



UPPSALA
UNIVERSITET

U.U.D.M. Project Report 2009:1

Bayesian Analysis of a Stochastic Volatility Model

Yu Meng

Examensarbete i matematik, 30 hp
Handledare och examinator: Johan Tysk
Februari 2009



Department of Mathematics
Uppsala University

1 Introduction

The stochastic volatility (*SV*) model introduced by and Taylor (1982) provides an alternative to the ARCH-type models of Engle (1982). The *SV* model is more realistic and flexible than the ARCH-type models, since it essentially involves two random processes, one for the observations, and one for the latent volatilities. The model is given by:

$$y_t = \exp(h_t/2)u_t, \quad u_t \stackrel{i.i.d.}{\sim} N(0, 1), \quad t = 1, \dots, T, \quad (1)$$

$$h_{t+1} = \mu + \phi(h_t - \mu) + \sigma_\eta \eta_{t+1}, \quad \eta_t \stackrel{i.i.d.}{\sim} N(0, 1), \quad t = 1, \dots, T, \quad (2)$$

where y_t is the observation at time t , $h_1 \sim N(\mu, \frac{\sigma_\eta^2}{1-\phi^2})$ and $N(a, b)$ denotes the normal distribution with mean a and variance b . The log-volatility h_t is assumed to follow a stationary AR(1) process with the persistent parameter $|\phi| < 1$. The observation error u_t captures the measurement and sampling errors, whereas the process error η_t assesses the variation in the underlying volatility dynamics.

The parameter estimation of the *SV* model is not easy due to the intractable form of the likelihood $p(y|\theta) = \int p(y|h, \theta)p(h|\theta)dh$, where $y = (y_1, \dots, y_T)$, $h = (h_1, \dots, h_T)$ is the vector of latent volatilities and $\theta = (\mu, \phi, \sigma_\eta^2)$ is the set of parameters. The likelihood is a T -dimensional integration with respect to the unknown latent volatilities and its analytical form, in general, is unknown. Several estimation methods have been proposed, including generalized method of moments, quasi-maximum likelihood, efficient method of moments and simulated maximum likelihood. In Bayesian context, Markov chain Monte Carlo (MCMC) technique has been suggested by Jacquier, Polson, and Rossi (1994) and Kim, Shephard, and Chib (1998). In a comparative study of estimation methods, Andersen, Chung, and Sorensen (1999) showed that Markov Chain Monte Carlo is the most efficient estimation technique for the *SV* model.

In this paper, we give an empirical framework for the estimation of the *SV* model using MCMC technique. This includes a detailed sampling proce-

dure, volatility filtering, convergence diagnostics and a misspecification test. The data used in this study is China's stock index. Its characteristics may not be well documented up to date. Our results show that this data exhibits some stylized facts of stock returns, such as volatility clustering and excess kurtosis.

The remainder of this paper is organized as follows. In Section 2 we give a short overview of Bayesian theory and MCMC, specifically the focus is on the Gibbs sampler (Geman and Geman (1984)). Section 3 gives the method for misspecification test. Section 4 describes the data and prior specifications. In Section 5 we report the empirical results, and Section 6 concludes our study.

2 Bayesian inference and MCMC

2.1 Bayesian inference

In a Bayesian context, we are interested in the conditional joint density $p(\theta, h|y)$, called the posterior distribution. It summarizes all information in the observations and in the model, and it also provides the basis for inference and decision making. The Bayes rule factors the posterior distribution into its constituent components:

$$p(\theta, h|y) \propto p(y|\theta, h)p(h|\theta)p(\theta),$$

where, $p(\cdot)$ denotes the probability density function, \propto stands for proportion, $p(y|\theta, h)$ is the full-information (or data-augmented) likelihood function, $p(h|\theta)$ is the conditional distribution of state variables and $p(\theta)$ is the joint distribution of model parameters, commonly called the prior.¹

¹The density version of Bayes's theorem is $p(\theta, h|y) = \frac{p(y|\theta, h)p(\theta, h)}{p(y)}$. The shape of the posterior density $p(\theta, h|y)$ is irrelevant to the predictive density $p(y) = \int p(y|\theta, h)p(\theta, h)d\theta dh$, since, for fixed y , $p(y)$ does not depend on θ and h and can be viewed as the integrating constant.

The prior distribution allows us to incorporate any information into the parameters of interest, prior to the observation y . For example, we can give a positivity constraint on some parameter of interest to make it economically meaningful. Specifically, for the *SV* model we can use an uniform prior $\text{Unif}(0, 1)$ on the persistent parameter ϕ to rule out the near unit-root behavior.

The joint posterior density $p(\theta, h|y)$ combines all information in the model and the observations y . It can be used to estimate model parameters. For example, if we can directly generate a sequence of i.i.d. random variates $\{\theta^{(g)}, h^{(g)}\}_{g=1}^G$ from the joint posterior distribution $p(\theta, h|y)$, then the sample mean of this i.i.d. sequence is the consistent estimate of the model parameter θ and the state variable h . Unfortunately, direct independent sampling from the joint posterior density $p(\theta, h|y)$ is not possible, in general, due to the penalty of high-dimensional integration. Note that the joint posterior density $p(\theta, h|y)$ depends on the predictive density $p(y) = \int p(y|\theta, h)p(\theta, h)d\theta dh$. For the *SV* model, calculating the predictive density involves high-dimensional integration.

2.2 The Gibbs sampler

As we mentioned in Section 2.1, the joint posterior $p(\theta, h|y)$ is an extremely complicated, high-dimensional distribution and directly generating samples from this distribution is prohibited. MCMC attacks the curse of dimensionality by breaking the joint posterior $p(\theta, h|y)$ into its complete set of conditional distributions, from which samples can be easily simulated. Specifically, for the single-move Gibbs sampler, which is employed in this study, the complete set of full conditional distribution (sometimes called full conditional posteriors) are:

$$p(h_t|y, h_{-t}, \mu, \phi, \sigma_\eta^2), \quad t = 1, \dots, T, \quad (3)$$

$$p(\sigma_\eta^2|y, h, \mu, \phi), \quad p(\phi|y, h, \sigma_\eta^2, \mu), \quad p(\mu|y, h, \phi, \sigma_\eta^2), \quad (4)$$

where h_{-t} denotes the elements of $h = (h_1, \dots, h_T)$ excluding h_t . These full conditional posteriors, based on Clifford-Hammersley theorem, uniquely determine the joint posterior $p(\theta, h|y)$.² In other words, knowledge of these full conditional posteriors is equivalent to knowledge of the joint posterior, up to a constant of proportionality.

