is available for free download at http://scythe.wustl.edu.

5.2 Parameter estimates

In all posterior computations, we simulate the h_t's in a multi-move fashion with stochastic knots based on the method described by [58, 9, 2, 3, 4]. We fix the number of blocks K to be 30 in such a way that each block contains 62 h_t's on average. We set the prior distributions of the common parameters as: $\mu \sim \mathcal{N}(0, 100)$, $\varphi \sim \mathcal{N}_{(-1,1)}(0.95, 100)$ and $\sigma_\eta^2 \sim \mathcal{IG}(2.5, 0.025)$. For φ, its prior mean and variance are 0.0032 and 0.3328. This prior setup is equivalent to the uniform distribution on interval $(-1, 1)$, which gives zero mean and variance of 0.3333. We assume that $\lambda \sim t_{0.5}(0, \frac{\pi^2}{4})$, a Jeffreys' prior suggested by Bayes and Branco [10]. Finally, for ν, we assume the prior suggested by Fonseca et al. [31].

Table 1. Summary statistics for SZSE-CI mean corrected returns

	mean	s.d.	max	min	skewness	kurtosis
Returns	0.00	1.98	9.16	-9.75	-0.32	5.38



(a) The plot of the mean corrected series

(b) The histogram of returns

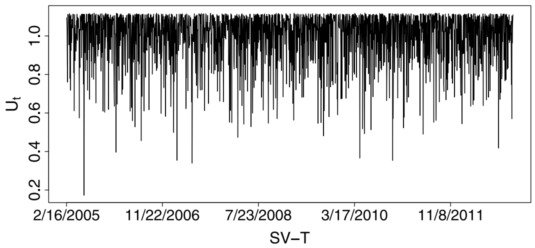
Figure 2. SZSE-CI returns with sample period from February, 16, 2005 to December 12, 2012.

For all of the models we considered, we generated 70,000 MCMC iterations. In all cases, the first 20,000 draws were discarded as a "burn-in" period. In order to reduce the autocorrelations between successive values of the simulated chain, only every 20th values of the chain were stored. With the resulting 2,500 values, we calculated the posterior means, the 95% highest posterior density (HPD) intervals [19, p. 219], and the convergence diagnostic (CD) statistics [33]. If the sequence of the recorded MCMC output is stationary, it converges in distribution to the standard normal. According to the CD, the null hypothesis that the sequence of 2500 draws is stationary is accepted at the 5% level, i.e., CD $\in (-1.96, 1.96)$, for all the parameters in all the models considered here. Table 2 summarizes the results.

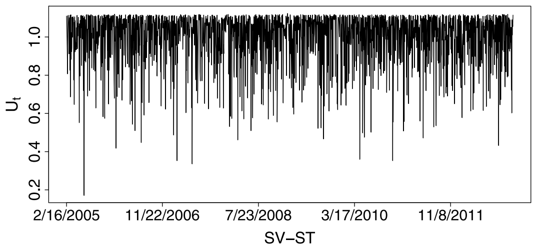
It is easy to see from Table 2 that the posterior means of φ and 95% HPD intervals are very close to the unity, which is consistent with the existing evidence of great persistence in the log-volatility process. Additionally, the pos-

Table 2. SZSE-CI returns: Estimation results for the SV-N, SV-T, SV-GT, SV-SN, SV-ST and SV-GST models. First row: Posterior mean. Second row: 95% HPD interval. Third row: CD statistics. Fourth row: Inefficiency factors

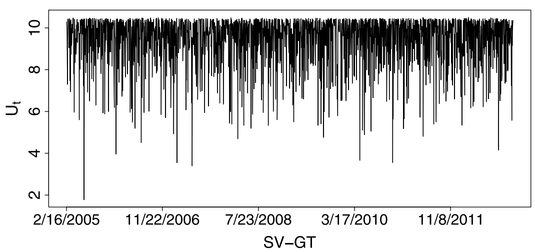
Parameter	Model					
	SV-N	SV-T	SV-GT	SV-SN	SV-ST	SV-GST
	1.1163	1.1218	1.1275	1.1051	1.1188	1.1028
μ	(0.7131,1.5011)	(0.5361,1.6635)	(0.5456,1.6864)	(0.7112,1.5342)	(0.5781,1.6230)	(0.5386,1.5908)
	−1.31	−0.51	−0.30	0.26	−0.52	0.29
	1.00	1.00	1.00	1.00	1.00	1.06
	0.9820	0.9891	0.9890	0.9832	0.9889	0.9887
φ	(0.9689,0.9939)	(0.9801,0.9984)	(0.9787,0.9978)	(0.9710,0.9942)	(0.9794,0.9979)	(0.9793,0.9979)
	−0.04	−0.27	0.94	−0.86	−0.53	−0.66
	3.12	3.48	2.99	3.52	3.48	3.35
	0.0174	0.0101	0.0100	0.0158	0.0097	0.0101
σ_η^2	(0.0092,0.0264)	(0.0055,0.01536)	(0.0054,0.0153)	(0.0085,0.02390)	(0.0049,0.0155)	(0.0054,0.0157)
	−0.12	0.08	−1.31	−0.35	1.15	0.54
	6.11	6.18	6.97	6.44	6.79	6.78
	–	9.1516	9.5762	–	9.3616	8.8054
ν	–	(5.9118,12.6834)	(6.1546,13.3752)	–	(6.1809,13.2712)	(6.3472,12.1114)
	–	−1.68	0.88	–	1.64	0.72
	–	5.48	50.06	–	7.06	30.52
	–	–	–	−0.9271	−0.4612	−0.4655
λ	–	–	–	(−1.4190,−0.0901)	(−0.8516,−0.0365)	(−0.8624,−0.0468)
	–	–	–	−0.74	1.79	−1.61
	–	–	–	18.83	3.97	3.88



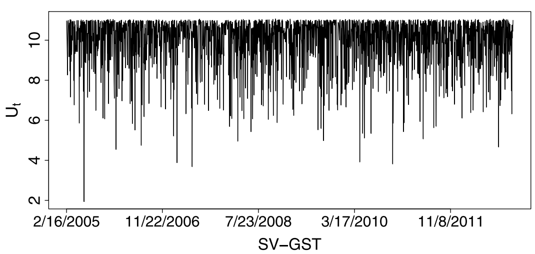
(a) For SV-T models



(b) For SV-ST models

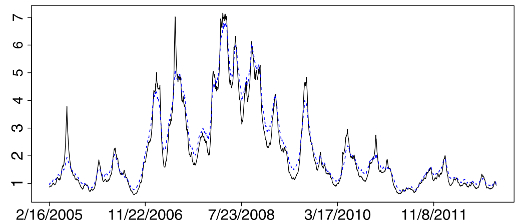


(c) For SV-GT models

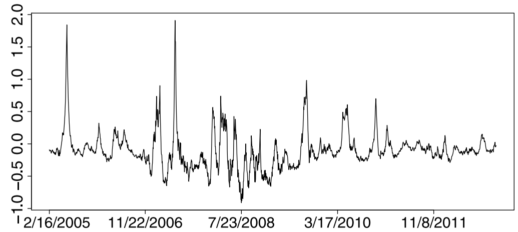


(d) For SV-GST models

Figure 3. SZSE-CI data set: posterior smoothed mean of mixture variable U_t.



