

Modeling skewness and kurtosis in Stochastic Volatility Models

Georgios Tsiotas*

University of Crete, Department of Economics, GR

December 19, 2006

Abstract

Stochastic volatility models have been seen as a real alternative to conditional variance models assuming that volatility follows a different than the observed stochastic process. However, issues such as data's normality violations in the form of excess kurtosis and skewness can give rise to the use of distributional assumptions away from normality. Here, the noncentral t-distribution is used in the stochastic volatility model set-up. By nesting both excess kurtosis and skewness in the same specification, we derive the noncentral-t stochastic volatility model which counts for two types of normality violations. Thus, we generalise stochastic volatility analysis, in a way that the non-skewed stochastic volatility model nests the skewed one. In this framework, a fully Bayesian estimation approach is followed where the Markov Chain Monte Carlo engine is used for parametric and log-volatility estimation. The new model is then investigated for its performance using real financial data series.

Key words: Stochastic volatility, non-central t-distribution, Metropolis-Hastings , MCMC, DIC, Model selection.

*Present Address: University of Crete, Department of Economics, Panepistimioupolis, Rethymnon 74100, GR, Tel. +30 28310 77426, Fax. +30 28310 77406.

1 Introduction

Stochastic volatility (SV) models have come as a natural alternative to conditional variance models of the ARCH family. They allow volatility to be seen as a different stochastic process than the observed one in a way that the observed and the latent-volatility processes are driven by separate error terms. It has attracted much interest as a way of generalising the Black-Scholes option pricing formula that allows volatility persistence in asset returns (Hull and White (1987), Jacquier et al. (1994)). In its basic version SV model assumes that error disturbances are stationary uncorrelated Gaussian white noise ones (Harvey and Ruiz (1993), Jacquier et al.(1994)). Some additional extensions involve the use of either mixtures of normals or the t-distribution as an approximate to the asset's return deviations from normality towards fat-tailness (Geweke (1994), Shephard and Pitt (1997)). Some other SV model extensions involve the "leverage" effect assumption in which the observed and latent volatility process innovations are correlated. Under this assumption observed and volatility innovations are negatively correlated in which negative return shocks are associated with volatility increases (Gallant et al. (1994), Jacquier et al.(2004)).

Despite the big number of papers devoted to the long-tailed probability nature of observed financial return process, little to no attention has been put on its asymmetric nature. Fernandez and Steel (1998) have considered the use of an asymmetric distribution with different scale parameters to derive the left and the right skewness of the distribution. However, issues like the co-existence of both asymmetry and long-tailness has only recently been considered in financial econometrics literature (Tsionas (2002)). Skewness, is well documented in many economics and financial data series such as exchange rates and stock returns (Harvey and Siddique (1999)(2000), Jondeau and Rockinger (2003)). Negative skewness in returns can be viewed as the case where negative returns of a given magnitude are more likely than positive ones of the same magnitude (Harvey and Siddique (1999)). Thus, in portfolio analysis, negative skewness makes a portfolio less preferable than a positively skewed one.

In conditional and SV literature there have been some attempts of modeling skewness in a Gaussian or an asymptotically Gaussian framework (Higgins et al. (1992), Tsiotas (2002, 2007)). However, no attention has been given to the introduction of distribution away from normality. In this paper, we intent to incorporate these normality violation assumption within the SV framework by using the noncentral t-distributional assumption. In doing so, we will create

a generalised t-distribution SV model where the symmetric t-distribution SV model will be nested in the asymmetric one. Therefore, data long-tailness can be treated together with asymmetric frequency distribution in data with volatility persistence.

At the estimation stage, although the observed return process is not assumed to follow a Gaussian process, the latent log-volatility process is restricted in a Gaussian space. Parametric and log-volatility estimation is then implemented in a Markov Chain Monte Carlo (MCMC) set-up. This overcomes the non-conjugacy of the conditional distribution creating very efficient simulation results (Shephard (1994), Shephard and Pitt (1997)). Due to normality departures in the observed process, we use the Metropolis-Hastings algorithm within the MCMC engine (Shephard and Pitt (1997)).

The SV model is demonstrated in terms of its specification and its Bayesian inference using three different models. These are then compared for their ability to capture data normality violations using standard Bayesian model selection estimators such as the DIC one (Springelhalter et al. (2002)). In doing so, we will measure the models' ability to capture return series second order dependency. Finally, the robustness of the posterior density results of the specification selected via the DIC estimator is further tested. Results show that the noncentral-t SV model outperforms all the other competing models.

The structure of the paper is the following. Section 2 outlines the existing SV model approaches together with the newly derived noncentral-t SV model. Special reference is given to the Bayesian inference strategies, the priors, and the MCMC algorithms used in the the three competing models. Section 3 describes the empirical results based on daily exchange rates data series. We focus on the model selection issues based on the DIC estimator. Finally, a sensitivity testing experiment of the skewness prior is demonstrated for the robustification of the inference results.