The Gibbs sampler provides a way to construct a Markov Chain by recursively sampling from the above full conditional posteriors. In this study we employ the single-move Gibbs sampler, which has been studied, for example, by Jacquier, Polson, and Rossi (1994) and Kim, Shephard, and Chib (1998). The algorithm is given by:

1. Initialize $h^{(0)}, \mu^{(0)}, \phi^{(0)}$ and $\sigma_\eta^{2(0)}$.
2. For $g = 1, \dots, G$,
 - For $t = 1, \dots, T$,

Sample $h_t^{(g)}$ from $p(h_t|y, h_{<t}^{(g)}, h_{>t}^{(g-1)}, \mu^{(g-1)}, \phi^{(g-1)}, \sigma_\eta^{2(g-1)})$.
 - Sample $\sigma_\eta^{2(g)}$ from $p(\sigma_\eta^2|y, h^{(g)}, \mu^{(g-1)}, \phi^{(g-1)})$.
 - Sample $\phi^{(g)}$ from $p(\phi|y, h^{(g)}, \sigma_\eta^{2(g)}, \mu^{(g-1)})$.
 - Sample $\mu^{(g)}$ from $p(\mu|y, h^{(g)}, \phi^{(g)}, \sigma_\eta^{2(g)})$.

where the subscribes of $h_{<t}^{(\cdot)}$ and $h_{>t}^{(\cdot)}$ denote the date before and after t , respectively. A sequence of dependent random variates $\{\theta^{(g)}, h^{(g)}\}_{g=1}^G$ can then be generated from the Gibbs sampler, which is Markov (the next state only depends on current state) and the chain is characterized by both its starting value $\{\theta^{(0)}, h^{(0)}\}$ and its transition kernel $P(\{\theta^{(g+1)}, h^{(g+1)}\}|\{\theta^{(g)}, h^{(g)}\})$. This Markov chain, under mild regularity conditions, *converges to its equilibrium distribution* (Ergodic Theorem, for proofs of the convergence of Gibbs

²Clifford-Hammersley theorem states that the joint posterior $p(\theta, h|y)$ can be uniquely determined by its complete set of conditionals $p(\theta|y, h)$ and $p(h|\theta, y)$. By successively applying Clifford-Hammersley theorem on $p(\theta|y, h)$ and $p(h|\theta, y)$, we can get the full conditional posteriors.

sampler, see Tierney (1994)) and *this unique equilibrium distribution is just the joint posterior* $p(\theta, h|y)$ (Clifford-Hammersley theorem, under positivity condition). To see the latter point, notice that the transition kernel of Gibbs sampler used for generating the Markov Chain is just the product of all the full conditional posteriors, that uniquely determines the joint posterior $p(\theta, h|y)$. The g -step transition kernel will converge (the convergence of Gibbs sampler), as $g \rightarrow \infty$, to the *unique* equilibrium distribution.³ Hence the convergence is equivalent to the joint posterior.

Other than the convergence of the Markov Chain, we are typically interested in the convergence of the sample average of $\{\theta^{(g)}, h^{(g)}\}_{g=1}^G$. The Ergodic Theorem guarantees that the sample average will converge to its population counterpart for any initial distribution, regardless of the rate of convergence.

2.2.1 The full conditional posteriors

We now turn to specifying the full conditional posteriors for the basic *SV* model. For the full conditional posteriors for model parameters: Assuming that all priors are independent, Baye's rule implies that

$$\begin{aligned} p(\sigma_\eta^2|y, h, \mu, \phi) &\propto p(h|\mu, \phi, \sigma_\eta^2)p(\sigma_\eta^2), & p(\phi|y, h, \sigma_\eta^2, \mu) &\propto p(h|\mu, \phi, \sigma_\eta^2)p(\phi), \\ p(\mu|y, h, \phi, \sigma_\eta^2) &\propto p(h|\mu, \phi, \sigma_\eta^2)p(\mu), \end{aligned} \quad (5)$$

where $p(\sigma_\eta^2)$, $p(\phi)$ and $p(\mu)$ are the priors. We assume $p(\sigma_\eta^2) \sim \mathcal{IG}(\alpha_\sigma, \beta_\sigma)$, $p(\phi) \sim \mathcal{N}(\alpha_\phi, \beta_\phi^2)I_{(-1, +1)}(\phi)$ and $p(\mu) \sim \mathcal{N}(\alpha_\mu, \beta_\mu^2)$, where $\mathcal{IG}(\cdot, \cdot)$ denotes the inverse-gamma distribution and $\alpha_{(\cdot)}$ and $\beta_{(\cdot)}$ are called hyperparameters. Like the prior distribution, the hyperparameters also allow us to incorporate non-sample information in the model parameters, and should be specified

³The g -step transition probability is, $P^{(g)}(x, A) = \text{Prob}[X^{(g)} \in A | X^{(0)} = x]$, and the equilibrium distribution is defined as, $\lim_{g \rightarrow \infty} \text{Prob}[X^{(g)} \in A | X^{(0)}] = \pi(A)$. The Ergodic Theorem for Markov Chain guarantees the convergence of the distribution of the chain to its equilibrium distribution, regardless the initial distribution.

by researcher prior to observations. We use a truncated normal distribution for parameter ϕ in order to rule out the near unit-root behaviour.