(a) Posterior smoothed mean of e^{h_t}



(b) Posterior smoothed mean of the difference of e^{h_t} in both models

Figure 4. SZSE-CI data set. SV-N (solid line) and SV-GST (dotted line).

terior means of φ under the SV-N and the SV-SN models are slightly smaller than those under the other four models. As expected, the posterior means of σ_η^2 under the SV-N and SV-SN models are higher than those under the SV-T, SV-GT, SV-ST and the SV-GST models, indicating that the log-volatility process of the last four models is less variable than that of the SV-N and SV-SN models. Under the SV-T, SV-GT, SV-ST and SV-GST models, the magnitude of the tail-heaviness is measured by the ν parameter. Moreover, under these four models the posterior means of ν are 9.1516, 9.5762, 9.3616 and 8.8054, respectively. Regarding the skewness parameter, the posterior means of λ, in the SV-SN, SV-ST and SV-GST models are -0.9271, -0.4612 and -0.4665, respectively. For all the models considered here, λ is significantly below 0, since all the three 95% HPD intervals do not contain zero. For the SV-ST and SV-GST, these results support the necessity to model asymmetry and heavy tails simultaneously. The magnitudes of the mixing parameter U_t are associated with extremeness of the corresponding observations. In the Bayesian paradigm, the posterior mean of the mixing parameter can be used to identify a possible outlier (see, for instance, Rosa et al. [56]). The SV-T, SV-GT, SV-ST and SV-GST models can accommodate an outlier by inflating the variance component for that observation in the conditional distribution with smaller U_t value. This fact is shown in Figure 3 where we plot the posterior mean of the mixing variable U_t for the SV-T (left top panel), SV-ST (right top panel), SV-GT (left bottom panel) and SV-GST (right bottom panel) models, respectively. In Figure 4, we draw the smoothed mean of e^{h_t} obtained from the MCMC output for the SV-N model (solid line) and the SV-GST model (dotted line). From a practical point of view, we are mainly interested in whether there is a significant difference between the two series. Therefore, in the bottom panel of Figure 4, we show the smoothed mean of the difference of e^{h_t} obtained from the SV-N and SV-GST models. Some extreme returns make the differences more clear. This can have a substantial impact, for instance, in the evaluation of derivative instruments and several strategic or tactical asset allocation topics.

In order to evaluate the sensitivity to the choice of the priors, we use a set of 10 priors for φ ans σ_η^2. In Table 3, we report these results. Essentially the posterior estimates are almost the same, implying that the results are not sensitive to the priors we choose.

5.3 Model comparison and diagnostics

The main purpose of this section is to compare the SV-GST with the other competing models. To assess the

Table 3. SZSE-CI returns: Prior Sensitivity

	Priors			Posteriors				
	μ	ϕ	σ_η^2	μ	ϕ	σ_η^2	λ	ν
1	$\mathcal{N}(0,100)$	$\mathcal{N}_{(-1,1)}(0.95,100)$	$\mathcal{IG}(5.0,0.025)$	1.1052 (0.4768,1.70422)	0.9909 (0.9820,0.9985)	0.0079 (0.0044, 0.0132)	−0.4545 (−0.8427,−0.0384)	8.8054 (6.3472,12.1114)
2	$\mathcal{N}(0,100)$	$\mathcal{N}_{(-1,1)}(0.90,100)$	$\mathcal{IG}(2.5,0.025)$	1.1176 (0.3678,1.7652)	0.9917 (0.9819,0.9987)	0.0072 (0.0036,0.0130)	−0.4462 (−0.8494,−0.0542)	8.9864 (6.3599,13.9385)
3	$\mathcal{N}(0,100)$	$\mathcal{N}_{(-1,1)}(0.91,100)$	$\mathcal{IG}(6.0,0.025)$	1.1127 (0.4368,1.7604)	0.9908 (0.9805, 0.9985)	0.0082 (0.0042,0.0148)	−0.4526 (−0.8246,−0.0539)	8.9103 (6.3909,12.9119)
4	$\mathcal{N}(-1,100)$	$\mathcal{N}_{(-1,1)}(0.92,100)$	$\mathcal{IG}(4.0,0.025)$	1.1035 (0.2301,1.7498)	0.9929 (0.9849,0.9991)	0.0067 (0.0031,0.0109)	−0.4236 (−0.7944,−0.0192)	8.3768 (6.00989,12.5564)
5	$\mathcal{N}(-1,100)$	$\mathcal{N}_{(-1,1)}(0.93,100)$	$\mathcal{IG}(1.5,0.025)$	1.1138 (0.4858,1.6592)	0.9900 (0.9792, 0.9981)	0.0080 (0.0044.0.0163)	−0.4644 (−0.8397,−0.0426)	9.0574 (6.15416,13.6562)
6	$\mathcal{N}(-1,100)$	$\mathcal{N}_{(-1,1)}(0.94,100)$	$\mathcal{IG}(4.5,0.025)$	1.0953 (0.0624,1.8416)	0.9930 (0.9843,0.9992)	0.0060 (0.0030,0.0113)	−0.4389 (−0.8268, −0.0265)	8.7106 (6.1684,13.3611)
7	$\mathcal{N}(-1,100)$	$\mathcal{N}_{(-1,1)}(0.95,100)$	$\mathcal{IG}(5.0,0.025)$	1.0965 (0.1521,1.8172)	0.9933 (0.9853,0.9992)	0.0058 (0.0027,0.0099)	−0.4456 (−0.8065, −0.0602)	8.6432 (5.8861,12.4514)
8	$\mathcal{N}(-1,100)$	$\mathcal{N}_{(-1,1)}(0.98,100)$	$\mathcal{IG}(0.02,0.02)$	1.1167 (0.4904,1.7144)	0.9900 (0.9791,0.9983)	0.0089 (0.0044,0.0156)	−0.4592 (−0.8481, −0.0272)	9.0251 (6.0687,15.1378)
9	$\mathcal{N}(-1,100)$	$\mathcal{N}_{(-1,1)}(0.97,100)$	$\mathcal{IG}(0.04,0.04)$	1.1074 (0.5552,1.5998)	0.9894 (0.9752,0.9968)	0.0112 (0.0058,0.01952)	−0.4727 (−0.8865, −0.0520)	9.3807 (6.7123.15.1211)
10	$\mathcal{N}(-1,100)$	$\mathcal{N}_{(-1,1)}(0.98,100)$	$\mathcal{IG}(0.08,0.08)$	1.1168 (0.6621,1.5637)	0.9843 (0.9709, 0.9953)	0.0108 (0.0064,0.0222)	−0.4681 (−0.8659,−0.0039)	10.2256 (7.0907,15.7129)

Table 4. SZSE return data set. DIC: deviance information criterion, BPIC: Bayesian predictive information criterion, and LPS: log predictive score

	DIC		BPIC		LPS	
Model	Value	Ranking	Value	Ranking	Value	Ranking
SV-N	7,568.1	6	11,190.0	6	2.9557	5
SV-T	7,504.8	5	7,605.6	4	2.0388	4
SV-GT	7,492.2	3	7,599.8	3	2.0384	3
SV-SN	7,495.3	4	11,000.6	5	3.0018	6
SV-ST	7,419.2	2	7,598.9	2	2.0370	1
SV-GST	7,398.2	1	7,595.5	1	2.0383	2

Table 5. SZSE returns. P-values of the diagnostics test using the standardized innovations ς_t. The BDS test developed by Brock et al. [13] is used to test for the null hypothesis of independent and identical distribution (iid)

	Box-Ljung Test	Jarque-Bera Test	BDS Test
	(p-value)	(p-value)	(p-value)
SV-N	0.2351	0.0000	0.0000
SV-T	0.1172	0.0186	0.1640
SV-GT	0.1200	0.0305	0.1267
SV-SN	0.2060	0.0000	0.0003
SV-ST	0.0984	0.4271	0.4471
SV-GST	0.0952	0.6317	0.4894