2 SV models

2.1 The basic SV model

In literature SV models have received much attention due to treating volatility as a stochastic process different than the observed one taking into account its variability as an additional to that of the financial series returns. The basic SV model consists of

$$y_t = e^{h_t/2} \cdot \epsilon_t, \quad h_t = \mu + \phi(h_{t-1} - \mu) + \sigma_u \cdot u_t \quad t = 1, \dots, T \quad (1)$$

$$\epsilon_t \sim Niid(0, 1), \quad u_t \sim Niid(0, 1) \quad t = 1, \dots, T \quad (2)$$

where h_t represents log-volatility and the $(\epsilon_t, u_t) \sim Niid(0, I)$ Gaussian iid process with zero mean and variance equals to one. We let $\theta = (\mu, \phi, \sigma_u)$ as the parameter vector, with μ be the intercept, ϕ the log-volatility's autocorrelation coefficient and σ_u the log-volatility's standard deviation. First we set the Bayesian hierarchical structure of the model's conditional density functions, $p(y \mid h)$, $p(h \mid \theta)$ and $p(\theta)$ where $\mathbf{y} = (y_1, \dots, y_T)$ and $\mathbf{h} = (h_1, \dots, h_T)$ the observed and unobserved log-volatility vector. Second we generate estimates from the unobserved h and θ using the conditional density engine of Markov Chain Monte Carlo method.

2.2 The fat-tailed SV model

The existence of fat tailness, widely documented in conditional variance literature (Geweke (1994), Gallant et al. (1997)), has been considered either as an outlier approach using mixtures of normals (Shephard (1994)) or as a purely t-distribution representation (Shephard and Pitt (1997), Jacquier et al. (2004)). In the later case, the fat tailed model takes the form

$$y_t = e^{h_t/2} \cdot \epsilon_t \equiv e^{h_t/2} \cdot \sqrt{\lambda_t} z_t, \quad h_t = \mu + \phi(h_{t-1} - \mu) + \sigma_u \cdot u_t \quad t = 1, \dots, T \quad (3)$$

$$z_t \sim Niid(0, 1), \quad \lambda_t \sim IG(k/2, k/2) \quad (4)$$

where λ_t is an *i.i.d* Inverse Gamma process which implies that $\epsilon_t = \sqrt{\lambda_t} z_t \sim t_k$ is a t-Student random process with k degrees of freedom. Here the parameter vector becomes equal to (θ, λ_t, k) . Having h as the sufficient statistics for the θ parameter vector the posterior densities for it is not affected by the introduction of fat tailness. Having independent λ_t along the observed index, its joint density function will be a product of its marginal ones. Using an Inverse Gamma prior for the λ_t parameter value we can generate posterior densities from the conjugate family. Details of this posterior density implementation of $\lambda = (\lambda_1, \dots, \lambda_T)$ is fully demonstrated in ¶2.6.

2.3 The noncentral t-distributed SV model

The co-existence of fat tailness and asymmetry can be considered using the noncentral t-distribution (Johnson et al. (1995)).

$$y_t = e^{h_t/2} \cdot \epsilon_t \equiv e^{h_t/2} \cdot \sqrt{\lambda_t} (z_t + \delta), \quad h_t = \mu + \phi(h_{t-1} - \mu) + \sigma_u \cdot u_t \quad t = 1, \dots, T \quad (5)$$

$$z_t \sim Niid(0, 1), \quad \lambda_t \sim IG(k/2, k/2) \quad (6)$$

where λ_t is an *i.i.d* Inverse Gamma process which implies that $\epsilon_t = \sqrt{\lambda_t}(z_t + \delta) \sim t_k$ is a noncentral t-distribution random process with k degrees of freedom and skewness parameter $\delta \in \mathfrak{R}$. Here the parameter vector becomes equal to $(\theta, \lambda_t, k, \delta)$.

2.4 Priors

We assume a flat Inverse Gamma (IG) prior for the log-volatility's variance σ_u with $v_o = 1$ degrees of freedom and a reasonably small sum of squares of $s = 0.005$ such as to secure a sparse random draw. The above prior will then result in an IG posterior where sampling is then straightforward. Also, for the μ parameter a flat Gaussian prior, such as $\mu \sim N(0, 100)$ together with the Gaussian u_t assumption will generate posterior density from a Gaussian process.

In terms of h_t 's autoregressive coefficient ϕ , here we can use both a Gaussian flat prior, truncated in the rang of $(-1, +1)$, denoted as TrN(0,100) and a Beta one with prior parameters 20 and 1.5. In the later case, letting $\phi = 2\phi^* - 1$, with ϕ^* distributed as Beta with parameters (ϕ_1, ϕ_2) Shephard and Pitt (Shephard and Pitt (1999)) specify a prior density of $p(\phi) \propto \{.5(1 + \phi)\}^{\phi_1 - 1} \cdot \{.5(1 - \phi)\}^{\phi_2 - 1}$ which supports a ϕ draw within the $(-1, +1)$ range and with prior mean of $\{2\phi_1/(\phi_1 + \phi_2) - 1\}$. This ϕ treatment is due to the fact that we intend to guarantee stationary conditions for the log-volatility process, although as Jacquier et al. (2004) note non-stationarity in stochastic volatility is to be seen as unrealistic since it implies that portfolio managers should adjust long-term option values after each volatility shock.