By successively conditioning on $p(h|\mu, \phi, \sigma_\eta^2)$, we can get:

$$p(h|\mu, \phi, \sigma_\eta^2) = p(h_1|\mu, \phi, \sigma_\eta^2) \prod_{t=1}^{T-1} p(h_{t+1}|h_t, \mu, \phi, \sigma_\eta^2). \quad (6)$$

Then we can insert the prior densities $p(\sigma_\eta^2)$, $p(\phi)$, $p(\mu)$ and (6) into (5), after some manipulations, the full conditional posteriors can be reformulated as (for derivations see Appendix A):

$$\begin{aligned} p(\sigma_\eta^2|y, h, \mu, \phi) &\propto \mathcal{IG}(\hat{\alpha}_\sigma, \hat{\beta}_\sigma), \\ \hat{\alpha}_\sigma &= \alpha_\sigma + \frac{T}{2}, \quad \hat{\beta}_\sigma = \beta_\sigma + \frac{1}{2} \left[\sum_{t=1}^{T-1} (h_{t+1} - \mu - \phi(h_t - \mu))^2 + (h_1 - \mu)^2 (1 - \phi^2) \right], \\ p(\phi|y, h, \sigma_\eta^2, \mu) &\propto \mathcal{N}(\hat{\alpha}_\phi, \hat{\beta}_\phi^2) I_{(-1, +1)}(\phi), \\ \hat{\alpha}_\phi &= \hat{\beta}_\phi^2 \left(\frac{\sum_{t=1}^{T-1} (h_{t+1} - \mu)(h_t - \mu)}{\sigma_\eta^2} + \frac{\alpha_\phi}{\beta_\phi^2} \right), \\ \hat{\beta}_\phi^2 &= \left(\frac{\sum_{t=1}^{T-1} (h_t - \mu)^2 - (h_1 - \mu)^2}{\sigma_\eta^2} + \frac{1}{\beta_\phi^2} \right)^{-1}, \\ p(\mu|y, h, \phi, \sigma_\eta^2) &\propto \mathcal{N}(\hat{\alpha}_\mu, \hat{\beta}_\mu^2), \\ \hat{\alpha}_\mu &= \hat{\beta}_\mu^2 \left(\frac{h_1(1 - \phi^2) + (1 - \phi) \sum_{t=1}^{T-1} (h_{t+1} - \phi h_t)}{\sigma_\eta^2} + \frac{\alpha_\mu}{\beta_\mu^2} \right), \\ \hat{\beta}_\mu^2 &= \left(\frac{1 - \phi^2 + (T-1)(1 - \phi)^2}{\sigma_\eta^2} + \frac{1}{\beta_\mu^2} \right)^{-1}. \end{aligned}$$

The samples can then be drawn from their corresponding full conditional posteriors.

The most difficult part of the Gibbs sampler is to effectively sample the latent state h_t from its full conditional posterior. We employ the accept-reject sampling procedure introduced by Kim, Shephard, and Chib (1998). The Baye's theorem implies:

$$p(h_t|y, h_{-t}, \theta) \propto p(y_t|h_t, \theta) p(h_t|h_{-t}, \theta), \quad t = 1, \dots, T,$$

We are interested in the right-hand side of the above equation. First, by employing the state equation (2) for h_t and h_{t+1} we can easily derive:

$$\begin{aligned} p(h_t|h_{-t}, \theta) &= p(h_t|h_{t-1}, h_{t+1}, \theta) = p_N(h_t|\alpha_t, \beta^2), \\ \alpha_t &= \mu + \frac{\phi\{(h_{t-1} - \mu) + (h_{t+1} - \mu)\}}{(1 + \phi^2)}, \quad \beta^2 = \frac{\sigma_\eta^2}{1 + \phi^2}, \end{aligned} \quad (7)$$

where $p_N(x|a, b)$ denotes the normal density function with mean a and variance b .

Second, the density $p(y_t|h_t, \theta)$, by taking the logarithm, can be written as:

$$\log p(y_t|h_t, \theta) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}h_t - \frac{y_t^2}{2}\exp(-h_t) = \text{const} - \log f^*(y_t, h_t, \theta),$$

and, due to the property of convexity, $\exp(-h_t)$ can be bounded by a function linear in h_t , we apply Taylor expansion for $\exp(-h_t)$ around α_t and get:

$$\log f^*(y_t, h_t, \theta) \leq -\frac{1}{2}h_t - \frac{y_t^2}{2}\{\exp(-\alpha_t)(1 + \alpha_t - h_t \exp(-\alpha_t))\} = \log g^*(y_t, h_t, \theta),$$

hence, due to (7) that $p(h_t|h_{-t}, \theta) = p_N(h_t|\alpha_t, \beta^2)$, we have:

$$p(h_t|h_{-t}, \theta)f^*(y_t, h_t, \theta) \leq p_N(h_t|\alpha_t, \beta^2)g^*(y_t, h_t, \theta),$$

the right-hand side of the above equation, after some manipulations, can be shown as:

$$p_N(h_t|\alpha_t, \beta^2)g^*(y_t, h_t, \theta) \propto p_N(h_t|\alpha_t^*, \beta^2), \quad \alpha_t^* = \alpha_t + \frac{\beta^2}{2}(y_t^2 \exp(-\alpha_t) - 1).$$

The accept-reject method can then be applied to sample h_t from $p(h_t|y, h_{-t}, \theta)$ (for details, see Appendix B). The procedure is summarized as follows: For $t = 1, \dots, T$, (1) Drawn h_t^* from $p_N(h_t|\alpha_t^*, \beta^2)$, (2) $U \leq f^*(y_t, h_t^*, \theta)/g^*(y_t, h_t^*, \theta)$, accept and set $h_t = h_t^*$; else, go to step 1.

2.3 Convergence diagnostics

Testing if the Markov chain $\{\theta^{(g)}, h^{(g)}\}_{g=1}^G$, generated from the MCMC algorithm, converges to the posterior distribution $p(\theta, h|y)$ is very important in empirical study. In this paper we employ the following two methods to check the convergence of the Markov chain.

Geweke (1992)'s Z-scores test is based on a test for equality of the means of the first and last part of a Markov chain (Geweke (1992) suggested using the first 10% and the last 50%). If the samples are drawn from the stationary distribution of the chain, the two means are equal and Geweke's statistic has an asymptotically standard normal distribution.

Heidelberger and Welch (1983)'s stationarity test uses the Cramer-von-Mises statistic to test the null hypothesis that the sampled values come from a stationary distribution. The half-width test calculates a 95% confidence interval for the mean, using the portion of the chain which passed the stationarity test. Half the width of this interval is compared with the estimate of the mean. If the ratio between the half-width and the mean is lower than a small value, for example 0.01, the halfwidth test is passed.