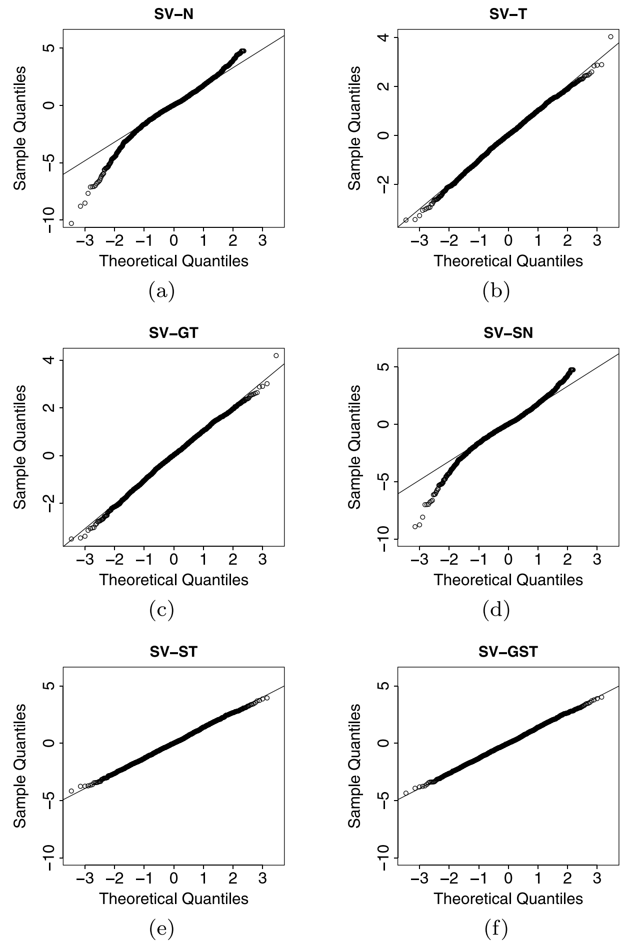


Figure 5. SZSE-CI data set: quantile-quantile plot of the residuals ς_t. The solid line plots the quantiles of the $\mathcal{N}(0,1)$ against the quantiles of the standard normal and the points were the sorted values of ς_t against the quantiles of the standard normal.

goodness-of-fit of the estimated models, we calculate the values for the corresponding DIC, BPIC and LPS. For the SV-N, SV-T, SV-GT, SV-SN, SV-ST and SV-GST models, the log-likelihood function, $\log p(\mathbf{y}_{1:T}|\boldsymbol{\theta})$, was estimated by using the APF with 10,000 particles. From Table 4, the DIC and BPIC values indicate the SV-GST model is the best model among all the models considered here, suggesting that the SZSE-CI return data demonstrate a sufficient departure from underlying normality assumptions and symmetry.

Finally, Table 5 reports the diagnostics test using the standardized innovations, ς_t (see its definition in Appendix D). The rejections of Jarque-Bera test and Brock-Dechert-Scheinkman (BDS) test imply the misspecification of the model for the SV-N, SV-SN, SV-T, SV-GT. We accept the SV-ST and SV-GST models. Figure 5 shows the quantile-quantile plot for standardized innovations ς_t. From this plot, we can see that the SV-GST outperforms the other SV models for the SZSE-CI returns.

5.4 Forecasting and Value at Risk

In order to examine the performance of VaR and ES forecast for the competing models, we use the data from December 13, 2012 to April 1, 2014 as validation period, giving $m = 310$ trading days. In the moving window approach, we use the first T observations in the period February 16, 2005 to December 12, 2012 to estimate the model and to forecast the $(T+1)$th observation; the sample is then rolled forward by one observation, so that the second to the $(T+1)$th observations are used to forecast the $(T+2)$th observation. This process is repeated until the end of the sample, i.e., the $(T+m)$th observation. We thus obtain 310 volatility forecasts, VaR and ES estimates with confidence levels of 5% and 95%. The competing models are: RiskMetrics, SV-N,

Table 6. SZSE-CI return data set. Violation rate in 310 one-step-ahead forecast, P-values of the unconditional coverage test, and the rank of the value-at-risk (VaR) at the 5% level

	Violation Rate (%)	LR_{uc}	D(0.05)	D(0.95)
RiskMetriks	0.073	0.1022	0.3511	0.2543
SV-N	0.062	0.1728	0.2030	0.1655
SV-T	0.068	0.1728	0.1060	0.1498
SV-GT	0.068	0.1728	0.1059	0.1351
SV-SN	0.068	0.1728	0.1039	0.0608
SV-ST	0.058	0.5248	0.0482	0.0158
SV-GST	0.061	0.3775	0.0500	0.0142

Table 7. MSPE for SZSE-CI returns

SV-N	SV-T	SV-GT	SV-SN	SV-ST	SV-GST
4.3722	4.3455	4.3131	4.3321	4.3088	4.2867

SV-T, SV-N, SV-ST and SV-GST. The results of 310 one-step-ahead forecasts are presented in Table 6. According to the unconditional coverage test we accept the null hypothesis that the achieved violation rate is equal to 5% for all the models. According to the violation rate the SV-ST and SV-GST give better performance than the other competing models. Similar results are obtained by comparing the $D(0.05)$ and $D(0.95)$.

In order to evaluate the prediction accuracy of the SV-N, SV-T, SV-N, SV-ST and SV-GST models, we use the same moving window approach as in the VaR estimation. We define the Mean Square Predictive Error (MSPE) as

$$MSPE^g = \sum_{t=T+1}^{T+m} \frac{1}{m} \sum_{i=1}^{N} \frac{1}{N} (y_t^{(i,g)} - y_t)^2$$

where $y_t^{(i,g)}$ is obtained by simulation using the MCMC procedure described in section 3.3 and $g \in$ {SV-N, SV-T, SV-N, SV-ST,SV-GST} denotes the model.

Table 7 shows that the SV-GST model outperforms the other models using the MSPE. Thus, according to the MSPE, the SV-GST model gives the best out-of-sample fit. It is also important to emphasize that, in general, we do not advocate the use of the SV-GST model in all situations but recommend using the model discussed here to assess the robustness of the conclusions, replacing the normal assumption with a more flexible model if this provides a more appropriate analysis.

6. DISCUSSION

In this article, we have proposed the stochastic volatility model with generalized skew-t (SV-GST) errors as an alternative to the normal (symmetric) assumption in the conditional distribution of the returns. The SV-GST model allows a parsimonious yet flexible treatment of both the skewness and the heaviness of the tails of the error distribution. Within the Bayesian framework, we have developed a fast and efficient MCMC sampling procedure to estimate all the parameters and latent quantities in our proposed SV-GST model. We used objective priors for the shape and the skewness parameters, ν and λ, based on Fonseca et al. [31] and Bayes and Branco [10], respectively. As a by-product of the MCMC algorithm, we were able to produce an estimate of the latent information process which can be used in financial modeling. The use of mixing variables, $\mathbf{U}_{1:T}$ not only simplifies the full conditional distributions required for the Gibbs sampling algorithm, but also provides a mean for the outlier diagnostics. We applied our methods to the analysis of the SZSE-CI return series, which showed that the SV-GST model provides a better fit than the SV-N, SV-T, SV-GT and SV-SN models in terms of parameter estimates, interpretation and robustness aspects. On the other hand, under the SV-GST model, the posterior mean and the 95% HPD interval of the parameter ν were respectively 8.8054 and $(6.3472, 12.1114)$, and the posterior mean and the entire 95% HPD interval of the parameter λ were below 0, indicating that there was a strong evidence of the skewness and heavy tails of the error distribution in the SZSE-CI data set. This fact was also found in the S&P 500, but it did not appear in the FTSE100 index returns (see Supplementary Materials). We found that the SV-GST model outperforms the other models using the MSPE given the best out-of-sample fit and it can be used to VaR and ES forecast.