2.5 The Metropolis-Hastings Algorithm

To sample from the multivariate non-Gaussian random vector \mathbf{h} we employ a Monte Carlo Markov Chain sampler such as the Metropolis-Hastings. We aim to simulate the T -dimensional distribution $\pi^*(\mathbf{h})$, $\mathbf{h} \in H \subseteq \mathfrak{R}^T$ that has density $\pi(\mathbf{h})$ with respect to some dominating measure. To define the algorithm, let $q(\mathbf{h}, \mathbf{h}')$ denote a candidate density for a candidate draw \mathbf{h}' given the current value \mathbf{h} in the sampled sequence. The density $q(\mathbf{h}, \mathbf{h}')$ is referred to as the proposal or candidate density function. Then, the M-H algorithm is defined by two steps: a first step in which a proposal value is drawn from the candidate density and a second step in which the proposal value is accepted as the next iterate in the Markov Chain according to the probability $\alpha(\mathbf{h}, \mathbf{h}')$,

where

$$\alpha(\mathbf{h}, \mathbf{h}') = \begin{cases} \min \left[\frac{\pi(\mathbf{h}')q(\mathbf{h}', \mathbf{h})}{\pi(\mathbf{h})q(\mathbf{h}, \mathbf{h}')}, 1 \right] & \text{if } \pi(\mathbf{h})q(\mathbf{h}, \mathbf{h}') > 0 ; \\ 1 & \text{otherwise .} \end{cases} \quad (7)$$

If the proposal value is rejected, then the next sampled value is taken to be the current value.

2.6 The full Algorithm

At this stage, we will describe the MCMC algorithm that will implement model estimation in the three competing models. The parameter vector θ in the basic SV model takes the values (μ, ψ, σ_u) . Then, in the its augmented form for the t-SVM and the noncentral t-SVM becomes equal to $(\mu, \psi, \sigma_u, \lambda, k)$ and $(\mu, \psi, \sigma_u, \lambda, k, \delta)$ respectively. To simulate the full posterior density function expressed by $p(\mathbf{h}, \theta \mid \mathbf{y})$, $p(\mathbf{h}, \theta, \lambda, k \mid \mathbf{y})$ and, $p(\mathbf{h}, \theta, \lambda, k, \delta \mid \mathbf{y})$, we need to sample from the full conditional densities in each SV model case. Therefore, for the basic SV model we simulate from the $p(\theta \mid \mathbf{h}, \mathbf{y})$ and, $p(\mathbf{h} \mid \theta, \mathbf{y})$, for the t-SVM from the $p(\theta \mid \mathbf{h}, \mathbf{y}, \lambda, k)$, $p(\lambda \mid \theta, \mathbf{h}, \mathbf{y}, \lambda, v)$, $p(\mathbf{h} \mid \theta, \mathbf{y}, \lambda, v)$, and $p(v \mid \theta, \mathbf{h}, \mathbf{y}, \lambda, v)$ and for the noncentral-t SV model the $p(\theta \mid \mathbf{h}, \mathbf{y}, \lambda, k, \delta)$, $p(\lambda \mid \theta, \mathbf{h}, \mathbf{y}, \lambda, v, \delta)$, $p(\mathbf{h} \mid \theta, \mathbf{y}, \lambda, v, \delta)$, and $p(k \mid \theta, \mathbf{h}, \mathbf{y}, \lambda, \delta)$, and $p(\delta \mid \theta, \mathbf{h}, \mathbf{y}, \lambda, k)$.

- $p(\mathbf{h} \mid \theta, \mathbf{y})$: Applying a Gaussian prior for the \mathbf{h} process, we can not get a posterior draw from the conjugate family. For this reason, we apply the M-H algorithm within the MCMC engine. The candidate density for the simulation draws is a Gaussian random-walk based on mean and variance generated from the Laplace density approximation (see Appendix).
- $p(\mu \mid \theta_{-\mu}, \mathbf{h}, \mathbf{y})$: Applying a uniform prior over \Re , we can generate a Gaussian full conditional density for the μ parameter, such as

$$p(\mu \mid \theta_{-\mu}, \mathbf{h}, \mathbf{y}) \sim N(\hat{\mu}, \sigma_\mu^2)$$

with mean

$$\hat{\mu} = \frac{\sigma_\mu^2}{\sigma_u^2} \{ (1 - \phi^2)h_1 + (1 - \phi) \sum_{t=2}^T (h_t - \phi h_{t-1}) \}$$

and variance $\sigma_\mu^2 = \sigma_u^2 \{ (T - 1)(1 - \phi)^2 + (1 - \phi^2) \}^{-1}$.

- $p(\phi \mid \theta_{-\phi}, \mathbf{h}, \mathbf{y})$: Applying the beta prior analysed in Section 2.2, we can guarantee the stationarity assumption. However, since this doesn't allow us a conjugate match with the Gaussian h process we apply the M-H algorithm within the MCMC engine. The candidate

density for the simulation draws is a Gaussian random-walk based on mean and variance least square components, such that

$$p(\phi \mid \theta_{-\phi}, \mathbf{h}, \mathbf{y}) \sim N(\hat{\phi}, \sigma_{\phi}^2)$$

with mean

$$\hat{\phi} = \frac{\sum_{t=2}^T (h_t - \mu)(h_{t-1} - \mu)}{\sum_{t=2}^T (h_{t-1} - \mu)^2}$$

and variance $\sigma_{\phi}^2 = \sigma_u^2 \{\sum_{t=1}^T (h_t - \mu)^2\}$.