3 Misspecification test

The misspecification test is based on the probability integral transform (see, Rosenblatt (1965)) of the realizations y_{t+1}^o taken with respect to the one-step-ahead prediction density $f(y_{t+1}|Y_t, \theta)$. The probability integral transform, ξ_{t+1} , is simply the cumulative distribution function corresponding to the prediction density $p(y_{t+1}|Y_t, \theta)$ evaluated at y_{t+1}^o : $\xi_{t+1} = \text{Prob}(y_{t+1} \leq y_{t+1}^o | Y_t, \theta)$. For $t = 1, \dots, n$, under the null hypothesis that the true distribution of y_{t+1}^o is $p(y_{t+1}|Y_t, \theta)$ (or equivalently, the model is correctly specified), the ξ_{t+1} converges in distribution to independent and identically distributed uniform random variables. By letting $\varsigma_{t+1} = \Phi^{-1}(\xi_{t+1})$, where Φ is the standard normal cumulative distribution function, a sequence of indepen-

dent $N(0, 1)$ random variables ς_{t+1} is obtained, which are the standardized innovations. The series ς_{t+1} can be used to carry out Box-Ljung, normality and heteroscedasticity tests, among others. This approach of one-day ahead prediction density based misspecification test has been studied, for example, by Smith (1985).

The standardized innovations can be easily calculated. By definition, the one-step ahead prediction density is given by:

$$\begin{aligned} p(y_{t+1}|Y_t, \theta) &= \int p(y_{t+1}|Y_t, h_{t+1}, \theta) p(h_{t+1}|Y_t, \theta) dh_{t+1}, \\ &= \int p(y_{t+1}|Y_t, h_{t+1}, \theta) \left[\int p(h_{t+1}|Y_t, h_t, \theta) p(h_t|Y_t, \theta) dh_t \right] dh_{t+1}, \end{aligned}$$

which can be estimated consistently by: $\frac{1}{M} \sum_{m=1}^M p(y_{t+1}|Y_t, h_{t+1}^{(m)}, \theta)$, where the predicted volatility $h_{t+1}^{(m)}|h_t^{(m)}$ is drawn from $N(\mu + \phi(h_t^{(m)} - \mu), \sigma_\eta^2)$ and $h_t^{(m)}$ is the filtered volatility generated from $p(h_t|\mathcal{F}_t, \mu, \phi, \sigma_\eta^2)$ using auxiliary particle filter (for auxiliary particle filter, see Appendix C). The probability $\text{Prob}(y_{t+1} \leq y_{t+1}^o|Y_t, \theta)$ can then be approximated by:

$$\text{Prob}(y_{t+1} \leq y_{t+1}^o|Y_t, \theta) \cong \frac{1}{M} \sum_{m=1}^M \text{Prob}(y_{t+1} \leq y_{t+1}^o|Y_t, h_{t+1}^{(m)}, \theta).$$

4 The data and priors

The data series consists of 1,584 daily continuously compounded returns, $y_t = \log p_t - \log p_{t-1}$, on China's ShangZheng stock index p_t from January 4, 1999 to August 12, 2005. The annualized mean and annualized standard deviation of the data are 0.67% and 22.89%, respectively. The data exhibits positive skewness with value 0.73, and its kurtosis is 8.35. The p-value of Box-Ljung's serial correlation test on the raw returns is 0.14, and the data rejects both the ARCH test (Engle (1982)) and Jarque-Bera's normality test.

For the parameter σ_η^2 , we use an inverse-gamma conjugate prior with shape $\alpha_\sigma = 2.5$ and scale $\beta_\sigma = 0.025$, which has a mean of 0.167 and a

standard deviation of 0.024.⁴ For the parameter μ we specify a normal prior with hyperparameters $\alpha_\mu = 0$ and $\beta_\mu^2 = 100$. In order to rule out near unit-root behavior, a truncated normal prior with mean $\alpha_\phi = 0$ and variance $\beta_\phi^2 = 1$ is used for the parameter ϕ .

5 Empirical results

5.1 Estimation results

We choose a burn-in period of 2,000 iterations, a follow-up period of 10,000. The sample generated during burn-in period are discard, in order to reduce the influence of the choice of starting point. The parameter estimates are the sample mean of the stored 10,000 posteriors.⁵ In Table 1, we re-

Table 1: Parameter estimates for the *SV* model

Parameter	Mean	SD	NSE	95% CI
μ^{SV}	-8.8892	0.1325	0.0009	(-9.1510,-8.6240)
ϕ^{SV}	0.9373	0.0166	0.0003	(0.9003, 0.9656)
σ_u^{SV}	0.3029	0.0388	0.0008	(0.2336, 0.3856)
$\ln L$	4578.4			

Note: the NSE standards for numerical standard error. The 95% CI denotes the 95% credible interval of posterior distribution. The $\ln L$ denotes the Chib's marginal likelihood.

port the parameter estimates of the basic *SV* model. The estimated μ is -8.8892 , implying the annualized volatility is 0.2251, this value is very close

⁴A prior is conjugate if the full conditional posterior has the same function form as the prior. For example, normal prior with normal likelihood is conjugate, since the full conditional posterior is also normally distribution.

⁵The code is written in C++ and compiled using VC 7.0. The running time using our C++ code for 10,000 iterations is less than two minutes, and the estimates are very close to those gotten from WINBUGS. Both C++ code and WINBUGS code, and in addition the R code which were written for algorithm debugging, are all available by sending E-mail to me.

to the empirical volatility 0.2289. The implied kurtosis of SV model is 6.39, whereas the empirical kurtosis of the data is 8.35, indicating the basic SV model may not be able to capture some large observations. In Appendix D, we give the derivations of unconditional moments and kurtosis of y_t . The volatility process is highly persistent indicated by the estimated ϕ with value 0.9373. This evidence is consistent with the stylized fact of stock returns. In Table 1, the NSE stands for the numerical standard error (see, Geweke (1992)) which measures the inefficiency of the estimated sample average approximating the population mean. As a rule of thumb, the simulation should be run until the NSE for each parameter of interest is less than about 5% of the sample standard deviation.