A potential interesting future research topic is the further investigation of the large observations by introducing jump components or considering asymmetry threshold models.

ACKNOWLEDGEMENTS

The authors are grateful to the Editor, Associate Editor and referees for numerous suggestions that significantly improved the paper. Dr. M.-H. Chen's research was partially supported by NIH grant #GM 70335.

APPENDIX A. THE FULL CONDITIONALS

In this appendix, we describe the full conditional distributions of the parameters and the mixing latent variables $\mathbf{U}_{1:T}$ and $\mathbf{W}_{1:T}$ under the SV-GST model.

Full conditional distributions of μ, φ and σ_η^2

The prior distributions of the common parameters are specified as: $\mu \sim N(\bar{\mu}, \sigma_\mu^2)$, $\varphi \sim \mathcal{N}_{(-1,1)}(\bar{\varphi}, \sigma_\varphi^2)$, $\sigma_\eta^2 \sim \mathcal{IG}(\frac{T_0}{2}, \frac{M_0}{2})$. We have the following full conditional for μ:

$$\mu \mid \mathbf{h}_{1:T}, \varphi, \sigma_\eta^2 \sim \mathcal{N}(\frac{b_\mu}{a_\mu}, \frac{1}{a_\mu}), \quad \text{(A.1)}$$

where $a_\mu = \frac{1}{\sigma_\mu^2} + \frac{(T-1)(1-\varphi)^2}{\sigma_\eta^2} + \frac{1-\varphi^2}{\sigma_\eta^2}$ and $b_\mu = \frac{\bar{\mu}}{\sigma_\mu^2} + \frac{(1-\varphi^2)}{\sigma_\eta^2}h_1 + \frac{\sum_{t=1}^{T-1}(h_{t+1}-\varphi h_t)(1-\varphi)}{\sigma_\eta^2}$. In a similar way, the conditional distribution of φ is given by

$$p(\varphi \mid \mathbf{h}_{1:T}, \mu, \sigma_\eta^2) \propto Q(\varphi)\exp\{-\frac{a_\varphi}{2}(\varphi - \frac{b_\varphi}{a_\varphi})^2\}\mathbb{I}(|\varphi|<1), \tag{A.2}$$

where $Q_\varphi = \sqrt{1-\varphi^2}\exp\{-\frac{1}{2\sigma_\eta^2}[(1-\varphi^2)(h_1-\mu)^2\}$, $a_\varphi = \frac{\sum_{t=1}^{T-1}(h_t-\mu)^2}{\sigma_\eta^2} + \frac{1}{\sigma_\varphi^2}$, $b_\varphi = \frac{\sum_{t=1}^{T-1}(h_t-\mu)(h_{t+1}-\mu)}{\sigma_\eta^2} + \frac{\bar{\varphi}}{\sigma_\varphi^2}$ and $\mathbb{I}(\cdot)$ is the indicator function. Since the closed form expression of $p(\varphi \mid \mathbf{h}_{1:T}, \mu, \sigma_\eta^2)$ in (A.2) is not available, we sample from it by using the Metropolis-Hastings algorithm with a truncated $\mathcal{N}_{(-1,1)}(\frac{b_\varphi}{a_\varphi}, \frac{1}{a_\varphi})$ distribution as the proposal density.

Finally, the full conditional of σ_η^2 is $\mathcal{IG}(\frac{T_1}{2}, \frac{M_1}{2})$, where $T_1 = T_0 + T$ and $M_1 = M_0 + [(1-\varphi^2)(h_1-\mu)^2] + \sum_{t=1}^{T-1}[h_{t+1} - \mu - \varphi(h_t-\mu)]^2$.

Full conditional distributions of ν, λ, U_t and W_t

We set ζ and ω in a way such that $E(y_t \mid h_t) = 0$ and $V(y_t \mid h_t) = \mathrm{e}^{h_t}$. Thus, we have $\zeta = -\sqrt{\frac{2}{\pi}}k_1\delta\omega$ and $\omega^2 = [k_2 - \frac{2}{\pi}k_1^2\delta^2]^{-1}$, where $k_1 = \sqrt{\frac{1}{2}}\frac{\Gamma(\frac{\nu-1}{2})}{\Gamma(\frac{\nu}{2})}$, $k_2 = \frac{1}{\nu-2}$ and $\delta = \frac{\lambda}{\sqrt{1+\lambda^2}}$. Then the full conditional distributions of ν and λ are given as follows:

$$\begin{aligned} & p(\nu \mid .) \\ \propto\ & \left(\frac{\nu}{\nu+3}\right)^{\frac{1}{2}}\left\{\psi'\left(\frac{\nu}{2}\right) - \psi'\left(\frac{\nu+1}{2}\right) - \frac{2(\nu+3)}{\nu(\nu+1)^2}\right\}^{\frac{1}{2}} \\ \times\ & \left(\frac{1}{2}\right)^{\frac{T\nu}{2}}\mathrm{e}^{\frac{\nu}{2}\sum_{t=1}^T \log U_t}[\Gamma(\frac{\nu}{2})]^{-T}\left(\frac{1}{\omega}\right)^T \\ \times\ & \mathrm{e}^{-\frac{1}{2\omega^2(1-\delta^2)}\sum_{t=1}^T U_t \mathrm{e}^{-h_t}(y_t-\zeta-\omega\delta W_t U_t^{-\frac{1}{2}}\mathrm{e}^{\frac{h_t}{2}})^2}, \\ & p(\lambda \mid .) \\ \propto\ & \left(1+\frac{2\lambda}{\frac{\pi^2}{4}}\right)^{-\frac{3}{4}}\left(\frac{1}{1-\delta^2}\right)^{\frac{T}{2}} \\ \times\ & \mathrm{e}^{-\frac{1}{2\omega^2(1-\delta^2)}\sum_{t=1}^T U_t \mathrm{e}^{-h_t}(y_t-\zeta-\omega\delta W_t U_t^{-\frac{1}{2}}\mathrm{e}^{\frac{h_t}{2}})^2}. \end{aligned}$$

Since the above full conditional distributions are not in any known closed form, we must simulate ν and λ using the Metropolis-Hastings algorithm. The proposal density used for ν and λ are $\mathcal{N}_{(2,\infty)}(\mu_\nu, \tau_\nu^2)$ and $\mathcal{N}(\mu_\lambda, \tau_\lambda^2)$, respectively, with $\mu_\upsilon = x - \frac{q'(x)}{q''(x)}$ and $\tau_\upsilon^2 = \max\{0.001, (-q''(x))^{-1}\}$ for $\upsilon = \nu$ or λ, where x is the value of the previous iteration, $q(.)$ is the logarithm of the conditional posterior density, and $q'(.)$ and $q''(.)$ are the first and second derivatives, respectively.