- $p(\sigma_u^2 \mid \theta_{-\sigma_u^2}, \mathbf{h}, \mathbf{y})$: Setting a conjugate Inverse Gamma (IG) prior, such that $\sigma_u^2 \sim IG(\sigma_r/2, S_r/2)$, with $\sigma_r = 5$ and $S_r = .01 \cdot \sigma_r$, than the posterior for σ_u^2 becomes:

$$\sigma_u^2 \mid \theta_{-\sigma_u^2}, \mathbf{h}, \mathbf{y} \sim IG\left\{\frac{T + \sigma_r}{2}, \frac{S_r + (h_1 - \mu)^2(1 - \phi^2) + \sum_{t=1}^T (h_t - \mu - \phi(h_{t-1} - \mu))^2}{2}\right\}$$

- $p(\lambda_t \mid \theta, \mathbf{h}, \mathbf{y})$: Setting a conjugate IG prior, such that $\lambda_t \sim IG(k/2, 2/k)$, than the posterior for λ_t becomes:

$$p(\lambda_t \mid \theta_{-\sigma_u^2}, \mathbf{h}, \mathbf{y}) \propto \frac{1}{\lambda_t^{\frac{1+k}{2}+1}} \exp\left\{-\frac{y_t/\sigma_t^2 + k}{2\lambda_t}\right\} \sim IG\left(\frac{k+1}{2}, \frac{2}{y_t^2/\sigma_t^2 + k}\right)$$

for the t-SV model, and

$$\lambda_t \mid \theta, \mathbf{h}, \mathbf{y} \sim IG\left(\frac{k+1}{2}, \frac{2}{(\frac{y_t}{\sigma_t} - \delta)^2 + k}\right)$$

for the noncentral t-SV one.

- $p(k \mid \theta, \mathbf{h}, \mathbf{y}, \lambda)$: Setting a $p(k)$ as any conjugate prior, we have $p(k \mid \theta, \mathbf{h}, \mathbf{y}, \lambda) \sim p(k \mid \lambda)$ since λ is a sufficient statistic for the parameter k . Thus, we can get posterior draws from

$$p(k \mid \lambda) \propto p(k) \prod_{t=1}^T p(\lambda_t \mid k) \propto \left(\frac{k^{k/2}}{\Gamma(k/2)}\right)^T \exp\left\{-\frac{k}{2} \sum_{t=1}^T \left(\frac{1}{2} + \log \lambda_t\right)\right\}$$

- $p(\delta \mid \theta, \mathbf{h}, \mathbf{y}, \lambda, k)$: Setting a $p(\delta) \sim N(0, \sigma_{\delta}^2)$, we can derive posterior draws from a conditional Gaussian process, such that

$$p(\delta \mid \theta, \mathbf{h}, \mathbf{y}, \lambda, k) \propto \exp\left\{-\frac{1}{2} \frac{\delta^2}{\sigma_{\delta}^2}\right\} \cdot \exp\left\{-\frac{1}{2} \left(\frac{y_t}{\sigma_t \sqrt{\lambda}} - \delta\right)^2\right\}$$

$$\delta \mid \theta, \mathbf{h}, \mathbf{y}, \lambda, k \sim N\left\{\left(1 + 1/\sigma_{\delta}^2\right)^{-1} \cdot \left(\frac{y_t}{\sigma_t \sqrt{\lambda}}\right), \left(1 + 1/\sigma_{\delta}^2\right)^{-1}\right\}.$$

3 Empirical results

To illustrate the new stochastic volatility model, we focus on an example involving real data series that demonstrate strong second order dependency. We consider daily exchange rate data for the British Pound (GBP) against U.S. Dollar (USD). The data series cover the period starting from the 28th of June 1985 and ending the 28th of April 1989. Before we proceed with the model's estimation, we need to demonstrate some stylised statistical properties of the analysed series. Table 1 demonstrates the mean, standard deviation, minimum, maximum, skewness and kurtosis coefficients of the logarithmically transformed return series. As far as Gaussian assumptions are concerned, the return exchange rate series demonstrate a considerable amount of asymmetry where excess kurtosis is marginally higher than the 3 measure of the Gaussian process. This gives as the proxy for analysing to what extend this normality violation can be seen as significant so as to incorporate the skewness and kurtosis assumption in the "standard" SV specification.

Simulation results for the three competing models are demonstrated in Table 2. These represent the last 40,000 of the total 60,000 iterations using the MCMC engine. These demonstrate the mean, standard deviation, median and the 5% confidence interval for each estimator in the three SV models. The results demonstrate high significance for all the estimated parameters in each model. Additionally, due to the right prior selection for the autoregressive coefficient ϕ the posterior mean as well as its confidence interval is away from the non-stationary assumption. Concerning the level of symmetry in the posterior estimates, the median seem to coincide with the mean estimates in most of the cases. Finally, the measures of normality violations expressed by δ and κ , seem to show a systematic deviation from normality. More specifically, in the t-SV model the confidence interval of the kurtosis parameter κ show a considerable normality violation for the analysed data. However, the level of the upper and lower confidence interval bound is such that may penalise the specification at the model selection stage of inference. In the noncentral-t SV model, the kurtosis parameter κ as well as the skewness parameter δ demonstrate data's normality violation but this time with much improved confidence interval bound compared with the t-SV one's.

Figure 2 demonstrate the histogram of the posterior simulation results in the noncentral-t SV model for the parameters, δ , k , μ , ϕ and τ which is the square root of σ_u^2 .