Figure 1 shows the filtered and smoothed volatility.⁶ The smoothed volatility are obtained directly from MCMC runs by taking the sample average of $\{h_t^{(g)}\}_{g=1}^G$ for each t . The filtered volatility are simulated using the auxiliary particle filter, see Appendix C. In this study, we use 50,000 particles and 250,000 auxiliary variables. The graph shows the expected feature of the filtered volatility lagging the smoothed volatility. Together with the filtered and smoothed volatility, we also give the plot of $|y|$. It shows that the estimated volatility has similar movements as $|y|$.

5.2 Misspecification test

In Table 2 we report the misspecification test using the standardized innovations. The rejections of Jarque-Bera test and BDS test imply the misspecification of the model. From the graph (b) and graph (d) given in Figure 2, we can see that the standardized innovations reveal some outliers, implying

⁶The definitions of smoothing, filtering and forecasting posterior densities are given by:

$$\begin{aligned} \text{Smoothing} & : p(h_t|\mathcal{F}_T), & t = 1, \dots, T, \\ \text{Filtering} & : p(h_t|\mathcal{F}_t), & t = 1, \dots, T, \\ \text{Forecasting} & : p(h_{t+1}|\mathcal{F}_t), & t = 1, \dots, T, \end{aligned}$$

where, \mathcal{F}_t denotes the observations up to time t and h_t is the volatility.

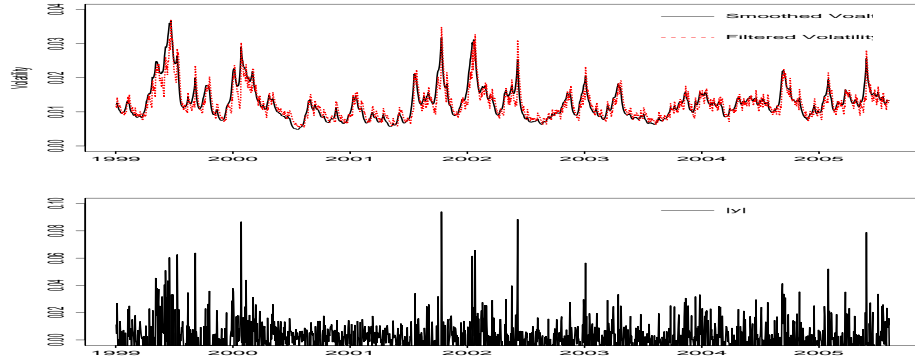


Figure 1: Top: the filtered and smoothed estimate of volatility $\exp(h_t/2)$. Bottom: $|y_t|$, the absolute values of returns.

Table 2: Misspecification test using the standardized innovations

	Box-Ljung Test	Jarque-Bera Test	ARCH test	BDS Test
	(p-value)	(p-value)	(p-value)	(p-value)
<i>SV</i>	0.0943	0.0002	0.2288	0.0110

Note: The BDS test developed by Brock, Dechert, Scheinkman, and LeBaron (1996) is used to test for the null hypothesis of independent and identical distribution (iid).

the basic SV with normal error fails to accommodate some of the data values that have limited daily movements, and a fat-tail error may need. We report the parameter estimates and misspecification test for the SV model with Student-t error in Appendix E. The SV_t passes all misspecification test and it outperforms the basic SV model according to the Chib's marginal likelihood. The implied kurtosis of the SV_t model is 7.96. This value is larger than the one implied from the SV model 6.39, and very close to the empirical kurtosis 8.35, implying the SV_t is more outlier resistant than the basic SV model.

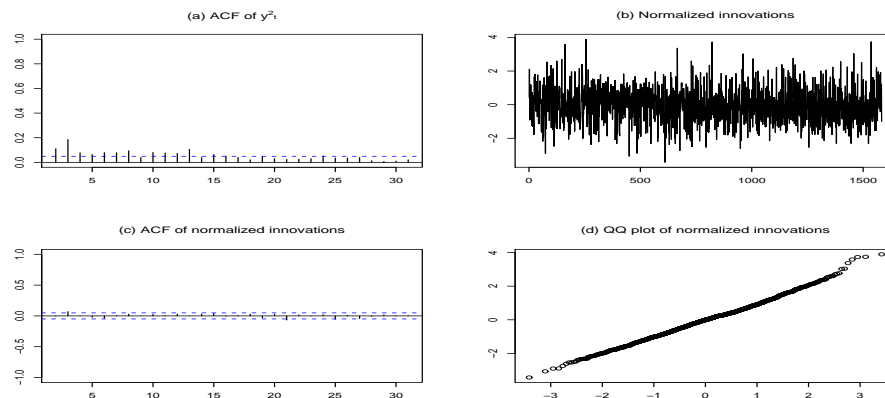


Figure 2:

5.3 Convergence diagnostic

The results of convergence diagnostic are reported in Table 3. All parameters pass both the Geweke's z-scores test and the Heibelberger-Welch's stationarity and half-width tests, that implies the convergence of the chain to its equilibrium distribution, or equivalently to the joint posterior, and consequently confirms the correctness of our estimated results. We implement these tests in R (an open source statistical software under GNU General Public License) using CODA package.

Table 3: Convergence diagnostics

Parameter	Z-scores test		Stationarity and Half-mean test			
	z-scores	p-value	p-value	Mean	Half-width	Ratio
μ^{SV}	0.1987	0.84	0.645	-8.895	0.0028	-0.0003
ϕ^{SV}	0.0329	0.97	0.897	0.928	0.0011	0.0012
σ_u^{SV}	-0.4135	0.68	0.979	0.329	0.0032	0.0096

Note: the half-width test is passed if the ratio less than 0.01.

6 Conclusion

In this study, we give an empirical framework for the estimation of SV model using MCMC technique. This includes the detailed sampling procedure, volatility filtering, convergence diagnostic and misspecification test. The empirical results show that the SV model is misspecified. We attribute this failure to the lack of ability to resist large observations of returns. As a by product of our work, we also estimated the SV model with Student-t error. The SV_t is more outlier resistant than the basic SV model, and it outperforms the basic SV according to Chib's marginal likelihood. Our results also show that the China's stock index exhibits volatility clustering and fat-tail behaviour, these evidences are consistent with the stylized facts of stock returns. But, this data has a positive skewness. Not surprisingly, the Italian and Japanese stock indices also exhibit positive skewness.

The further study can focus on investigating the large observation by introducing fat-tail error, such as Student-t or Generalized error distribution, or by introducing jump components. The study of the positive skewness might be also an interesting topic.