Since $U_t \sim \mathcal{G}(\frac{\nu}{2}, \frac{1}{2})$, the conditional distribution of U_t is given by

$$\begin{aligned} & p(U_t \mid h_t, W_t, \nu, \lambda) \\ \propto\ & Q(U_t)U_t^{\frac{\nu+1}{2}-1}\mathrm{e}^{-\frac{U_t}{2}[1+\frac{\mathrm{e}^{-h_t}(y_t-\zeta\mathrm{e}^{\frac{h_t}{2}})^2}{\omega^2(1-\delta^2)}]}, \end{aligned} \tag{A.3}$$

where $Q(U_t) = \mathrm{e}^{\frac{U_t^{\frac{1}{2}}\delta W_t \mathrm{e}^{-\frac{h_t}{2}}(y_t-\zeta\mathrm{e}^{\frac{h_t}{2}})}{\omega(1-\delta^2)}}$. However, $p(U_t \mid h_t, W_t, \nu, \lambda)$ in (A.3) does not have a closed form expression, we shall sample from it by using the Metropolis-Hastings algorithm with $\mathcal{G}(\frac{\nu+1}{2}, \frac{1}{2}[1+\frac{\mathrm{e}^{-h_t}(y_t-\zeta\mathrm{e}^{\frac{h_t}{2}})^2}{\omega^2(1-\delta^2)}])$ as the proposal density. Finally, using Equations (5a) and (5c), we obtain the full conditional distribution of W_t given by $\mathcal{N}_{[0,\infty)}(\frac{\delta U_t^{\frac{1}{2}}\mathrm{e}^{-\frac{h_t}{2}}[y_t-\zeta\mathrm{e}^{\frac{h_t}{2}}]}{\omega}, \frac{1}{1-\delta^2})$.

APPENDIX B. THE BLOCK SAMPLER

In order to simulate $\mathbf{h}_{1:T} = (h_1, \ldots, h_T)'$ in the SV-ST model, we consider a two-step process. First, we simulate h_1 conditional on $\mathbf{h}_{2:T}$ and then draw $\mathbf{h}_{2:T}$ conditional on h_1. To sample the vector $\mathbf{h}_{2:T}$, we develop a multi-move block algorithm. In our block sampler, we divide it into $K+1$ blocks, $\mathbf{h}_{k_{l-1}+1:k_l-1} = (h_{k_{l-1}+1}, \ldots, h_{k_l-1})'$ for $l = 1, \ldots, K+1$, with $k_0 = 1$ and $k_{K+1} = T$, where $k_l - 1 - k_{l-1} \geq 2$ is the size of the $l-$th block. We sample the block of disturbances $\boldsymbol{\eta}_{k_{l-1}:k_l-2} = (\eta_{k_{l-1}}, \ldots, \eta_{k_l-2})'$ given the end conditions $h_{k_{l-1}}$ and h_{k_l} instead of $\mathbf{h}_{k_{l-1}+1:k_l-1}$. In order to facilitate the exposition, we omit the dependence on $\boldsymbol{\theta}$, $\mathbf{W}_{t+1:t+k}$ and $\mathbf{U}_{t+1:t+k}$, and suppose that $k_{l-1} = t$ and $k_l = t+k+1$ for the $l-$th block, such that $t+k<T$. Then $\boldsymbol{\eta}_{t:t+k-1} = (\eta_t, \ldots, \eta_{t+k-1})'$ are sampled at once from their full conditional distribution $f(\boldsymbol{\eta}_{t:t+k-1}|h_t, h_{t+k+1}, \mathbf{y}_{t:t+k})$, which without the constant terms is expressed in the log scale as

$$\begin{aligned} & \log f(\boldsymbol{\eta}_{t:t+k-1}|h_t, h_{t+k+1}) \\ =\ & \text{const} - \frac{1}{2}\sum_{r=t}^{t+k-1}\eta_r^2 + \sum_{r=t+1}^{t+k} l(h_r) \\ -\ & \frac{1}{2\sigma_\eta^2}[h_{t+k+1} - \mu - \varphi(h_{t+k}-\mu)]^2\mathbb{I}(t+k<T). \end{aligned} \tag{B.1}$$

We denote the first and second derivatives of $l(h_r)$ with respect to h_r by l' and l'', where $l(h_r) = \log p(y_r \mid \nu, \lambda, W_r, U_r, h_r)$ is obtained from Equation (5a). As (B.1) does not have a closed form, we use the Metropolis-Hastings acceptance-rejection algorithm (64, 21) to sample from (B.1). We propose to use the following artificial Gaussian state space model as a proposed density to simulate the block $\boldsymbol{\eta}_{t+1:t+k}$

$$\hat{y}_r = h_r + \xi_r,\ \xi_r \sim \mathcal{N}(0, d_r),\ r = t+1, \ldots, t+k, \tag{B.2}$$

$$\begin{aligned} & h_{r+1} = \mu + \varphi(h_r-\mu) + \sigma_\eta\eta_r,\ \eta_r \sim \mathcal{N}(0,1), \\ & r = t, t+1, \ldots, t+k-1, \end{aligned} \tag{B.3}$$

where the auxiliary variables d_r and $\hat{y}_r$ for $r = t+1, \ldots, t+k-1$ and $t+k=T$ are defined as follows:

$$\begin{aligned} d_r &= -\frac{1}{l_F''(\hat{h}_r)}, \\ \hat{y}_r &= \hat{h}_r + d_r l'(\hat{h}_r). \end{aligned} \tag{B.4}$$

For $r = t+k < T$, it follows that

$$d_r = \frac{\sigma_\eta^2}{\varphi^2 - \sigma_\eta^2 l_F''(\hat{h}_{t+k})},$$

$$\hat{y}_r = d_r\left[l'(\hat{h}_r) - l_F''(\hat{h}_r)\hat{h}_r + \frac{\varphi}{\sigma_\eta^2}[h_{r+1} - \mu(1-\varphi)]\right]. \tag{B.5}$$

We obtain the measurement equation (B.2) by a second-order expansion of l_r around some preliminary estimate of η_r, denoted by $\hat{\eta}_r$, where $\hat{h}_r$ is the estimate of h_r equivalent to $\hat{\eta}_r$, and

$$l_F''(h_r) = E[l''(h_r)] = -\frac{1}{2} - \frac{(\zeta + \omega\delta W_t U_t^{-\frac{1}{2}})^2}{4\omega^2(1-\delta^2)} U_r, \tag{B.6}$$

which is everywhere strictly negative. The expectation in (B.6) is taken with respect to y_r conditional on h_r, W_r, U_r, $\boldsymbol{\theta}$. Since (B.2)–(B.3) define a Gaussian state space model, we can apply de Jong and Shephard's simulation smoother (27) to perform the sampling. We denote this density by g. Since f is not bounded by g, we use the Metropolis-Hastings acceptance-rejection algorithm to sample from f, as recommended by Chib and Greenberg [21]. In the SV-SN case, we use the same procedure with $U_t = 1$ for $t = 1, \ldots, T$.

The procedure to select the expansion block $\hat{\mathbf{h}}_{t+1:t+k}$ is described in the Algorithm B.1.

Algorithm B.1.

1. Initialize $\hat{\mathbf{h}}_{t+1:t+k}$.
2. Evaluate recursively $l'(\hat{h}_r)$ and $l_F''(\hat{h}_r)$ for $r = t+1, \ldots, t+k$.
3. Conditional on the current values of the vector of parameters $\boldsymbol{\theta}$, $\mathbf{U}_{t+1:t+k}$, $\mathbf{W}_{t+1:t+k}$, h_t and h_{t+k+1}, define the auxiliary variables $\hat{y}_r$ and d_r using equations (B.4) or (B.5) for $r = t+1, \ldots, t+k$.
4. Consider the linear Gaussian state-space model in (B.2) and (B.3). Apply the Kalman filter and a disturbance smoother [44] and obtain the posterior mean of $\boldsymbol{\eta}_{t:t+k}$ ($\mathbf{h}_{t:t+k}$) and set $\hat{\boldsymbol{\eta}}_{t:t+k}$ ($\hat{\mathbf{h}}_{t:t+k}$) to this value.
5. Return to Step 2 and repeat the procedure until achieving convergence.