3.1 Model selection

As a model choice criterion we will adopt the Deviance Information Criterion (DIC). This criterion is based on the posterior deviance statistic

$$D(\theta) = -2\log f(y \mid \theta) + 2f(y)$$

where $f(y \mid \theta)$ stands for the likelihood function and $f(y)$ for the standardising term. Spiegelhalter et al. (2002) propose that this deviance statistic apart from the goodness of fit measure and to the analogy of the “classical” Akaike Information Criterion, should have a penalising term for the possible model complexity increase. Thus, the authors propose for the fitness measure the posterior expectation of the deviance, $\bar{D} = E_{\theta|y}[D]$, and for that of the penalising term the posterior mean of the parameters, such as

$$p_D = E_{\theta|y}[D] - D(E_{\theta|y}[\theta]) = \bar{D} - D(\bar{\theta})$$

therefore the DIC is defined as

$$DIC = \bar{D} - p_D = 2\bar{D} - D(\bar{\theta}) \quad (8)$$

with smaller values of DIC indicating a better-fitting model.

In the context of the SV model choice, the DIC estimator has also been used in the past (Berg et al. (2004)). They have argued that traditional Bayesian model selection criteria, such as the Bayes factor, the familiar BIC and the penalised likelihood ratio model choice criterion AIC, suffer from their dependence to the number of parameters used. In hierarchical Bayesian models such as the SV one, the number of unknowns outnumbers the number of observations, something that makes the model choice issue a very complex one.

In our model comparison case, we demonstrate the estimator’s results in Table 3 for the three competing models. These show that the proposed noncentral-t SV model is favoured against the other two SV models as it manages to minimise the DIC estimator. Additionally, the t-SVM model seems to fail in its comparison to the “standard” SVM as it demonstrates a very small p_D value.

3.2 Sensitivity issues

Having selected the noncentral-t SV model at the best performed model, we can now make a sensitivity “testing” for this model. This will refer to the choice of different priors assumptions

for the skewness parameter δ .

First, we work on the same type of prior assumption, the Gaussian one, and instead of unity variance we increase it to 100. Second, we change the prior assumption to the Uniform distribution within the range of values $(-10, 10)$. Third, we increase the range of values the Uniform distribution prior distribution to $(-100, 100)$. Our intention is to show that for large variance values for the prior of δ , such as 100, 33.33 and, 833.33 for the three consecutive assumptions, we can get robust posterior density results.

Table 4, demonstrates the results from this sensitivity experiment. It shows that in all three cases, the posterior means of the estimated parameter in the noncentral-t-SV model show a considerable robustness. The parameters' significance level as well as the confidence interval is not affected in a way that could affect the model selection stage. However, we can easily observe than the Uniform distribution prior choice for the δ parameter has slightly affected the posterior variance of the k parameter, as it has increased the level of its statistical significance.

4 Conclusion

In this paper we have seen how we can incorporate both log-tailed and asymmetric frequencies in the standard SV model by the use of the noncentral t-distribution. The full Bayesian estimation framework has been worked out using both the Gibbs sampler and the Metropolis-Hastings algorithm when conditionally conjugacy was not available. An empirical investigation is then displayed focusing on the model selection issues among three competing SV models. The standard SV one, the t-SVM, and the noncentral-t SV one. Results, based on the DIC estimator shows that the noncentral-t SV model, managing to reveal data's normality violations, out-performs the other two specifications.

Future work need to be focused on the comparison, apart the within sample, the forecasting performance of the present noncentral-t SVM against its main competitors. Also, a direction towards comparing the noncentral t-distribution with other asymmetric ones such as the Skewed Normal or the Skewed t-distribution (Azzallini (1985)) within the SV framework can also be an interesting task.

References

- Azzallini, A. (1985) A class of distributions which includes the normal one. *Scandinavian Journal of Statistics* 12, 171-178.
- Berg, A., Meyer, R. Yu, J. (2004) The DIC as a model comparison criterion for stochastic volatility models. *Journal of Business and Economics Statistics*, 22, 107-120.
- Fernandez, C., Steel, M.F.J. (1988) On bayesian modelling of fat tails and skewness. *Journal of the American Statistical Association* 93, 359-371.
- 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.
- Geweke, J. (1994) Comments on bayesian analysis of stochastic volatility. *Journal of Business and Economics Statistics*, 12, 371-417.
- Harvey, A.C., Ruiz, E. (1993) Multivariate stochastic volatility models. *Review of Economics Studies* 61, 247-264.
- Harvey, C. and Siddique, A. (1999) Autoregressive conditional skewness. *Journal of Finance and Quantitative Analysis*, 34, 465-487.
- Harvey, C. and Siddique, A. (2000) Conditional skewness in asset pricing tests. *Journal of Finance*, 55, 1263-1296.
- Higgins, M.L. and Bera, A. (1992) A class of nonlinear ARCH models. *International Economic Review* 62, 137-158.
- Hull, J., White, A. (1987) The pricing of options on assets with stochastic volatility. *Journal of Finance* 3, 281-300.
- Johnson, N.L., Kotz, S., Balakrishnan, N. (1995) *Continuous Univariate Distributions, Vol. 2* 2nd Edition NY: Wiley.

- Jacquier, E., Polson, N.G. and Rossi, P.E. (1994) Bayesian analysis of stochastic volatility models (with discussion). *Journal of Business and Economics Statistics*, 12, 371-417.
- 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, 185-212.
- 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.
- Shephard, N. (1994) Partial non-Gaussian state space. *Biometrika* 81, 115-131.
- Shephard, N., Pitt, M.K. (1997) Likelihood analysis of non-Gaussian measurement time series. *Biometrika* 84, 653-667.
- Spiegelhalter, D.J., Best, N., Carlin, B.P., van der Linde, A. (2002) Bayesian measurements of model flexibility and fit (with discussion). *Journal of Royal Statistical Society B*, 64, 583-639.
- Tsionas, E. F. (2002) Bayesian inference in the noncentral student-t model. *Journal of Computational and Graphical Statistics* 11, 208-221.
- Tsiotas, G. (2002) Nonlinearities in Stochastic Volatility Models. Ph.D thesis, University of Essex.
- Tsiotas, G. (2007) On the use of the Box-Cox transformation on conditional variance models. *Finance Research Letters* (forthcoming).