A The derivation of full conditional posteriors

The full conditional posterior for σ_η^2 :

$$\begin{aligned} p(\sigma_\eta^2|y, h, \mu, \phi) &\propto p(h|\mu, \phi, \sigma_\eta^2)p(\sigma_\eta^2), \\ &\propto p(h_1|\mu, \phi, \sigma_\eta^2) \prod_{t=1}^{T-1} p(h_{t+1}|h_t, \mu, \phi, \sigma_\eta^2) \mathcal{IG}(\alpha_\sigma, \beta_\sigma), \end{aligned}$$

the inverse gamma distribution with shape α and scale β has a support $(0, \infty)$, its density is $p(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma} x^{-(\alpha+1)} e^{-\beta/x}, x > 0$,

$$\begin{aligned} &\propto \left(\frac{1}{\sigma_\eta^2} \right)^{\frac{T}{2}} \exp \left(-\frac{(h_1 - \mu)^2(1 - \phi^2)}{2\sigma_\eta^2} - \frac{\sum_{t=1}^{T-1} (h_{t+1} - \mu - \phi(h_t - \mu))^2}{2\sigma_\eta^2} \right) \\ &\quad \frac{(\beta_\sigma)^{\alpha_\sigma} e^{-\beta_\sigma/\sigma_\eta^2}}{\Gamma(\alpha_\sigma)(\sigma_\eta^2)^{\alpha_\sigma+1}}, \end{aligned}$$

where, β_σ and α_σ are the hyperparameters, which are constant and should be specified by researcher, hence, the terms $(\beta_\sigma)^{\alpha_\sigma}$ and $\Gamma(\alpha_\sigma)$ are constant,

$$\begin{aligned} &\propto \exp \left(-\frac{\beta_\sigma + \frac{1}{2}(h_0 - \mu)^2(1 - \phi^2) + \frac{1}{2} \sum_{t=1}^{T-1} (h_{t+1} - \mu - \phi(h_t - \mu))^2}{\sigma_\eta^2} \right) \\ &\quad \left(\frac{1}{\sigma_\eta^2} \right)^{(\alpha_\sigma + \frac{T}{2})+1}, \\ &\propto \mathcal{IG}(\hat{\alpha}_\sigma, \hat{\beta}_\sigma), \end{aligned}$$

where,

$$\hat{\alpha}_\sigma = \alpha_\sigma + \frac{T}{2},$$

and,

$$\hat{\beta}_\sigma = \beta_\sigma + \frac{1}{2}(h_1 - \mu)^2(1 - \phi^2) + \frac{1}{2} \sum_{t=1}^{T-1} (h_{t+1} - \mu - \phi(h_t - \mu))^2.$$

The methodology of deriving the full conditional posteriors μ is similar to σ_η^2 .

$$\begin{aligned}
p(\mu|y, h, \phi, \sigma_\eta^2) &\propto p(h|\mu, \phi, \sigma_\eta^2)p(\mu), \\
&\propto p(h_1|\mu, \phi, \sigma_\eta^2) \prod_{t=1}^{T-1} p(h_{t+1}|h_t, \mu, \phi, \sigma_\eta^2) \text{N}(\alpha_\mu, \beta_\mu), \\
&\propto \exp \left\{ -\frac{(h_1 - \mu)^2(1 - \phi^2)}{2\sigma_\eta^2} - \frac{\sum_{t=1}^{T-1} (h_{t+1} - \mu - \phi(h_t - \mu))^2}{2\sigma_\eta^2} \right\} \\
&\quad \exp \left\{ -\frac{(\mu - \alpha_\mu)^2}{2\beta_\mu^2} \right\}, \\
&\propto \exp \left\{ -\frac{1}{2} \left[\underbrace{\mu^2 \left(\frac{1 - \phi^2 + (T-1)(1-\phi)^2}{\sigma_\eta^2} + \frac{1}{\beta_\mu^2} \right)}_A \right. \right. \\
&\quad \left. \left. - 2\mu \left(\underbrace{\frac{h_1(1 - \phi^2) + (1 - \phi) \sum_{t=1}^{T-1} (h_{t+1} - \phi h_t)}{\sigma_\eta^2} + \frac{\alpha_\mu}{\beta_\mu^2}}_B \right) \right] \right\}, \\
&\propto \text{N} \left(\frac{B}{A}, \frac{1}{A} \right).
\end{aligned}$$

The full conditional posterior for ϕ :

$$\begin{aligned}
p(\phi|y, h, \sigma_\eta^2, \mu) &\propto p(h|\mu, \phi, \sigma_\eta^2)p(\phi), \\
&\propto p(h_1|\mu, \phi, \sigma_\eta^2) \prod_{t=1}^{T-1} p(h_{t+1}|h_t, \mu, \phi, \sigma_\eta^2) \text{N}(\alpha_\phi, \beta_\phi^2) I_{(-1, +1)}(\phi), \\
&\propto \exp \left\{ -\frac{(h_1 - \mu)^2(1 - \phi^2)}{2\sigma_\eta^2} - \frac{\sum_{t=1}^{T-1} (h_{t+1} - \mu - \phi(h_t - \mu))^2}{2\sigma_\eta^2} \right\} \\
&\quad \exp \left\{ -\frac{(\phi - \alpha_\phi)^2}{2\beta_\phi^2} \right\} I_{(-1, +1)}(\phi), \\
&\propto \exp \left\{ -\frac{1}{2} \left[\phi^2 \underbrace{\left(\frac{-(h_1 - \mu)^2 + \sum_{t=1}^{T-1} (h_t - \mu)^2}{\sigma_\eta^2} + \frac{1}{\beta_\phi^2} \right)}_C \right. \right. \\
&\quad \left. \left. - 2\phi \underbrace{\left(\frac{\sum_{t=1}^{T-1} (h_{t+1} - \mu)(h_t - \mu)}{\sigma_\eta^2} + \frac{\alpha_\phi}{\beta_\phi^2} \right)}_D \right] \right\} I_{(-1, +1)}(\phi), \\
&\propto \text{N} \left(\frac{D}{C}, \frac{1}{C} \right) I_{(-1, +1)}(\phi).
\end{aligned}$$

B Acceptance-rejection method

The most difficult part of the Gibbs sampler is to effectively sample the latent state h_t from its full conditional posterior. We employ the accept-reject sampling procedure introduced by Kim, Shephard, and Chib (1998).