Finally, we describe the updating procedure for h_1 and the knot conditions h_{k_l}, for $l = 1, \ldots, K$. First, we simulate h_1 from $p(h_1|h_2, \boldsymbol{\theta}, \mathbf{y}_{1:T})$ by using the Metropolis-Hasting (MH) algorithm with the normal density, $\mathcal{N}(\mu + \varphi[h_2 - \mu], \sigma_\eta^2)$, as a proposal. Then, the acceptance probability is given by $\alpha_{MH} = \min\{1, \frac{Q(h_1^p)}{Q(h_1^{(i-1)})}\}$, where $Q(h_1)$ is the conditional density of $y_1 \mid \boldsymbol{\theta}, W_1, U_1, h_1$. Let h_1^p and $h_1^{(i-1)}$ denote the proposal and the previous iteration values. As the density $p(h_{k_l}|h_{k_l-1}, h_{k_l+1})$ does not have a closed form, we use the MH algorithm with the proposal $\mathcal{N}(\frac{\mu(1-\varphi)^2 + \varphi(h_{k_l-1} + h_{k_l+1})}{1+\varphi^2}, \frac{\sigma_\eta^2}{1+\varphi^2})$. As before, $h_{k_l}^p$ and $h_{k_l}^{(i-1)}$ denote the proposal and the previous iteration values, respectively. Thus, the acceptance probability is given by $\alpha_{MH} = \min\{1, \frac{Q(h_{k_l}^p)}{Q(h_{k_l}^{(i-1)})}\}$, where $Q(h_{k_l})$ is the conditional density of y_{k_l} given $\boldsymbol{\theta}, W_{k_l}, U_{k_l}, h_{k_l}$.

APPENDIX C. THE AUXILIARY PARTICLE FILTER

In the filtering problem the goal is to sample random variates $\{h_t^{(1)}, \ldots, h_t^{(N)}\}$ from the filtering distribution $p(h_t|\boldsymbol{\theta}, \mathbf{y}_{1:t})$. We employ the auxiliary particle filter (APF) introduced by Pitt and Shephard [54], which allows us to draw samples from the filtering distribution $p(h_t|\boldsymbol{\theta}, \mathbf{y}_{1:t})$ by numerical approximation.

First, let us consider $\{(h_{t-1}^{(1)}, w_{t-1}^{(1)}), \ldots, (h_{t-1}^{(N)}, w_{t-1}^{(N)})\} \sim p(h_{t-1}|\boldsymbol{\theta}, \mathbf{y}_{1:t-1})$, where the probability density function, $p(h_{t-1}|\boldsymbol{\theta}, \mathbf{y}_{1:t-1})$, of the continuous random variable, h_{t-1}, is approximated by a discrete variable with random support. It then follows that the one-step ahead predictive distribution $p(h_t|\boldsymbol{\theta}, \mathbf{y}_{1:t-1})$ can be approximated as:

$$\begin{aligned} p(h_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1}) &= \int p(h_t \mid h_{t-1}, \boldsymbol{\theta}) p(h_{t-1} \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1}) dh_{t-1} \\ &\approx \sum_{i=1}^{N} p(h_t \mid \boldsymbol{\theta}, h_{t-1}^{(i)}) w_{t-1}^{(i)}, \end{aligned} \tag{C.1}$$

where $h_{t-1}^{(i)}$ is a sample from $p(h_{t-1} \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1})$ with weight $w_{t-1}^{(i)}$. The one-step ahead density, $p(y_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1})$, is then estimated by Monte Carlo averaging of $p(y_t \mid \boldsymbol{\theta}, h_t)$ over the draws of $h_t^{(i)} \sim p(h_t \mid \boldsymbol{\theta}, h_{t-1}^{(i)})$ as follows:

$$\begin{aligned} p(y_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1}) &= \int p(y_t \mid h_t, \boldsymbol{\theta}) p(h_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1}) dh_t \\ &\approx \sum_{i=1}^{N} p(y_t \mid \boldsymbol{\theta}, h_t^{(i)}) w_{t-1}^{(i)}. \end{aligned} \tag{C.2}$$

This recursive procedure needs to draw h_t sequentially from the filtered distribution, $p(h_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t})$, which is updated as described in the APF Algorithm.

The APF Algorithm

1. Posterior at $t-1$:
   $\{(h_{t-1}^{(1)}, w_{t-1}^{(1)}), \ldots, (h_{t-1}^{(i)}, w_{t-1}^{(i)}), \ldots, (h_{t-1}^{(N)}, w_{t-1}^{(N)})\} \sim p(h_{t-1} \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1})$.
2. For $i = 1, \ldots, N$, calculate $\mu_t^{(i)} = \mu + \varphi(h_{t-1}^{(i)} - \mu)$.
3. Sampling (k, h_t):
   For $i = 1, \ldots, N$

Indicator: k^i such that $P(k^i = k) \propto p(y_t \mid \mu_t^{(k^i)}) w_{t-1}^{(k^i)}$.
Evolution:

$$h_t^{(i)} \sim \mathcal{N}(\mu_t^{k^i}, \sigma^2).$$

Weights: compute $w_t^{(i)}$ as follows

$$w_t^{(i)} \propto \frac{p(y_t \mid \boldsymbol{\theta}, h_t^{(i)})}{p(y_t \mid \boldsymbol{\theta}, \mu_t^{(k^i)})}.$$

4. Posterior at t: $\{(h_t^{(1)}, w_t^{(1)}), \ldots, (h_t^{(i)}, w_t^{(i)}), \ldots, (h_t^{(N)}, w_t^{(N)})\} \sim p(h_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t})$.

The Log-likelihood Estimation Algorithm

1. Set $t = 1$ and obtain a sample $h_{t-1}^{(i)}$.
2. For each value of $h_{t-1}^{(i)}$ sample

$$h_t^{(i)} \sim \mathcal{N}(\mu + \varphi(h_{t-1}^{(i)} - \mu), \sigma_\eta^2).$$

3. Estimate the one-step ahead density as

$$p(y_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1}) \approx \sum_{i=1}^{N} p(y_t \mid \boldsymbol{\theta}, h_t^{(i)}) w_{t-1}^{(i)}.$$

4. Apply the filtering procedure in the APF Algorithm to obtain $\{(h_t^{(1)}, w_t^{(1)}), \ldots, (h_t^{(i)}, w_t^{(i)}), \ldots, (h_t^{(N)}, w_t^{(N)})\}$.
5. Return the log likelihood ordinate

$$\log p(\mathbf{y}_{1:T} \mid \boldsymbol{\theta}) = \frac{1}{T} \sum_{t=1}^{T} \log p(y_t \mid \boldsymbol{\theta}, \mathbf{y}_{1:t-1}).$$