Appendix

Using the Laplace method, h_t 's posterior density is approximated around its mode. Here, we take the case of approximating the log-density function $\forall t \in \{1, \dots, T\}$. Therefore, the conditional density function for the basic SV model, being the exponential function of the log-density one it becomes:

$$\begin{aligned} p(h_t | y^{t-1}) &= e^{\ell(h_t | h_{t-1})} \approx e^{\ell(\theta | \epsilon_t, h_t) + \frac{\partial \ell}{\partial h_t}(h_t - \hat{h}_{t|t-1}) + \frac{\partial^2 \ell}{\partial h_t^2} \frac{1}{2}(h_t - \hat{h}_{t|t-1})^2} \\ &\approx \exp\left\{-\frac{(h_t - \mu(h_{t-1}))^2}{2\sigma_v^2} - \frac{y_t^2}{2} \exp\{-\hat{h}_t\}[1 + (h_t - \hat{h}_t)^2]\right\} \\ &\propto \exp\left\{-\frac{(h_t - \mu_h)^2}{2\sigma_h^2}\right\} \end{aligned}$$

Thus, the posterior density function of log-volatility, h_t , is approximately distributed as a normal with a mean of $\mu_h = \sigma_h^2(\mu(h_{t-1})v_t^{-1} + \hat{h}_t u_t^{-1})$ and a variance of $\sigma_h^2 = (v_t^{-1} + u_t^{-1})^{-1}$, i.e.:

$$h_t \sim N(\mu_h, \sigma_h^2)$$

where $v_t^{-1} = \sigma_v^{-2}$ and $u_t^{-1} = y_t^2 \exp\{-\hat{h}_t\}$ respectively denote the variances of the prior density function and that of the likelihood function.

For the t-SV model, the $u_t^{-1} = y_t^2 \lambda^{-1} \exp\{-\hat{h}_t\}$ and in the noncentral-t SV model the $u_t^{-1} = (y_t \lambda^{-1/2} \exp\{\frac{1}{2}\hat{h}_t\} - \delta)^2$.

Table 1: Summary statistics for GBP/USD return series

	<i>Data series</i>
	GBP/USD
Mean	0.02964
Standard deviation	0.2900
Minimum	-0.9877
Maximum	1.66900
Skewness	0.2332166
Kurtosis	4.696187

Table 2: MCMC results in SV models using Great Britain Pound against U.S. dollars daily exchange rates data

models	estimators	<i>MCMC results</i>				
		mean	s.d.	2.5% CI	median	97.5% CI
SV						
	μ	-2.558	0.1765	-2.873	-2.571	-2.175
	ϕ	0.9795	0.01175	0.9492	0.9822	0.995
	τ	0.102	0.02742	0.06517	0.09601	0.1724
t-SV						
	k	10.88	2.416	7.193	10.49	16.63
	μ	-2.711	0.1812	-3.049	-2.721	-2.291
	ϕ	0.9802	0.01039	0.9548	0.982	0.9954
	τ	0.08886	0.01817	0.05668	0.08902	0.1288
noncentral-t SV						
	δ	0.111	0.03261	0.0474	0.111	0.1753
	k	10.3	2.747	6.457	9.828	17.42
	μ	-2.755	0.1899	-3.112	-2.763	-2.337
	ϕ	0.9805	0.01018	0.9561	0.982	0.9955
	τ	0.09876	0.02146	0.06611	0.09581	0.1492

Table 3: Deviance Estimators for the SV models using daily GBP/USD exchange rates data.

Models	<i>Estimators</i>			
	$\bar{D}(\theta)$	$D(\hat{\theta})$	DIC	p
SV	239.7	208.6	270.8	31.06
t-SV	258.7	241.5	276.0	17.27
noncentral-t SV	142.9	31.39	254.4	111.5

Table 4: Sensitivity MCMC results in the noncentral-t SV model using Great Britain Pound against U.S. dollars daily exchange rates data

		<i>MCMC results</i>				
δ Priors	estimators	mean	s.d.	2.5% CI	median	97.5% CI
N(0,100)						
	δ	0.1134	0.03276	0.04963	0.1136	0.1774
	k	10.88	2.822	6.495	10.54	19.38
	μ	-2.714	0.1879	-3.074	-2.717	-2.316
	ϕ	0.9817	0.007558	0.9648	0.9823	0.9946
	τ	0.1026	0.01642	0.07396	0.1011	0.1384
U(-10,10)						
	δ	0.1088	0.03243	0.04555	0.1086	0.1722
	k	10.18	1.773	7.117	10.03	13.9
	μ	-2.734	0.1985	-3.133	-2.737	-2.339
	ϕ	0.988	0.00562	0.9758	0.9887	0.9972
	τ	0.07359	0.009611	0.05755	0.07331	0.09318
U(-100,100)						
	δ	0.1112	0.03254	0.04736	0.1111	0.1745
	k	9.633	1.769	6.507	9.509	13.64
	μ	-2.749	0.1829	-3.108	-2.754	-2.368
	ϕ	0.9857	0.006832	0.9702	0.9866	0.9967
	τ	0.08267	0.01842	0.05713	0.07893	0.1262