Given a target distribution $f(x)$, from which we want to generate random variables, but directly sampling is difficult. Assuming we have another instrumental distribution $g(x)$, and we can easily simulate random variables from this distribution. Then the acceptance-rejection method can be applied to generate random variates from the target distribution, once the instrumental distribution times a real valued constant $c \geq 1$ blankets the target distribution, that is $1 \leq \frac{cg(x)}{f(x)}$. The algorithm is given by:

1. Sample x from $g(x)$ and u from $\text{Unif}(0, 1)$,

2. If $u < f(x)/cg(x)$, accept x as a realization of $f(x)$

else, repeat 1.

For the SV model, the full conditional posterior of h_t is:

$$\begin{aligned} & p(h_t|y, h_{-t}, \theta), \\ & \propto p(y_t|h_t, \theta)p(h_t|h_{-t}, \theta), \\ & = \frac{1}{\sqrt{2\pi \exp(h_t)}} \exp\left(-\frac{y_t^2}{2 \exp(h_t)}\right) p(h_t|h_{-t}, \theta), \\ & = f^{**}(y_t, h_t, \theta)p(h_t|h_{-t}, \theta), \end{aligned}$$

where, $\log f^{**}(y_t, h_t, \theta) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} h_t - \frac{y_t^2}{2} \exp(-h_t)$. Due to the convexity of $\exp(-h_t)$, we can apply Taylor expansion for $\exp(-h_t)$ around α_t and get $\log f^{**}(y_t, h_t, \theta) \leq \log g^{**}(y_t, h_t, \theta)$, where $\log g^{**}(y_t, h_t, \theta) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} h_t - \frac{y_t^2}{2} \{\exp(-\alpha_t)(1 + \alpha_t - h_t \exp(-\alpha_t))\}$. Since $p(h_t|h_{-t}, \theta) = p_N(h_t|\alpha_t, \beta^2)$, see (7), we have:

$$\begin{aligned} f^{**}(y_t, h_t, \theta)p(h_t|h_{-t}, \theta) & \leq g^{**}(y_t, h_t, \theta)p_N(h_t|\alpha_t, \beta^2), \\ & = kp_N(h_t|\alpha_t^*, \beta^2), \end{aligned}$$

where, α_t is given in (7), k is a real valued constant, and $p_N(h_t|\alpha_t^*, \beta^2)$ is the instrumental distribution, which is normally distributed with mean $\alpha_t^* = \alpha_t + \frac{\beta^2}{2}(y_t^2 \exp(-\alpha_t) - 1)$ and its variance β^2 is given in (7). Note that the constant term $\frac{1}{k}$ can be omitted, since $p(h_t|y, h_{-t}, \theta) \propto f^{**}(y_t, h_t, \theta) \propto \frac{1}{k} f^{**}(y_t, h_t, \theta)$.

The instrumental distribution $p_N(h_t|\alpha_t^*, \beta^2)$ blankes the target distribution $f^{**}(y_t, h_t, \theta)p(h_t|h_{-t}, \theta)$, and hence the acceptance-rejection method can be applied. The acceptance probability is:

$$\begin{aligned} & \text{Prob}\left(\frac{f^{**}(y_t, h_t, \theta)p(h_t|h_{-t}, \theta)}{p_N(h_t|\alpha_t^*, \beta^2)} \geq U\right), \quad U \sim \text{Unif}[0, 1] \\ & = \frac{f^{**}(y_t, h_t, \theta)}{g^{**}(y_t, h_t, \theta)} \end{aligned}$$

C The auxiliary particle filter

The goal is to sample random variates $\{h_t^1, \dots, h_t^M\}$ from the filter distribution $p(h_t|\mathcal{F}_t, \theta)$, where \mathcal{F}_t denotes the available information up to time t . In this study, we employ the auxiliary particle filter introduced by Pitt and Shephard (1999).

1. Given $\{h_{t-1}^1, \dots, h_{t-1}^M\}$ from $p(h_{t-1}|\mathcal{F}_{t-1}, \mu, \phi, \sigma_\eta^2)$ calculate

$$\begin{aligned}\hat{h}_t^m &= \mu + \phi(h_{t-1}^m - \mu), \\ w_j &= p(y_t|\hat{h}_t^m), \quad m = 1, \dots, M\end{aligned}$$

2. Sample R times the indexes $1, 2, \dots, M$ with probability proportional to $\{w_m\}$ and get a sample $\{i_1, \dots, i_R\}$. Associate the sampled indexes with corresponding \hat{h}_t^m and get a sample $\{\hat{h}_t^{i_1}, \dots, \hat{h}_t^{i_R}\}$. Associate the sampled indexes with corresponding h_t^m and get another sample $\{h_t^{i_1}, \dots, h_t^{i_R}\}$. In this study, we take R to be five times larger than M .

3. For each value of i_r simulate

$$\check{h}_t^r \sim N(\mu + \phi(h_{t-1}^{i_r} - \mu), \sigma_\eta^2), \quad r = 1, \dots, R$$

4. Resample $\{\check{h}_t^{i_1}, \dots, \check{h}_t^{i_R}\}$ M times with probability proportional to

$$\frac{p(y_t|\check{h}_t^r)}{p(y_t|\hat{h}_t^{i_r})}, \quad r = 1, \dots, R,$$

to produce the filtered sample $\{h_t^1, \dots, h_t^M\}$ from $p(h_t|\mathcal{F}_t, \mu, \phi, \sigma_\eta^2)$.

Note that, for the basic SV model, $p(y_t|h_t) \sim N(0, \exp(h_t))$.

D The unconditional moments of y_t

The kurtosis is defined as: $y_{\text{kurtosis}} = \frac{E[(y - E[y])^4]}{E[(y - E[y])^2]^2}$.