APPENDIX D. DIAGNOSTICS

In order to check the distribution assumptions of the SV models, we use an approach similar to Kim et al. [43]. The diagnostics test is based on the probability integral transform of the realizations y_{t+1}^o taken with respect to the one-step-ahead prediction density $p(y_{t+1} \mid \mathbf{y}_{1:t}, \boldsymbol{\theta})$. The probability integral transform, ε_{t+1}, is simply the cumulative distribution function corresponding to the prediction density $p(y_{t+1} \mid \mathbf{y}_{1:t}, \boldsymbol{\theta})$ evaluated at $y_{t+1}^o : \varepsilon_{t+1} = Prob(y_{t+1} \leq y_{t+1}^o \mid \mathbf{y}_{1:t}, \boldsymbol{\theta})$. For $t = 1, \ldots, T$, under the null hypothesis that the true distribution of y_{t+1}^o is $p(y_{t+1} \mid \mathbf{y}_{1:t}, \boldsymbol{\theta})$ (or equivalently, the model is correctly specified), the ε_{t+1} converges in distribution to independent and identically distributed uniform random variables on $[0, 1]$ [see, 57, 59, 43, 32, 47, among others]. By letting $\varsigma_{t+1} = \Phi^{-1}(\varepsilon_{t+1})$, where $\Phi()$ denotes the standard normal cumulative distribution function, a sequence of independent standard normal random variables ς_{t+1} is obtained, which are the standardized innovations. The probability $Prob(y_{t+1} \leq y_{t+1}^o \mid \mathbf{y}_{1:t}, \boldsymbol{\theta})$ can then be approximated by

$$\begin{aligned} & Prob(y_{t+1} \leq y_{t+1}^o \mid \mathbf{y}_{1:t}, \boldsymbol{\theta}) \\ = \; & \frac{1}{N} \sum_{i=1}^{N} Prob(y_{t+1} \leq y_{t+1}^o \mid \mathbf{y}_{1:t}, h_{t+1}^{(i)}, \boldsymbol{\theta}). \end{aligned}$$

Received 5 September 2013

REFERENCES

[1] Aas, K., and Haff, I. H. [2006], "The Generalized Hyperbolic Skew Student's t-Distribution," *Journal of Financial Econometrics*, 4, 275–309.
[2] Abanto-Valle, C. A., Bandyopadhyay, D., Lachos, V., and Enriquez, I. [2010], "Robust Bayesian analysis of heavy-tailed stochastic volatility models using scale mixtures of normal distributions," *Computational Statistics and Data Analysis*, 54, 2883–2898. MR2727721
[3] Abanto-Valle, C. A., Lachos, V. H., and Dey, D. K. [2011], Bayesian estimation of a skew-t stochastic volatility model, Technical report, Department of Statistics, Federal University of Rio de Janeiro, Brazil.
[4] Abanto-Valle, C. A., Migon, H. S., and Lachos, V. H. [2012], "Stochastic volatility in mean models with heavy-tailed distributions," *Brazilian Journal of Probability and Statistics*, 26(4), 402–422. MR2949086
[5] Ando, T. [2006], "Bayesian Inference for nonlinear and non-gaussian stochastic volatility model with leverge effect," *Journal of the Japan Statistical Society*, 36, 173–197. MR2338878
[6] Ando, T. [2007], "Bayesian predictive information criterion for the evaluation of hierarchical Bayesian and empirical Bayes models," *Biometrika*, 94, 443–458. MR2380571
[7] Artzner, P., Delbaen, F., Eber, J. M., and Heath, D. [1997], "Thinking coherently," *Risk*, 10(11), 68–71.
[8] Artzner, P., Delbaen, F., Eber, J. M., and Heath, D. [1999], "Coherent measures of risk," *Mathematical Finance*, 9, 203–228. MR1850791
[9] Asai, M. [2009], "Bayesian analysis of stochastic volatility models with mixture-of-normal distributions," *Mathematics and Computers in Simulation*, 79(8), 2579–2596. MR2531471
[10] Bayes, C. L., and Branco, M. D. [2007], "Bayesian inference for the skewness parameter of the scalar skew-normal distribution," *Brazilian Journal of Probability and Statistics*, 21, 141–163. MR2397043
[11] Berg, A., Meyer, R., and Yu, J. [2004], "Deviance Information Criterion for comparing stochastic volatility models," *Journal of Business and Economic Statistics*, 22, 107–120. MR2019793
[12] Bollerslev, T. [1986], "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307–327. MR0853051
[13] Brock, W. A., D., W., Scheinkman, J. A., and LeBaron, B. [1996], "A test for independence based on the correlation dimension," *Econometric Reviews*, 15, 197–235. MR1410877
[14] Cappuccio, N., Lubian, D., and Raggi, D. [2004], "MCMC Bayesian estimation of a skew-GED stochastic volatility model," *Studies in Nonlinear Dynamics & Econometrics*, 8(2). MR2164136
[15] Cappuccio, N., Lubian, D., and Raggi, D. [2006], "Investigating asymmetry in U.S. stock market indexes: evidence from a stochastic volatility model," *Applied Financial Economics*, 16, 479–490.
[16] Carnero, M. A., Peña, D., and Ruiz, E. [2004], "Persistence and Kurtosis in GARCH and Stochastic volatility models," *Journal of Financial Econometrics*, 2, 319–342.
[17] Celeux, G., Forbes, F., Robert, C. P., and Titterington, D. M. [2006], "Deviance information criteria for missing data models," *Bayesian Analysis*, 1, 651–674. MR2282197

[18] Chen, C. W. S., Liu, F. C., and So, M. K. P. [2008], "Heavy-tailed-distributed threshold stochastic volatility models in financial time series," *Australian & New Zeland Journal of Statistics*, 50, 29–51. MR2414654

[19] Chen, M.-H., Shao, Q.-M., and Ibrahim, J. G. [2000], *Monte Carlo Methods in Bayesian Computation*, New York: Springer-Verlag. MR1742311

[20] Chen, Q., Gerlach, R., and Lub, Z. [2012], "Bayesian Value-at-Risk and expected shortfall forecasting via the asymmetric Laplace distribution," *Computational Statistics and Data Analysis*, 56, 3498–3516. MR2943909

[21] Chib, S., and Greenberg, E. [1995], "Understanding the Metropolis-Hastings algorithm," *The American Statistician*, 49, 327–335.

[22] Chib, S., Nardari, F., and Shephard, N. [2002], "Markov chain Monte Carlo methods for stochastic volatility models," *Journal of Econometrics*, 108(2), 281–316. MR1894758

[23] Choy, S., and Chan, C. [2000], "Bayesian estimation of stochastic volatility model via scale mixtures distributions," *Statist. Finance Interface*, 1, 186–204. MR1830258

[24] Choy, S., Wan, W. Y., and Chan, C. [2008], "Bayesian Student-t Stochastic Volatility Models via Scale Mixtures," *Advances in Econometrics*, 23, 595–618.

[25] Chunhachinda, P., Dandapani, K., Hamid, K., and Prakash, A. J. [1997], "Portfolio selection and skewness: evidence from international stock markets," *Journal of Banking and Finance*, 21, 143–167.

[26] Corrado, C. J., and Su, T. [1997], Implied volatility skews and stock index skewness and kurtosis implied by S&P500 index option prices, Technical report, Working Paper, University of Missouri, Columbia.