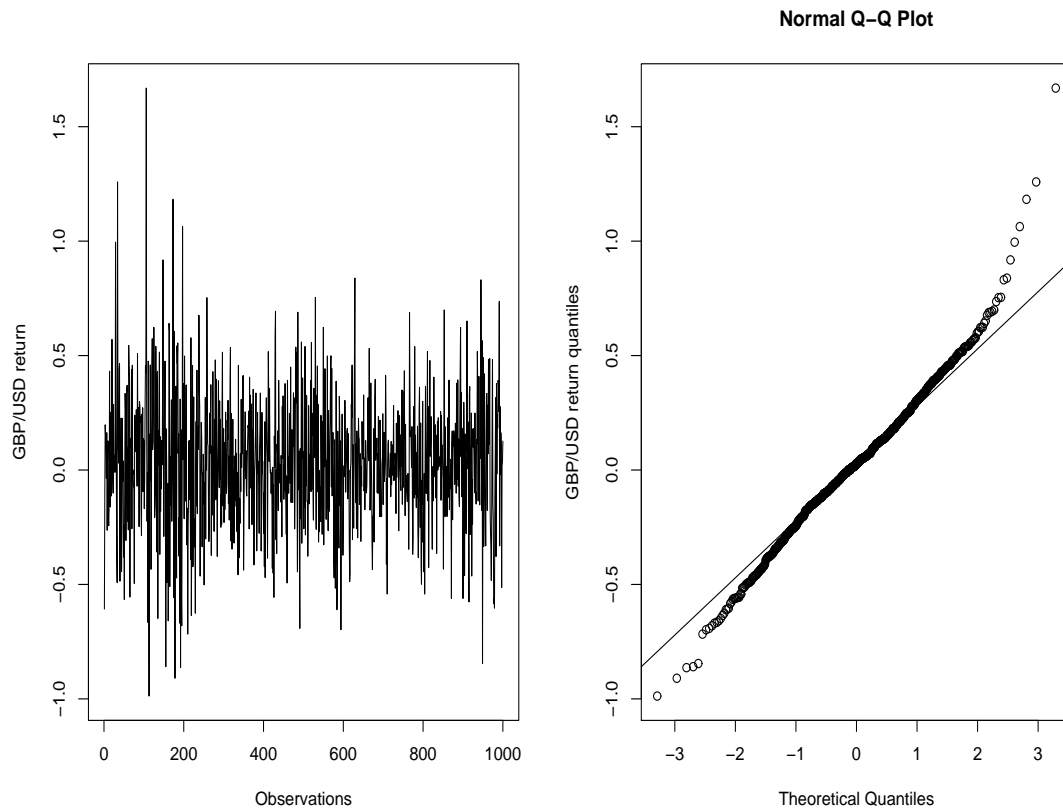


Figure 1: Plot and empirical quantile of the *GBP/USD* data series.