The SV model:

$$\begin{aligned} y_t &= \exp(h_t/2)u_t, & u_t &\sim N(0, 1), \\ h_t &\sim N(\mu_h, \sigma_h^2), \\ \mu_h &= \mu, & \sigma_h^2 &= \frac{\sigma_\eta^2}{1 - \phi^2}, \end{aligned}$$

the raw moments of y_t are:

$$\begin{aligned} E[y_t^2] &= E[\exp(h_t)]E[u_t^2] = \exp(\mu_h + 0.5\sigma_h^2), \\ E[y_t^4] &= E[\exp(2h_t)]E[u_t^4] = 3\exp(2\mu_h + 2\sigma_h^2). \end{aligned}$$

and the central moments of y_t are equal to their corresponding raw moments. Note that, the derivation of $E[\exp(ch_t)]$, where c is a real valued constant, $h_t \sim N(\mu_h, \sigma_h^2)$. Let $\log \psi_t \equiv ch_t$, then $\log \psi_t \equiv ch_t \sim N(c\mu_h, c^2\sigma_h^2)$ and the first moment of this log-normal distribution is $E[\psi_t] = E[\exp(ch_t)] = \exp(c\mu_h + 0.5c^2\sigma_h^2)$.

The SV_t model:

$$\begin{aligned} y_t &= \exp(h_t/2)\sqrt{\lambda_t}u_t, & u_t &\sim N(0, 1), \\ \lambda_t &\sim \mathcal{IG}(v/2, v/2), & v &> 2, \\ h_t &\sim N(\mu_h, \sigma_h^2), \end{aligned}$$

the raw moments of y_t are:

$$\begin{aligned} E[y_t^2] &= E[\exp(h_t)]E[\lambda_t]E[u_t^2] = \exp(\mu_h + 0.5\sigma_h^2)\frac{v}{v-2}, \\ E[y_t^4] &= E[\exp(2h_t)]E[\lambda_t^2]E[u_t^4] = 3\exp(2\mu_h + 2\sigma_h^2)\frac{v^2}{(v-2)(v-4)}. \end{aligned}$$

and the central moments of y_t are equal to their raw moments. Note that, the first and the second raw moments of $X \sim \mathcal{IG}(\alpha, \beta)$ are:

$$E[X] = \frac{\beta}{\alpha - 1} \quad E[X^2] = \frac{\beta^2}{(\alpha - 1)(\alpha - 2)}.$$

E The parameter estimates and misspecification tests of SV model with Student-t error

The model is given by:

$$\begin{aligned} y_t &= \exp(h_t/2)\sqrt{\lambda_t}u_t, \quad t = 1, \dots, T, \\ h_{t+1} &= \mu + \phi(h_t - \mu) + \sigma_\eta \eta_t, \quad t = 1, \dots, T, \\ \lambda_t &\stackrel{i.i.d.}{\sim} \mathcal{IG}(v/2, v/2), \quad v > 2(u_t, \eta_t) \stackrel{i.i.d.}{\sim} N(0, I_2), \end{aligned}$$

where we assume that λ_t is distributed as i.i.d. inverse-gamma random variable, or equivalently $v/\lambda_t \sim \chi_v^2$. This implies that the marginal distribution of $\sqrt{\lambda_t}u_t$ is Student-t with degree of freedom v . The assumption $v > 2$ is to ensure the existence of second moment.

Table 4: Parameter estimates for the SV_t model

Parameter	Mean	SD	ts-SE	95% CI
μ^{SV_t}	-9.0976	0.1635	0.0015	(-9.4190,-8.7740)
ϕ^{SV_t}	0.9642	0.0113	0.0002	(0.9391, 0.9830)
$\sigma_u^{SV_t}$	0.2068	0.0309	0.0007	(0.1535, 0.2740)
v	8.5034	2.1285	0.0408	(5.6160, 13.760)
$\ln L$	4583.9			

Note: the ts-SE standards for time-series standard error, due to Geweke (1992). The 95% CI denotes the 95% credible interval of posterior distribution. The $\ln L$ denotes the Chib's marginal likelihood given by $\ln L = \ln p(y|\theta^*) + \ln p(\theta^*) + \ln p(\theta^*|y)$.

Table 5: Misspecification test of SV_t model

	Box-Ljung Test	Jarque-Bera Test	ARCH test	BDS Test
	(p-value)	(p-value)	(p-value)	(p-value)
SV_t	0.1050	0.1010	0.1786	0.1732

Note: The BDS test developed by Brock, Dechert, Scheinkman, and LeBaron (1996) is used to test for the null hypothesis of independent and identical distribution (iid).

References

- Andersen, T., H. Chung, and B. Sorensen, 1999, Efficient method of moments estimation of a stochastic volatility model: A monte carlo study, *Journal of Econometrics* 91, 61–87.
- Brock, W.A., W.D. Dechert, J.A. Scheinkman, and B. LeBaron, 1996, A test for independence based on the correlation dimension, *Econometric Reviews* 15, 197–235.
- Engle, R.F., 1982, Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation, *Econometrica* 50, 987–1007.
- Geman, S., and D. Geman, 1984, Stochastic relaxation, gibbs distributions, and the bayesian restoration of images, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 721–741.
- Geweke, J., 1992, *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments* . in Bayesian Statistics 4 (Oxford University Press: U.K.) pp. 169-193.
- Heidelberger, P., and P. Welch, 1983, Simulation run length control in the presence of an initial transient, *Operations Research* 31, 1109–44.

- Jacquier, E., N.G. Polson, and P.E. Rossi, 1994, Bayesian analysis of stochastic volatility models, *Journal of Business and Economic Statistics*(with discussion) 12, 371–417.
- Kim, S., N. Shephard, and S. Chib, 1998, Stochastic volatility: likelihood inference and comparison with arch models, *Review of Economic Studies* 65, 361–393.
- Pitt, M., and N. Shephard, 1999, Filtering via simulation: auxiliary particle filter, *Journal of the American Statistical Association* 94, 590–599.
- Smith, J.Q., 1985, Diagnostic checks of non-standard time series models, *Journal of Forecasting* 4, 283–291.
- Taylor, S.J., 1982, *Financial returns modelled by the product of two stochastic processes—a study of daily sugar prices 1961-75* . In Anderson, O. D. (ed.), *Time Series Analysis: Theory and Practice* (1, 203-226, North-Holland: Amsterdam).
- Tierney, L., 1994, Markov chains for exploring posterior distributions, *The Annals of Statistics* 22, 1701–1762.