[27] de Jong, P., and Shephard, N. [1995], "The Simulation Smoother for Time Series Models," *Biometrika*, 82, 339–350. MR1354233

[28] Delatola, E.-I., and Griffin, J. E. [2011], "Bayesian Nonparametric Modelling of the Return Distribution with Stochastic Volatility," *Bayesian Analysis*, 6, 901–926. MR2869968

[29] Embrechts, P., Kaufmann, R., and Patie, P. [2004], "Strategic Long-term Financial Risks: Single Risk Factors," *Computational Optimization and Applications*, 22, 61–90. MR2157327

[30] Fan, T.-H., and Wang, Y.-F. [2013], "An empirical Bayesian forecast in the threshold stochastic volatility models," *Journal of Statistical Computation and Simulation*, 83, 486–500. MR3028378

[31] Fonseca, T. C. O., Ferreira, M. A. R., and Migon, H. S. [2008], "Objective Bayesian analysis for the Student-t regression model," *Biometrika*, 95, 325–333. MR2521587

[32] Gerlach, R., Carter, C., and Kohn, R. [1999], "Diagnostics for time series analysis," *Journal of Time Series Analysis*, 20, 309–330. MR1693161

[33] Geweke, J. [1992], "Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments," in *Bayesian Statistics. Vol. 4*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, Oxford, U.K.: Oxford University Press, pp. 169–193. MR1380276

[34] Gneiting, T., and Raftery, A. E. [2007], "Strictly proper scoring rules, prediction and estimation," *Journal of the American Statistical Association*, 6, 901–926. MR2345548

[35] Good, I. J. [1952], "Rational Decisions," *Journal of the Royal Statistical Society, Series B*, 14, 107–114. MR0077033

[36] Hansen, B. E. [1994], "Autoregressive Conditional Density Estimation," *International Economic Review*, 35, 705–730.

[37] Harvey, C., and Sidique, A. [1999], "Autoregressive conditional skewness," *Journal of Finance and Quantitative Analysis*, 34, 465–487.

[38] Harvey, C., and Sidique, A. [2000], "Conditional skewness in asset pricing tests," *Journal of Finance*, 55, 1263–1296.

[39] Jacquier, E., Polson, N. G., and Rossi, P. E. [2004], "Bayesian analysis of stochastic volatility models with fat-tails and correlated errors," *Journal of Econometrics*, 122(1), 185–212. MR2083256

[40] Jacquier, E., Polson, N., and Rossi, P. [1994], "Bayesian analysis of stochastic volatility models," *Journal of Business and Economic Statistics*, 12, 371–418.

[41] Jondeau, E., and Rockinger, M. [2003], "Conditional volatility, skewness and kurtosis: Existence, persistence and co-movements," *Journal of Economics Dynamics and Control*, 27, 1699–1737. MR1981727

[42] Kim, S., Chen, M.-H., and Dey, D. K. [2008], "Flexible generalized t-link models for binary response data," *Biometrika*, 95, 93–106. MR2409717

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

[44] Koopman, S. [1993], "Disturbance smoothers for State Space models," *Biometrika*, 80, 117–126. MR1225219

[45] Kupiec, P. H. [1995], "Techniques for verifying the accuracy of risk measurement models," *The Journal of Derivatives*, 3, 73–84.

[46] Liesenfeld, R., and Jung, R. C. [2000], "Stochastic volatility models: conditional normality versus heavy-tailed distributions," *Journal of Applied Econometrics*, 15(2), 137–160.

[47] Liesenfeld, R., and Richard, J.-F. [2003], "Univariate and multivariate stochastic volatility models: estimation and diagnostics," *Journal of Empirical Finance*, 10, 505–531.

[48] Mahieu, R., and Schotman, P. C. [1998], "Am empirical application of stochastic volatility models," *Journal of Applied Econometrics*, 13, 333–360.

[49] Melino, A., and Turnbull, S. M. [1990], "Pricing foreign options with stochastic volatility," *Journal of Econometrics*, 45, 239–265.

[50] Mittnik, S., and Paolella, M. S. [2000], "Conditional density and value-at-risk prediction of asian currency exchange rates," *Journal of Forecasting*, 19, 313–333.

[51] Nakajima, J. [2013], "Stochastic volatility model with regime-switching skewness in heavy-tailed errors for exchange rate returns," *Studies Nonlinear Dynamics Economics*, 175(5), 499–520. MR3141393

[52] Nakajima, J., and Omori, Y. [2012], "Stochastic volatility model with leverage and assymmetrically heavy-tailed error using GH skew Student's t-distribution," *Computational Statistics and Data Analysis*, 56, 3690–3704. MR2943921

[53] Pemstein, D., Quinn, K. V., and Martin, A. D. [2011], "The Scythe statistical library: An open source C++ library for statistical computation," *Journal of Statistical Software*, 42, 1–26.

[54] Pitt, M., and Shephard, N. [1999], "Filtering via simulation: Auxiliary particle filter," *Journal of the American Statistical Association*, 94, 590–599. MR1702328

[55] Robert, C. P., and Titterington, D. M. [2002], "Discussion on "Bayesian measures of model complexity and fit"," *Biometrical Journal*, 64, 573–590.

[56] Rosa, G. J. M., Padovani, C. R., and Gianola, D. [2003], "Robust linear mixed models with Normal/Independent distributions and Bayesian MCMC implementation," *Biometrical Journal*, 45, 573–590. MR1998137

[57] Rosenblatt, M. [1952], "Remarks on a multivariate transformation," *Annals of Mathematical Statistics Volume 23*, 23, 470–472. MR0049525

[58] Shephard, N., and Pitt, M. [1997], "Likelihood analysis of non-Gaussian measurements time series," *Biometrika*, 84, 653–667. MR1603940

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

[60] Spiegelhalter, D. J., Best, N. G., Carlin, B. P., and van der Linde, A. [2002], "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society, Series B*, 64, 621–622. MR1979380

[61] Stone, M. [2002], "Discussion on "Bayesian measures of model complexity and fit"," *Journal of the Royal Statistical Society, Series B*, 64, 621. MR1979380

[62] Tauchen, G. E., and Pitts, M. [1983], "The price variability-volume relationshis in speculative markets," *Econometrica*, 51, 485–506.

[63] Taylor, S. [1982], "Financial returns modelled by the product of two stochastic processes-a study of the daily sugar prices 1961–75," in *Time Series Analysis: Theory and Practice, Vol 1*, ed. O. Anderson, Amsterdam: North-Holland, pp. 203–226.

[64] Tierney, L. [1994], "Markov chains for exploring posterior distributions (with discussion)," *Annal of Statistics*, 21, 1701–1762. MR1329166

[65] Tsiotas, G. [2012], "On generalised asymmetric stochastic volatility models," *Computational Statistics and Data Analysis*, 56(1), 151–172. MR2833044

[66] Watanabe, T., and Omori, Y. [2004], "A multi-move sampler for estimate non-Gaussian time series model: Comments on Shepard and Pitt (1997)," *Biometrika*, 91, 246–248. MR2050475

Carlos A. Abanto-Valle
Departament of Statistics
Federal University of Rio de Janeiro
Caixa Postal 68530, CEP: 21945-970
Rio de Janeiro
Brazil
E-mail address: cabantovalle@im.ufrj.br

Caifeng Wang
School of Mathematics and Physics
University of Science and Technology Beijing
Beijing, 100083
China
E-mail address: cfwang@ustb.edu.cn

Xiaojing Wang
Department of Statistics
University of Connecticut
215 Glenbrook Rd, U-4120
Storrs, CT 06269
USA
E-mail address: xiaojing.wang@uconn.edu

Fei-Xing Wang
School of Mathematics and Physics
University of Science and Technology Beijing
Beijing, 100083
China
E-mail address: wfeixing@vip.sina.com

Ming-Hui Chen
Department of Statistics
University of Connecticut
215 Glenbrook Rd, U-4120
Storrs, CT 06269
USA
E-mail address: ming-hui.chen@uconn.edu