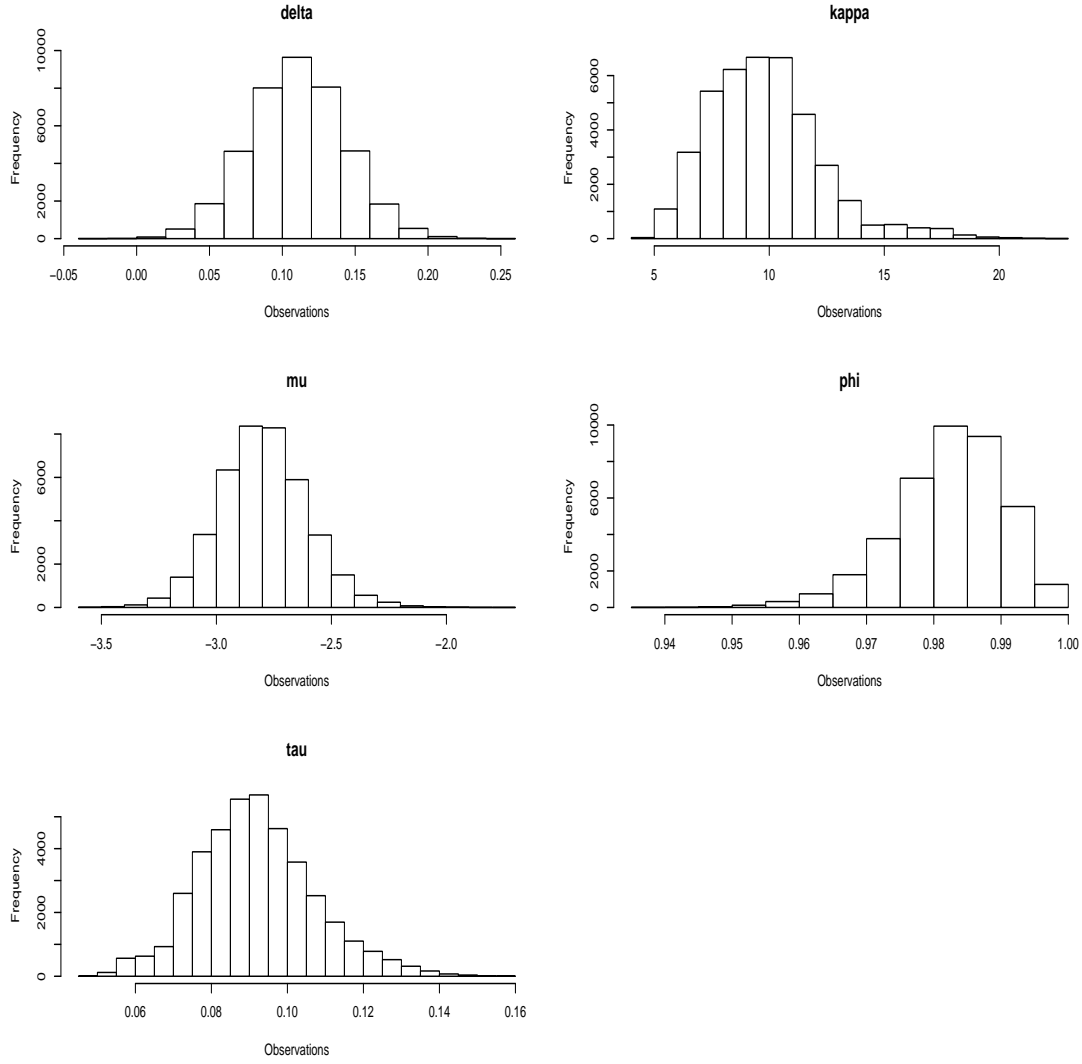


Figure 2: Histograms of the posterior distributions derived using from the last 40,000 iterations. These represent, from the top and from left to the right, the parameters δ , k , μ , ϕ and τ which is the square root of σ_u^2 .

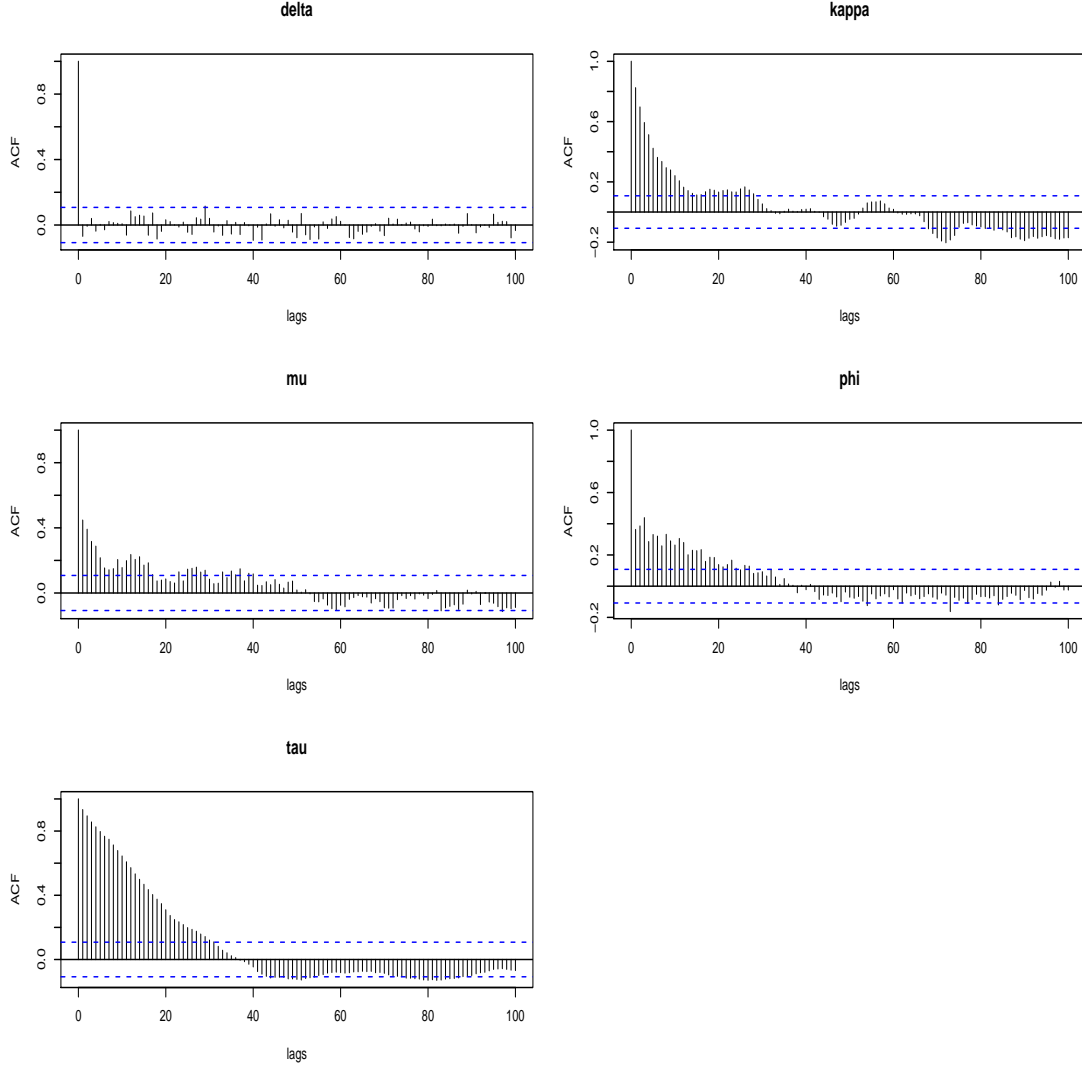


Figure 3: Autocorrelation functions derived using from the last 40,000 iterations. These represent, from the top and from left to the right, the parameters δ , k , μ , ϕ and τ which is the square root of σ_u^2 .